Essays on Human Capital in Macroeconomics

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

Lukas Mahler

Frühjahrs-/Sommersemester 2024

Abteilungssprecher: Prof. Klaus Adam, Ph.D.
Vorsitzende der Disputation: Prof. Michèle Tertilt, Ph.D.
Referent: Prof. Antonio Ciccone, Ph.D.
Koreferent: Prof. Minchul Yum, Ph.D.
Tag der Disputation: 3. Juni, 2024

Acknowledgments

I am profoundly grateful to my advisors, Antonio Ciccone, Michèle Tertilt, and Minchul Yum, for their unwavering support, encouragement, and continuous investment of time and effort during my Ph.D. studies. Under Antonio's mentorship, I learned to engage deeply with my areas of interest, and approach economic research with structure and purpose. His ability to challenge my perspectives while providing guidance in a friendly and supportive manner has been invaluable to my growth as an academic researcher. Antonio's high standard of academic excellence serves as a beacon for my future endeavors. Michèle's support was steadfast, and profoundly encouraging. Her mentorship extended beyond guidance, as she fostered an academic environment conducive to productivity and led by example. Her extensive network proved immensely beneficial, and her encouragement to take part in conferences and workshops, coupled with her insightful feedback on projects and presentations, significantly facilitated my academic journey. From the outset of my Ph.D., Minchul played a pivotal role in my development. Initially as my mentor, offering unwavering support in both academic and personal matters and later on as a co-author on the first and third chapter of this dissertation. I have tremendously benefited from working with him, as he not only generously imparted knowledge on various tools used in modern macroeconomic research but also provided invaluable guidance on navigating the academic profession as a young researcher. Much of my technical expertise is owed to him.

I am privileged to have collaborated with Suzanne Bellue on the second chapter of this dissertation, and I am sincerely grateful for our productive partnership. Additionally, I have had the pleasure of being part of a talented, supportive and fun cohort of Ph.D. students, including Andrés, Giovanni, Jacopo, Jasmina, Jonathan, Laura, Oliver, and Tommaso, whose camaraderie greatly enriched my doctoral experience. Furthermore, I extend my gratitude to many other (former) members of the economics department at the University of Mannheim, including Klaus Adam, Effrosyni Adamopoulou, Philip Ager, Miren Azkarate-Askasua, Antoine Camous, Hans-Peter Grüner, Andreas Gulyas, Anne Hannusch, David Koll, Matthias Meier, Ana Morena-Maldonado, Arthur Seibold, Sebastian Seitz, and the participants of various seminars, for their valuable feedback and engaging discussions. Their contributions have been pivotal in shaping my dissertation.

I would also like to express my appreciation to Ulrich Kehl, Golareh Khalilpour, Marion Lehnert, Stephanie Lilwall, Caroline Mohr, and Astrid Reich for their administrative support throughout my tenure at Mannheim University. Additionally, I am grateful to Michelle Sovinsky for her guidance during the job market process. I also acknowledge the financial support provided by the German Research Foundation (through the CRC-TR-224 project A04) and the SFB 884 Political Economy of Reforms during the initial stages of my Ph.D.

Lastly, I want to thank my family: my parents, siblings, relatives, and my partner, for their unwavering support, encouragement, and kindness throughout my academic journey.

Contents

Preface

In this dissertation, I examine how differences among individuals in various dimensions of human capital, such as health or education, evolve throughout all stages of the life cycle and endure across generations. I explore how these differences are influenced by policies and how they interact with economic inequalities, such as those in income or wealth. To investigate these questions, I employ quantitative methods used in modern macroeconomic research, linking models of individual decision-making and well-being with micro–, and macroeconomic data. The dissertation consists of three self-contained chapters.

Chapter 1 is titled "Lifestyle Behaviors and Wealth-Health Gaps in Germany" and is co-authored with Minchul Yum. In this chapter, we build and estimate a life-cycle model of endogenous health and wealth formation to study the observed strong positive association between wealth and health in Germany. Even though health insurance is universal in Germany and out-of-pocket medical expenses are low, the relationship between wealth and health that we document –and refer to as wealth-health gaps– is substantial. Data from a panel of representative households suggests that individual health behaviors or lifestyles may play a role in mediating this relationship: Richer households tend to engage in more health-promoting lifestyles, such as physical exercise, health-conscious nutrition, or abstention from smoking, than poorer households, even conditional on education, occupation, or other observable characteristics. For that reason, we include such lifestyles in our framework and model them as individual health efforts whose adjustment over time is costly. This allows us to capture the habitual nature of healthy and unhealthy lifestyles that we observe in the data.

Our estimation strategy requires the model to replicate the joint distribution of individual earnings, labor supply, and these health efforts across age, health, and education status but we do not target the distribution of wealth by health directly. Our model with endogenous health can rationalize between three quarters and up to the entire empirical wealth-health gaps, depending on age and the point of the wealth distribution we measure it. In contrast, a model with a purely exogenous health process, where individuals cannot influence the probability of being healthy in the future, can only account for around two-thirds of the gaps.

In the model, wealth-health gaps then arise on the one hand because good health outcomes make individuals more productive and labor supply less costly, leading to higher labor earnings and wealth. Similarly, good health outcomes affect the incentives to accumulate wealth because of higher expected longevity and quality of life in the future. Quantitatively, our estimated model shows that it is the second channel working through savings that contributes most to the observed wealth-health relationship, as it can account for around half of the wealth-health gap. The other channel, working through earnings is particularly relevant for the young and asset-poor for whom earnings provide the main basis for wealth accumulation. On the other hand, the wealth-health relationship can also be driven by channels that operate in the direction from wealth to health. In particular, we demonstrate theoretically and quantitatively that the perspective of future utility driven by high levels of wealth incentivizes agents to engage in more health-promoting behaviors. Thus, lifestyles can act as a dynamic amplification vehicle, which fuels the wealth-health gaps. Quantitatively, we find that eliminating variations in individual lifestyle behaviors reduces the wealth-health gaps by between 12% and 29%, as compared to the economy, in which individuals can adjust their health efforts.

Chapter 2 is titled "Efficiency and Equity of Education Tracking: A Quantitative Analysis," and is co-authored with Suzanne Bellue. In this chapter, we investigate the long-term aggregate and distributional effects of school tracking policies. Tracking policies describe the separation of school children into different school tracks or types of schools, that are typically associated with different curricula and often lead to different higher education and labor market outcomes. While tracking is a common feature of education policy in many countries, there is a striking variety in the age at which children are tracked across countries. In countries where tracking occurs very early, such as Germany, tracking policies are often held responsible for low intergenerational mobility and persistent inequality.

To evaluate the effects of a broad reform to the tracking age, we build and calibrate a model of overlapping generations, in which school children can attend different tracks in secondary school. The school track decision, made by parents, can affect the skills a school child accumulates in two ways: through the teaching level, or instruction pace, in each track, and through the skills of the peers in the track. The instruction pace in each track is determined by a policy choice aimed at maximizing aggregate skills. At the same time, the development of skills is subject to unforeseeable shocks in each period. We demonstrate analytically that this skill formation technology can rationalize the sometimes ambiguous evidence from the empirical literature regarding the effects of tracking on the distribution of learning outcomes. The chosen school track also influences higher education and subsequent labor market opportunities as college education is less costly for graduates from only one of the tracks. Wages for college and non-college skilled workers are determined in general equilibrium and are therefore influenced by the distribution of human capital across all workers. The wages also affect the school track incentives for the next generations.

We estimate the parameters of the skill formation technology using data from a representative panel of school children in Germany. This dataset comprises independently and repeatedly administered achievement tests across various domains, which serve as noisy measures of the latent skill variable. The remaining model parameters are calibrated on a variety of micro and macro data for Germany in the 2010s. The model matches well the cross-sectional distribution of school skills and labor earnings, intergenerational mobility in terms of intergenerational income elasticities and correlations between school tracks and parental education and income, as well as the transitions of school children through the education system.

Using counterfactual exercises, we demonstrate that a postponement of school tracking in Germany by four years, from age ten to age fourteen, entails an efficiency-mobility trade-off in the long run. On the one hand, intergenerational mobility improves, as the economic outcomes of children become less correlated with those of their parents. These mobility gains arise primarily because there is less heterogeneity in skill accumulation during longer comprehensive school, resulting in smaller differences in skills across children from different parental backgrounds and, across tracks once they are tracked later. On the other hand, postponing tracking leads to a drop in aggregate GDP, accompanied by a reduction in aggregate human capital in the long run. These losses arise because prolonged learning in a comprehensive school track foregoes efficiency gains from tailored instruction levels in an early tracking system. Moreover, we demonstrate that this trade-off only materializes in the "long-run", when the college wage premium has had time to adjust to the policy-induced change in the supply of skills.

Chapter 3 is titled "Aggregate and Distributional Effects of School Closure Mitigation Policies: Public versus Private Education", and is co-authored again with Minchul Yum. This shorter chapter is motivated by recent research demonstrating that the school closures that many countries implemented in the wake of the Covid pandemic can lead to substantial, and unequally distributed, long-term costs as measured for example in terms of income. We illustrate that the long-term effects of policies proposed to mitigate these school closure costs, such as additional periods of public schooling, crucially depend on the interaction between public and private investments into children's human capital. To that end, we set up a simple model of human capital formation, in which the human capital of school children can be raised both through parental investments, for example by paying for private tuition or by spending more time with their children, and through public investments in the form of public schooling. Importantly, these two types of inputs are imperfectly substitutable in the production of human capital. The degree of substitutability constitutes a central parameter in our analysis, as existing literature provides mixed findings, especially when considering different time horizons, different ages, or different domains of skills.

We calibrate the model to match existing estimates of the long-term effects of school closures on cross-sectional inequality and intergenerational mobility. We then illustrate that a targeted school closure mitigation strategy, such as a means-tested subsidy to private education, increases intergenerational mobility and reduces cross-sectional inequality. In contrast, the long-term effects of a program that universally prolongs school days depend crucially on the substitutability between private and public investments in human capital formation. Specifically, when private and public inputs are complements, longer schooling time can even hamper mobility and increase inequality, as it disproportionately benefits children from higher socio-economic backgrounds. Depending on the welfare function of the social planner, such policies may thus have undesirable long-term consequences.

Chapter 1

Lifestyle Behaviors and Wealth-Health Gaps in Germany[1]

WITH MINCHUL YUM[2]

Abstract: We document significant gaps in wealth across health status over the life cycle in Germany—a country with a universal healthcare system and negligible out-of-pocket medical expenses. To investigate the underlying sources of these wealth-health gaps, we build a heterogeneous-agent life-cycle model in which health and wealth evolve endogenously. In the model, agents exert efforts to lead a healthy lifestyle, which helps maintain good health status in the future. Effort choices, or lifestyle behaviors, are subject to adjustment costs to capture their habitual nature in the data. We find that our estimated model generates the great majority of the empirical wealth gaps by health and quantify the role of earnings and savings channels through which health affects these gaps. We show that variations in individual health efforts account for around a quarter of the model-generated wealth gaps by health, illustrating their role as an amplification mechanism behind the gaps.

[1] We thank three anonymous referees and the editor, Chad Jones, for their helpful comments that have improved the paper substantially. We also thank Yongsung Chang, In Choi, Antonio Ciccone, Youngsoo Jang, Soojin Kim, Sang Yoon (Tim) Lee, Chiara Malavasi, Serena Rhee, Michele Tertilt, Hitoshi Tsujiyama, Johanna Wallenius, Nicolas Ziebarth and participants at various conferences and seminars for useful comments and discussions. Financial support from the German Research Foundation (DFG) through CRC TR 224 (Projects A03 and A04) and the SFB 884 Political Economy of Reforms is gratefully acknowledged.

[2] Department of Economics, University of Southampton and CEPR (m.yum@soton.ac.uk)

1.1 Introduction

A large body of literature across economics, sociology, and public health demonstrates strong positive associations between financial and health status at the individual level. For example, De Nardi et al. (2023) document substantial differences in wealth over the life cycle in the United States between men with a high school degree who report being in good health and those in poor health. In this paper, we show that large gaps in wealth by health exist in Germany as well. These gaps appear not only within the nationally representative sample but also within education groups. The gaps begin to open up at around the age of 25 and grow over the life cycle before stabilizing after retirement. For example, median wealth among *healthy* 60-64-year-olds (100,000 EUR) amounts to more than three times that of *unhealthy* individuals in the same age group (31,000 EUR).

What explains such large gaps in a country like Germany, characterized by universal health insurance, low out-of-pocket medical expenses, and generous sickness benefits (OECD, 2019)? Existing studies on the positive relationship between health and wealth have tended to focus on the U.S., highlighting the role of large out-of-pocket medical expenditures and unequal access to health insurance (e.g. De Nardi et al., 2010), or the unilateral effect of health on labor supply and productivity coupled with the availability of disability insurance (Hosseini et al., 2021).[3] In this paper, we employ a structural framework in which individuals' wealth and health evolve endogenously over the life cycle to investigate lifestyle behaviors as potential drivers of these observed wealth-health gaps.

Our model explicitly allows the possibility of individuals influencing their health evolution through their health-related lifestyle behaviors (Cawley and Ruhm, 2011; Cole et al., 2019) in an otherwise standard heterogeneous-agent life-cycle framework. We include these endogenous health behaviors given that in Germany, as in most developed countries, morbidity and mortality are predominantly attributed to individuals' behavioral risk factors, including dietary risks, smoking, and physical inactivity (Darden et al., 2018; Kvasnicka et al., 2018; OECD, 2019). Furthermore, behavioral health risks tend to be more common among people of low socio-economic status, with evidence suggesting that divergences in health behaviors have accelerated in recent years (Lampert et al., 2018). It has thus become ever more important to understand the consequences of healthy lifestyles not only for health inequality, but also for wealth inequality. Our quantitative theoretical framework allows to shed greater light on these empirical observations on health and wealth inequality.

In the model, individual health efforts increase the probability of being healthy in the

[3] For a comprehensive review of the potential mechanisms underlying the positive relationship between health and socio-economic status more generally, see, for example, Cutler et al. (2011).

future. Good health, in turn, raises survival probability, affects labor income through productivity and the disutility of working, and complements utility from consumption. These channels influence economic resources through labor supply choices and affect savings decisions, both of which shape wealth and health inequality. As a higher fraction of individuals maintain the same lifestyle behaviors over time in the data, in our model health effort adjustment is subject to stochastic adjustment costs. This allows us to capture healthy (e.g., physical exercise) and unhealthy (e.g., smoking) lifestyle habits. Agents differ along several fixed dimensions including education, discount factor, productivity type, and health type. We include such ex-ante heterogeneity to account for additional forces driving the life cycle evolution of health and wealth.

We estimate our model using the method of simulated moments and information from the German Socio-Economic Panel. Our estimated model is consistent with a number of salient features in the data. For example, the model-generated data align with the observed joint evolution of labor supply and earnings by health and education over age, and match the empirical age pattern of average health effort choices by education. Furthermore, the model replicates the degree of wealth accumulation as well as wealth and income inequality seen in Germany. It also reproduces more detailed aspects of effort choices, such as its dispersion, persistence over time and the share of individuals making large positive and negative adjustments or no adjustments.

We find that the estimated model accounts for between 75% to 100% of the observed wealth-health gaps in the data, depending at which point of the distribution and which age this is measured. In contrast, an estimated model with comparable richness in heterogeneity but without lifestyle behaviors and thus purely exogenous health transitions explains less than two thirds of the empirical gaps, highlighting our baseline model's ability to rationalize observed wealth-health gaps. We then investigate two channels behind the wealth-health gaps that work primarily from health to wealth. On the one hand, good health outcomes are associated with higher labor earnings, as a result of both higher labor supply and higher productivity. This translates into larger wealth. On the other hand, good health outcomes also affect the incentives to accumulate wealth because of higher expected longevity and improved quality of life in the future. Having illustrated these channels using a conceptual, simple two-period model, we conduct counterfactual exercises using our estimated life-cycle model to quantify their relative importance. We find that the second channel working through savings contributes quantitatively most to the wealth-health relationship, accounting for, on average around 50% of the gap. The other channel that works through earnings is particularly relevant for the young and asset-poor agents for whom earnings provide the main basis for wealth accumulation.

Finally, motivated by our empirical evidence suggesting the potential role of lifestyle behaviors as a dynamic amplification vehicle which fuels the wealth-health gaps, we run another counterfactual experiment that quantifies the extent to which heterogeneity in lifestyle behaviors accounts for the wealth-health gaps. We find that eliminating variations in individual lifestyle behaviors reduces the wealth-health gaps by between 12% and 29%, as compared to the baseline model economy. This significant effect demonstrates the role of lifestyle behaviors that operate in the direction from wealth to health: wealthier individuals engage in more health-promoting efforts, which dynamically feeds back into better health in the presence of the earnings and savings channels. We further demonstrate, both theoretically and quantitatively, that the anticipation of future utility resulting for example from exogenous changes in wealth could prompt agents to modify current lifestyle behaviors, thereby influencing the health distribution and the wealth-health gaps.

Our paper primarily intersects with a growing literature that augments structural life-cycle models with idiosyncratic health risk to study the aggregate and distributional economic effects of health and health-related policies. Much early research in this direction has focused on the influence of health and mortality risk on the labor supply and savings of people around retirement age (French, 2005; French and Jones, 2011; De Nardi et al., 2010; Kopecky and Koreshkova, 2014). More recent studies analyze rising health care expenditures and explore specific questions regarding the implementation and economic consequences of health care programs in the U.S.[4] Capatina (2015) and De Nardi et al. (2023) endeavor to quantify the accumulated, life-time consequences of health, and calibrate their models to U.S. data. While Capatina (2015) highlights the importance of the productivity and time endowment channels that influence labor supply and precautionary savings, De Nardi et al. (2023) find that a substantial degree of ex-ante heterogeneity and a rich health process are required to be able to match the observed wealth-health gradient in the U.S. Building on their work, we empirically document and study inequality in health and wealth in the case of Germany. Notably, while De Nardi et al. (2023) study the interaction between wealth and health in an exogenous health framework, we study this in a model with endogenous health.

In this regard, our paper is closely related to several studies that endogenize health through some form of individuals' effort choices in a structural framework. We build, for example, on Cole et al. (2019), who similarly construct a model with endogenous effort choices but focus on a very different research question; namely, the interaction between labor

[4] See e.g., Hall and Jones (2007); Attanasio et al. (2010); Kitao (2014); Zhao (2014); Jung and Tran (2016); Pashchenko and Porapakkarm (2017); Jang (2023). Much work has also been devoted to understanding the dynamics of the insurance incentive trade-offs associated with health or disability insurance, again with a focus on the U.S., see e.g. Low and Pistaferri (2015); Cole et al. (2019).

market and health insurance policies. In addition to this work, a number of recent studies, including Capatina et al. (2020), Hai and Heckman (2022), and Margaris and Wallenius (2023), highlight the interaction between health and human capital accumulation and the role of the latter in explaining observed socio-economic gradients in health. We follow these insights by including two education groups in our analysis. We focus, however on the relation between health and wealth, rather than earnings, as wealth provides a more comprehensive assessment of the accumulated costs of poor health.

The aforementioned literature tends to look at the U.S., and often finds that health insurance is a crucial mechanism that amplifies the two-way relationship between health and earnings along the income distribution. For example, several studies, including Prados (2018), Chen et al. (2022), and Ozkan (2017), use structural models for policy counterfactual experiments and conclude that a switch to more universal health care coverage could substantially lower health-related income inequality.

Given this, Germany offers a particularly interesting case for studying the wealth-health relationship. Most notably, it is compulsory by law for all citizens and residents to have health insurance in Germany.[5] The country moreover mandates health insurance providers to cover a relatively generous package of benefits compared to international standards. In general, Germany reports low levels of self-reported unmet medical needs and low out-of-pocket medical expenses relative to its European neighbors (OECD, 2019).[6] Despite these, we document that gaps in health-related outcomes between members of low and higher socio-economic groups are sizeable in Germany. In examining a novel mechanism—lifestyle behaviors—our study thus offers complementary findings to a literature that has largely focused on mechanisms such as health insurance and medical expenses to explain wealth-health gaps.

Finally, our paper also relates to the voluminous empirical literature studying the relationship between socio-economic status and health. A survey and summary of the main empirical findings of this literature is provided in Cutler et al. (2011). We contribute to this body of work by providing an update on the state of health-related inequalities in Germany. In doing so, we complement other studies using this same data set, such as Lampert et al. (2018), who employ the latter to compare the socioeconomic-health gradient in Germany to other countries and across time.

The remainder of the paper is organized as follows. Section 1.2 sets forth a number of empirical observations related to wealth, health, and lifestyle behaviors that guide the

[5] See Appendix 1.A for a detailed discussion of the German healthcare system.

[6] The German healthcare system is also characterized by the highest per capita spending among EU countries and some of the highest rates of available beds, doctors, and nurses per population.

development of our structural model. We then present the model economy in Section 1.3 and describe its estimation in Section 1.4. Section 1.5 provides and discusses the main quantitative results. Section 3.4 concludes.

1.2 Empirical Observations on Health, Lifestyle Behaviors, and Wealth in Germany

Throughout this paper, we rely on data from the German Socio-Economic Panel (SOEP). The SOEP is an annual representative longitudinal panel study of private households, conducted by the German Institute for Economic Research, DIW Berlin. We use information from the 2004-2018 survey waves. We convert nominal variables into 2015 Euros using a CPI index for inflation adjustment.

1.2.1 Health and Lifestyle Behaviors

Health Status

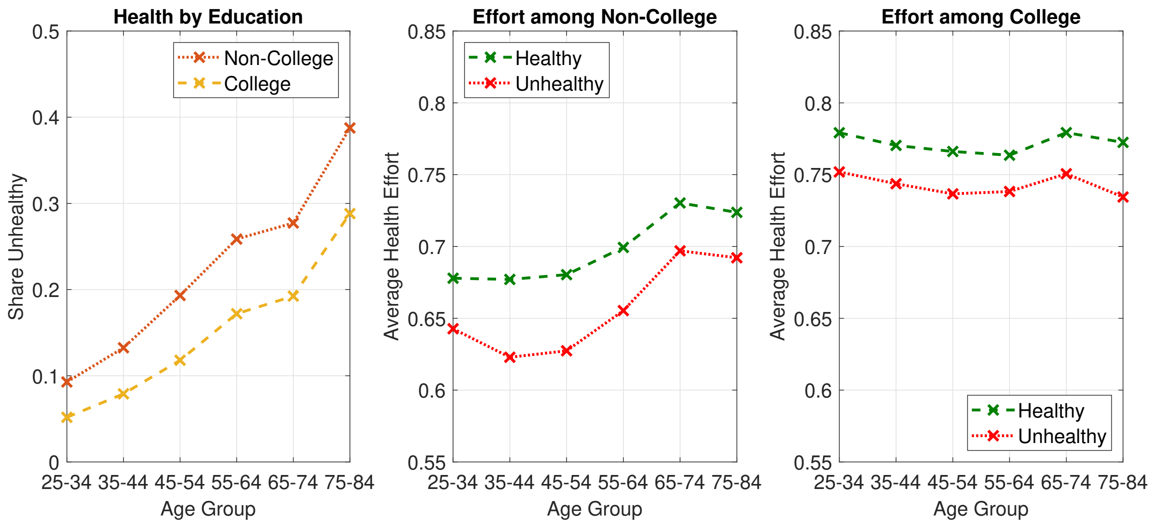
We measure individual health using information on self-rated health status in the SOEP.[7] In every survey wave, respondents are asked "How would you rank your current health?" to which respondents can answer *Very good, good, satisfactory, less well, or poor*. Consistent with much of the literature (De Nardi et al., 2023; French, 2005), we combine the first three categories into one *healthy* category and the last two into one *unhealthy* category.[8]

The left panel of Figure 1.1 shows the average share of unhealthy individuals by 10-year age groups, starting at ages 25-34 and ending with ages 75-84. We also distinguish between individuals according to their education level, where those in the college group have obtained a college degree, and those in the non-college group have not. Already at ages 25-34, members of the non-college education group are around 2 percentage points more likely to be unhealthy than the college-educated. This gap grows over the life course. At ages 75 and older, around

[7] In select survey waves, the SOEP also contains more objective health measures, such as a series of concrete diagnoses. We use this information to construct an index of *frailty*, similar to that in Hosseini et al. (2022), by adding one to the index each time an individual is diagnosed with a specific health condition. Moreover, since 2002, the SOEP includes questions that allow to construct generic indicators of perceived physical and mental health, called Physical and Mental Component Summary scores (PCS and MCS, respectively). In Appendix 1.B, we check the correlation of our benchmark binary health measure and these two alternatives. We focus on the self-reported health status measure because this maximizes the amount of data available for our empirical analysis, given that most of the more detailed questions about health deficits only started to be asked in 2011.

[8] This procedure could mitigate potential issues related to measurement errors and also reduces computational burden when we estimate our quantitative model presented in Section 1.3.

Figure 1.1: Average Health and Health Effort over the Life Cycle by Education



Notes: Left: Share of unhealthy people in the SOEP over 10-year age groups, distinguishing between the non-college-education and college-education groups. Center and Right: Average health effort by 10-year age groups for non-college (center) and college-educated (right) individuals in the SOEP, distinguished between unhealthy status and healthy status.

40% of non-college educated individuals are in poor health compared to around 30% of the college-educated.

Lifestyle Behaviors

We measure *lifestyle behaviors* by individual *health efforts*—a composite measure of three individual behaviors for which we have information. These behaviors include: (i) the frequency of sport or physical exercise; (ii) health-conscious nutrition; and (iii) the daily number of cigarettes smoked. In Germany, as in most developed countries, physical inactivity, smoking, and poor diet are recognized as the most important contributors to individual health risk (OECD, 2019). We first standardize each component so that they have mean zero and standard deviation one (Kling et al., 2007). We then construct health effort as a weighted sum of these, which we normalize to be in the unit interval.[9] Overall, individual health effort observations have a mean of around 0.71 and a standard deviation of 0.16. Moreover, we observe substantial path dependence in health efforts. For example, the autocorrelation of health efforts in a two-year interval is high at 0.76.

Figure 1.1 compares the average health effort levels for the non-college (central panel) and

[9] The weights are taken from the relative loadings of each behavior on the first principal component of all behaviors, after stripping them of variation coming from observable characteristics. Details are explained in Appendix 1.C.

the college-educated (right panel), separately for unhealthy and healthy individuals. Three patterns are worth noting. First, the life-cycle patterns for each group are relatively flat.[10] Second, there are large and persistent differences in average health effort across education groups. College-educated individuals are characterized by health efforts that are, on average, around half a standard deviation higher than those non-college-educated individuals. Third, conditional on education, unhealthy individuals consistently exert less health effort on average, than healthy ones. Unhealthy individuals could experience physical and mental difficulties exerting efforts (contributing to a higher health gap). At the same time, they could also have a greater incentive to exert more efforts to recover health (Verdun, 2022). These two countervailing forces could explain the relatively small yet still significant observed differences across health status (around 1/4 of a standard deviation).

1.2.2 The Relationship between Health and Wealth

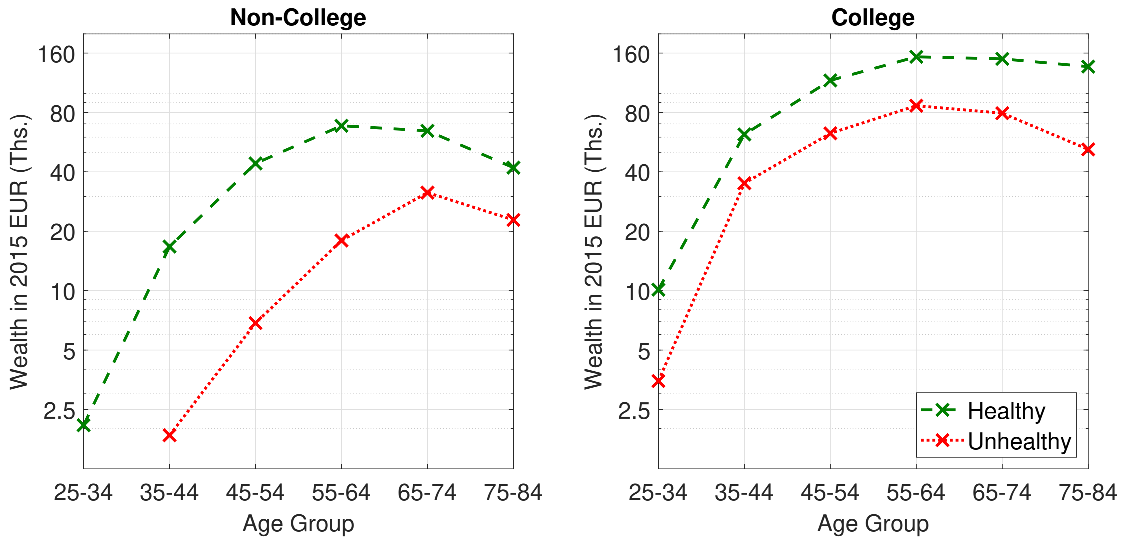
Germany is no different from many countries in the strong association we observe between financial well-being and health-related well-being. To illustrate, Figure 1.2 shows the evolution of median wealth over the life cycle, separately for healthy and unhealthy individuals in each education group (non-college and college). Wealth is measured as net worth, as is standard in the literature. It includes information on owner-occupied housing and other properties (net of mortgage debt), financial and business assets, tangible assets, private pensions (including life insurance) and consumer credits (Frick et al., 2007).[11] Wealth levels are plotted on a log/ratio scale, such that equal spaced points go up by a factor of 2.

For both education groups, the wealth levels of the healthy are consistently higher than those of the unhealthy. This *wealth-health gap* is already present early on in life. The percentage gap is generally higher among non-college educated individuals than among college educated ones. In both groups, the percentage gap is relative constant throughout the working years. It decreases slightly after retirement among the non-college educated whereas it increases slightly among the college educated. The existence of these significant wealth-health gaps in both education groups indicates that the association between wealth and health cannot be explained solely by education. Similar exercises can, in fact, be carried out using

[10] This does not preclude significant age-trends in lifestyle behaviors. For instance, while sport and exercise frequencies seem to decrease over age, healthy nutrition and abstention from smoking increase (see Figure 1.C.1).

[11] It does not include information on pension entitlements through both company pensions and the statutory German social pension fund as well as the pension entitlements for civil servants. Contrary to widely used surveys in other countries such as the Panel Study of Income Dynamics, the SOEP provides information on wealth at the individual level. This is achieved by asking the respondents for their personal share of ownership regarding each of the above components of wealth. In our analysis, we use an average of individual wealth across different imputation techniques.

Figure 1.2: Median Wealth Profiles of Healthy and Unhealthy Individuals by Education



Notes: Median wealth by 10-year age-groups and health status for non-college-educated (left panel) and college-educated (right panel) individuals in the SOEP, plotted on a log/ratio scale.

different dimensions of socioeconomic status. For instance, occupations could, through their potentially different toll on health, contribute to the wealth-health gap (see Figure 1.K.1). Yet, in all cases, an independent correlation between wealth and health seems to persist, suggesting the existence of other channels driving this relationship.

Perhaps the most natural channel of this type consists of the detrimental effect of poor health on an individual's ability to productively participate in the labor market. Indeed, a large empirical literature documents that health deficits significantly contribute to employment decline (Blundell et al., 2023b). Moreover, even when they are working, individuals in worse health tend to reduce their hours and are less productive, as reflected in their lower wages relative to healthy workers. Together, these factors contribute to the significantly lower labor incomes observed for unhealthy individuals.[12] Worse health thus leads to lower available resources to accumulate wealth over the life cycle.

Yet, as pointed out by Poterba et al. (2017) and De Nardi et al. (2023), a simple accumulation of lost labor income due to poor health over the lifetime does not explain the majority of the association between health and wealth.[13] In light of these results, we explore

[12]Relatedly, Hosseini et al. (2021) decompose the channels through which worse health leads to reduced labor income in the U.S. They find that the most important driver behind declines in earnings is exit from employment. In Appendix 1.D, we investigate the effect of health on labor income in the SOEP data using an instrumental variables approach. Our results indicate that being healthy increases the probability of being employed by an estimated 10.8%, even conditional on employment in the past two periods. Moreover, when working, good health increases labor income by around 10%. The majority of this increase is due to longer working hours, which increase by over 6%, while the rest is explained by higher wages.

[13]In their findings for the U.S., Poterba et al. (2017) argue that between 20 and 40% of the asset costs of poor

the importance of individual health behaviors as an additional mechanism underlying the wealth-health relationship.[14]

Health Efforts and the Wealth-Health Relationship

Given that an individual's health outcomes benefit from better health behaviors (Darden et al., 2018; Kvasnicka et al., 2018), variations in that latter could in part explain the considerable wealth-health gap observed in the data. Moreover, economic theory suggests that, in a world where survival is endogenous and can be influenced by healthy lifestyles or investments into health, the return to such efforts should increase in wealth, as richer people gain relatively more from prolonging their life.[15]

In line with this, Figure 1.3 illustrates that, indeed, healthy behaviors increase with wealth in the SOEP data. The figure displays the average level of our constructed health effort measure across wealth quartiles, conditional on education and age group. Health effort consistently rises in wealth. The increase is especially pronounced for non-college-educated 45-64-year-olds, where average effort increases by almost one standard deviation when going from the bottom to the top wealth quartile.

These effort differences by wealth might be driven by the fact that richer people can simply afford more or higher quality health investments thanks to their greater financial resources. We argue, however, that this is not the case here since our health effort measure contains variables that are mostly behaviorally driven. Moreover, in the case of abstention from smoking, higher health effort actually requires lower financial expenditure.

To further investigate the role of health-related behaviors in influencing the wealth-health relationship net of potentially confounding factors, we estimate the following equation:[16]

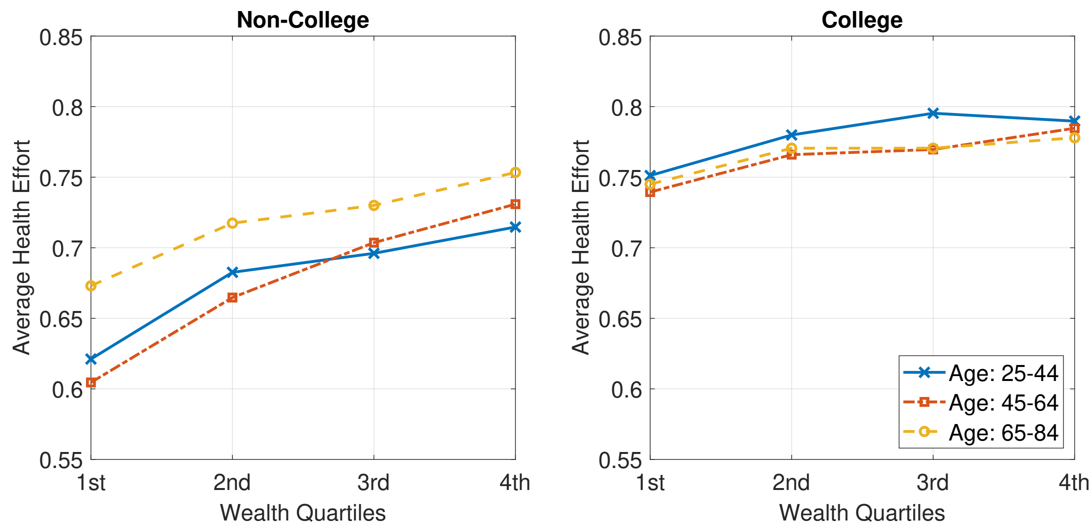
health are attributable to lower income and annuity income. We find similar effects in our quantitative results. De Nardi et al. (2023) estimate that even adding out-of-pocket medical expenses does not close the wealth-health gap.

[14] A number of other influences of wealth on health have been investigated in the literature, including the direct effects of material resources on health, such as living conditions, the affordability of better health care, or certain psychological effects that can translate into better physical health. These studies draw mixed conclusions, see for example Cesarini et al. (2016); Schwandt (2018), and a survey in O'Donnell et al. (2015).

[15] We illustrate this argument in a very simple model even without monetary investments in Section 1.5.2. The idea is that, when survival is endogenous, what matters for inter-temporal decisions is not just the marginal utility of consumption, but the *levels* of utility, which increase in wealth. Similar theories that typically include monetary investments into health (i.e. where health can be "bought") have been set forth in several seminal papers, such as Rosen (1988), Becker (2007), and Hall and Jones (2007), where they serve as the main explanation for the rising share in healthcare spending in the U.S.

[16] We note that we do not intend to estimate causal effects of wealth on health from this regression. Instead, the purpose of this exercise is to illustrate how dynamic correlations between current wealth and future health are affected by the presence of health efforts, which can play a role of a mediating force behind such dynamic relationships. In fact, it is the difficulty of estimating causal effects of wealth on health in a

Figure 1.3: Mean Health Effort by Wealth Quartiles



Notes: Average health effort by age group and wealth quartiles for non-college-educated (left panel) and college-educated (right panel) individuals in the SOEP data.

$$Health_{i,t+k} = \beta_1 Wealth_{i,t} + (\beta_2 Effort_{i,t}) + \gamma \mathbf{X_{i,t}} + u_{i,t}, \tag{1.1}$$

where $\mathbf{X_{i,t}}$ includes a constant, age, age^2, years of schooling, labor income, hours worked, lagged health, gender, marital status, labor force status, type of health insurance (private or public), year dummies, number of children in the households, as well as a measure of individual patience.[17] Row (1) in Table 1.1 reports the estimated coefficients $\hat{\beta}_1$ of wealth on health in the current year t and in a future year $t+k$ for $k = 1, 2, 3$. The coefficients confirm a persistent positive correlation between wealth and current and future health, net of other confounding influences.

Row (2) reports the estimated coefficients on wealth, while *including current health effort* as an additional regressor. The estimated coefficients on wealth, $\hat{\beta}_1$, decrease by 6-8% across all horizons of health. That is, a non-negligible share of the estimated effect of wealth on current and future health can be explained by variations in health effort. This suggests that health effort can mediate the positive relationship between wealth and health. At the same time, the estimated coefficients on health effort, $\hat{\beta}_2$ are all positive and increase with the horizon of health, indicating that our measure of health effort captures aspects of lifestyle

reduced-form way that, amongst other things, motivates our structural analysis in the following sections.

[17] We include patience in an attempt to control for unobserved discount factor heterogeneity that could be correlated with individual health evolution and health behaviors but also wealth. Due to the fact that detailed wealth information is only available every 5 years in the SOEP, we cannot directly estimate a version of (1.1) that includes individual fixed effects. Section 1.4 details how we measure patience from the data, as our quantitative model also features discount factor heterogeneity.

Table 1.1: Effect of Wealth on Current and Future Health, with and without Effort

		Effect on $Health_{i,t+k}$			
		(i)	(ii)	(iii)	(iv)
		$k = 0$	$k = 1$	$k = 2$	$k = 3$
(1)	$Wealth_{i,t}$	0.106	0.111	0.134	0.139
		(0.035)	(0.031)	(0.043)	(0.048)
(2)	$Wealth_{i,t}$	0.100	0.103	0.124	0.128
		(0.033)	(0.036)	(0.039)	(0.044)
	$Effort_{i,t}$	0.103	0.148	0.170	0.192
		(0.015)	(0.015)	(0.016)	(0.017)
(1) R^2		0.299	0.253	0.234	0.215
(2) R^2		0.301	0.256	0.238	0.219

Notes: Estimated coefficient $\hat{\beta}_1$ from equation (1.1) in row (1), and $\hat{\beta}_1$ and $\hat{\beta}_2$ in row (2). Columns (i) - (iv) correspond to separate regressions for $k = 0, 1, 2, 3$. Numbers in parentheses are heteroskedasticity-robust standard errors. The estimated coefficients and standard errors of wealth are multiplied by 10^7. $N = 24,928$.

behaviors that positively affect the probability of being healthy, and that these effects take time to materialize.

The empirical observations presented in this section paint a clear picture. There exists a strong association between individual health and financial resources in Germany. These wealth-health gaps grow substantially in absolute terms over the working career and persist even after controlling for obvious potential confounding factors, such as education and occupation. We provide suggestive evidence that variations in individual lifestyle behaviors play an important role in explaining these gaps. Over time, positive wealth gradients in efforts could translate into better health outcomes, which in turn are associated with higher earnings.

The dynamic nature and mutual dependencies of these effects make empirically assessing the relative importance of the different mechanisms underlying the wealth-health relationship particularly challenging without a structural framework. In the following sections, we therefore construct and estimate a model around the joint evolution of wealth and health of heterogeneous agents over the life cycle that allows us to disentangle the contribution of the different channels.

1.3 Model

1.3.1 Demographics

Agents enter the model at the beginning of their working career at age $j = 1$ and live at most for J periods. A period corresponds to two years. They decide how much to work for every period until age j_R, when they retire and consume out of their savings and pension benefits.

Agents are ex-ante heterogeneous along several dimensions. First, education status e can either be high ($e = 1$), corresponding to college education, or low ($e = 0$), corresponding to no college education. Second, agents also differ in their fixed discount factor β. Moreover, we allow agents to be different in their productivity type θ, which affects life-cycle wage offers (Storesletten et al., 2004). Finally, agents differ in their fixed health type η, which influences the health transition over the life cycle. We think of these health types as primarily capturing heterogeneity in health evolution that stems from factors that occur before agents enter the model (such as child and adolescent health and lifestyles or family environment and upbringing) or innate and genetic heterogeneity.[18]

1.3.2 Health and Lifestyle Behaviors

At every age j, agents can be either healthy ($h_j = 1$) or unhealthy ($h_j = 0$). Being unhealthy affects economic outcomes in several ways. First, it decreases the survival probability from age j to $j + 1$, denoted by $S_j(h_j, e)$, which also depends on age and education. Second, it results in productivity loss when working, which manifests in a constant education-specific productivity penalty. Third, poor health affects the disutility incurred from working and the marginal utility derived from consumption. Finally, it also affects the utility costs associated with maintaining a healthy lifestyle.

We view lifestyles as being the result of health effort choices $f_j \in [0, 1]$. Analogous to the definition in Section 1.2, we think of this level as a compound measure of all the efforts an individual makes to lead a healthy lifestyle. Agents enter every period j with a health effort level f_{j-1}, chosen in the past period. They then decide whether to change their health effort level from f_{j-1} or not. This decision is subject to a stochastic adjustment cost drawn from an age-dependent uniform distribution $\chi_j \in \mathcal{U}[0, B_j] \equiv H_j(\chi)$, which has to be paid if the agent decides to change her effort level relative to her previous level f_{j-1}.[19] The inclusion

[18] In their analysis of the joint wealth and health distribution in the U.S., De Nardi et al. (2023) find that inherent differences in time preferences across fixed health types are a substantial driving force of the observed wealth-health gradient. As detailed in Section 1.4.2, we also allow the initial conditions to be correlated with each other, in line with the data.

[19] Stochastic adjustment costs are widely used in different contexts such as firm investment and price ad-

of such a cost is motivated by the fact that a relatively high number of people in the data do not adjust their health efforts over time. Intuitively, this captures the idea of habits in health-related lifestyle behaviors.

Aside from a discrete decision on adjustment, we maintain the assumption that exerting health effort f_j comes at a direct contemporaneous utility cost, as in Cole et al. (2019). This utility cost $\varphi_j(f_j; h_j, e)$ is allowed to differ by age, health status and education. The dependence on education could capture any advantages more educated people have when exerting efforts, such as better neighborhoods or social networks, which could mitigate disutility of exerting healthy behaviors (Cutler and Lleras-Muney, 2010).

The benefit of leading a healthy lifestyle is that the latter increases the probability of being healthy in $j+1$, denoted by $\pi(h_{j+1} = 1|h_j, f_j, f_{j-1}, e, \eta)$. This probability firstly depends on the fixed health type η. Moreover, it depends not only on health efforts undertaken in period j, but also on those in the previous period. This assumption at least partially accommodates the fact that healthy lifestyles take time to materialize and may have health benefits that persist into the future (Cutler et al., 2011). Through its effect on health, higher health effort is then also associated with better survival prospects, given that survival probability increases in health. Finally, we let this probability be education-specific to allow for potential advantages in good health outcomes stemming from higher education net of its effect on efforts, for example through better living conditions.[20]

1.3.3 Preferences

Agents derive utility from consumption c and disutility from hours worked n. We assume that n can take a value from $\{0, n_p, n_f\}$, allowing for adjustments along both extensive and intensive margins. Working in age j implies a utility cost $\phi_j(n_j; h_j, e)$ that decreases in current health and is age- and education-dependent. This captures the fact that continuing to work when unhealthy may be more inconvenient.

Moreover, we assume that health affects the utility of consumption, where the effect is governed by $\kappa(h_j)$. This takes a value of one if healthy and $\tilde{\kappa}$, which is less than one if unhealthy. We include this complementarity between health and consumption utility as, for the great majority of goods and services, there is evidence that individuals enjoy their

justment in order to generate behaviors that often feature inaction. See Khan and Thomas (2008) for an overview.

[20]Moreover, this dependence on education allows us also to capture effects that cannot be picked up by our health effort measure, because of the way we construct it. For example, these could be more regular preventive doctor visits of the better educated because of better knowledge or access to information that would not show up in our data.

consumption more when healthy.[21]

Under these assumptions, per-period utility then takes the following form:

$$u(c_j, n_j, f_j; h_j, e) = \kappa(h_j)\left(\frac{c_j^{1-\sigma}}{1-\sigma} + b\right) - \phi_j(n_j; h_j, e) - \varphi_j(f_j; h_j, e), \tag{1.2}$$

where σ denotes the inverse of the elasticity of intertemporal substitution and b is a utility constant that is added to ensure that the value of being alive is always greater than the value of being dead (Hall and Jones, 2007). We let this utility constant be also dependent on health through $\kappa(h_j)$. Without this, the utility level would shift up for the unhealthy with an empirically reasonable value of $\sigma > 1$, which could result in higher utility of life for the unhealthy relative to the healthy.

The addition of this constant b has implications for the levels of future utility. Since survival is endogenous and can be influenced by health effort, the future utility levels play a role in shifting individual effort choices. This is in contrast to standard dynamic problems, where agents only care about marginal utility in each given period of life. The dependence on future utility levels through endogenous survival therefore incentivizes richer individuals (who can expect to have higher future utility levels through a longer life length) to increase their health efforts (Becker, 2007). This is because the return to health effort, namely the ability to enjoy a longer and healthier life, increases with wealth—one of the reasons why we expect our model to generate a wealth gradient in health efforts, as in the data. We explore this mechanism both theoretically and quantitatively in Section 1.5.2.

1.3.4 Earnings, Taxes and Transfers

When working in age j, agents receive gross labor income equal to $w_j(h_j, e, \theta, z_j)n_j$. The wage offer $w_j(h_j, e, \theta, z_j)$ consists of a deterministic component $\lambda_j(h_j, e)$ that depends on health h_j and education e as well as the fixed productivity type θ and persistent idiosyncratic productivity risk z_j:

$$w_j(h_j, e, \theta, z_j) = exp(\lambda_j(h_j, e) + \theta + z_j) \tag{1.3}$$

We include the fixed effects θ to allow for the possibility that factors beyond education, age, and health can shift wage profiles (Low and Pistaferri, 2015).

[21] For example, Finkelstein et al. (2013), using data from the U.S. Health and Retirement Survey, observe a decline in marginal utility of consumption when health deteriorates; medical goods and services, such as nursing care, being the exception. Similarly, Blundell et al. (2023a) find that the resulting consumption drop of non-durable goods after an adverse health shock comes mainly from a change in the utility of consuming them rather than from the effect of health on resources.

We incorporate progressive labor income taxation captured by $\mathcal{T}(y_j, \bar{y})$ (Heathcote et al., 2017), where y_j denotes gross labor income and $\bar{y}$ refers to its average in the economy. In addition, agents are provided with transfers $T(c_j, h_j, n_j)$ that incorporate two types of welfare programs. First, a minimum consumption $\tilde{c}$ is guaranteed by the government to every individual, so that $T(c_j, h_j, n_j)$ includes $\tilde{c} - c_j$ if $c_j < \tilde{c}$. This could capture various means-tested social safety programs in Germany that are especially relevant to those with zero labor income, in particular Germany's basic social security provisions. We also incorporate a state-contingent transfer to capture sickness benefits, which would provide insurance against adverse health shocks. Specifically, $T(c_j, h_j, n_j)$ includes $\widetilde{T} > 0$ if an agent is unhealthy ($h_j = 0$) and does not work ($n_j = 0$).[22] Finally, the government provides pension benefits $P(e)$, which are paid out in retirement periods.

1.3.5 Individual Optimization Problems

We first describe the individual optimization problem of a *working-age agent* ($j < j_R$). At the beginning of each period j, the agent learns about her current health realization h_j and productivity draw z_j. At this point, the state variables are composed of a vector given by $\mathbf{s}_j = (e, \beta, \theta, \eta, a_j, h_j, z_j)$. Given $(\mathbf{s}_j, f_{j-1})$, the value function at the beginning of age j is then given by:

$$V_j(\mathbf{s}_j, f_{j-1}) = \mathbb{E}_{\chi_j} M_j(\mathbf{s}_j, f_{j-1}, \chi_j), \tag{1.4}$$

where M_j denotes the interim value after the stochastic effort adjustment cost draw χ_j is realized. This is given by:

$$M_j(\mathbf{s}_j, f_{j-1}, \chi_j) = \max \left\{ \underbrace{W_j^{adj}(\mathbf{s}_j, f_{j-1}, \chi_j)}_{\text{value of } \textit{adjusting effort}}, \underbrace{W_j^{not}(\mathbf{s}_j, f_{j-1})}_{\text{value of } \textit{not adjusting effort}} \right\}. \tag{1.5}$$

Here, W_j^{adj} is the value of adjusting health effort relative to its level in the previous period, which is given by:

$$W_j^{adj}(\mathbf{s}_j, f_{j-1}, \chi_j) = \max_{\substack{c_j, a_{j+1} \geq 0 \\ f_j \in [0,1], n_j \in \{0, n_p, n_f\}}} \left\{ \begin{array}{c} u(c_j, n_j, f_j; h_j, e) - \chi_j \\ +\beta S_j(h_j, e) \mathbb{E}_{h_{j+1}, z_{j+1} | \Omega_j} V_{j+1}(\mathbf{s}_{j+1}, f_j) \end{array} \right\}, \tag{1.6}$$

[22] In Germany, an integral part of the health insurance system consists of sickness benefits provisions that are paid to insured people in case they become incapable of working due to sickness (disability).

subject to

$$
\begin{aligned}
c_j + a_{j+1} &\leq a_j(1+r) + T(c_j, h_j, n_j) + w_j(h_j, e, \theta, z_j)n_j - \mathcal{T}(w_j(h_j, e, \theta, z_j)n_j, \bar{y}) \\
h_{j+1} = 1 & \quad \text{with prob. } \pi_j(h_{j+1} = 1 | h_j, f_j, f_{j-1}, e, \eta) \\
= 0 & \quad \text{with prob. } 1 - \pi_j(h_{j+1} = 1 | h_j, f_j, f_{j-1}, e, \eta).
\end{aligned}
$$

That is, the adjustment cost χ_j must only be paid when an agent decides to change her health effort relative to her previous level. Ω_j refers to the relevant subset of the state variables in period j used for taking conditional expectations.

Finally, W_j^{not} is the value of not adjusting health effort:

$$
W_j^{not}(\mathbf{s}_j, f_{j-1}) = \max_{\substack{c_j, a_{j+1} \geq \underline{a}_j \\ n_j \in \{0, n_p, n_f\}}} \left\{ \begin{array}{c} u(c_j, n_j, f_{j-1}; h_j, e) \\ +\beta S_j(h_j, e) \mathbb{E}_{h_{j+1}, z_{j+1} | \Omega_j} V_{j+1}(\mathbf{s}_{j+1}, f_{j-1}) \end{array} \right\}, \tag{1.7}
$$

subject to

$$
\begin{aligned}
c_j + a_{j+1} &\leq a_j(1+r) + T(c_j, h_j, n_j) + w_j(h_j, e, \theta, z_j)n_j - \mathcal{T}(w_j(h_j, e, \theta, z_j)n_j, \bar{y}) \\
h_{j+1} = 1 & \quad \text{with prob. } \pi_j(h_{j+1} = 1 | h_j, f_{j-1}, f_{j-1}, e, \eta) \\
= 0 & \quad \text{with prob. } 1 - \pi_j(h_{j+1} = 1 | h_j, f_{j-1}, f_{j-1}, e, \eta).
\end{aligned}
$$

During retirement periods ($j \geq j_R$), the optimization problem reduces to a standard consumption-savings problem in combination with choosing whether or not to adjust health effort and, in the affirmative, to which level. Thus, the interim value function (1.5) becomes:

$$
M_j(\mathbf{s}_j, f_{j-1}, \chi_j) = \max \left\{ \underbrace{R_j^{adj}(\mathbf{s}_j, f_{j-1}, \chi_j)}_{\text{value of } \textit{adjusting}}, \underbrace{R_j^{not}(\mathbf{s}_j, f_{j-1})}_{\text{value of } \textit{not adjusting}} \right\} \tag{1.8}
$$

where the values of adjusting effort, R_j^{adj}, and not adjusting effort, R_j^{not}, during retirement are now defined as

$$
R_j^{adj}(\mathbf{s}_j, f_{j-1}, \chi_j) = \max_{\substack{c_j, a_{j+1} \geq 0 \\ f_j \in [0,1]}} \left\{ \begin{array}{c} u(c_j, 0, f_j; h_j, e) - \chi_j \\ +\beta S_j(h_j) \mathbb{E}_{h_{j+1} | \Omega_j} V_{j+1}(\mathbf{s}_{j+1}, f_j) \end{array} \right\}, \tag{1.9}
$$

$$R_j^{not}(\mathbf{s}_j, f_{j-1}) = \max_{c_j, a_{j+1} \geq 0} \left\{ \begin{array}{c} u(c_j, 0, f_{j-1}; h_j, e) \\ +\beta S_j(h_j) \mathbb{E}_{h_{j+1}|\Omega_j} V_{j+1}(\mathbf{s}_{j+1}, f_{j-1}) \end{array} \right\}, \tag{1.10}$$

subject to the constraints

$$\begin{aligned} c_j + a_{j+1} &\leq a_j(1+r) + P(e) \\ h_{j+1} &= 1 \quad \text{with prob. } \pi_j(h_{j+1} = 1 | h_j, f_j, f_{j-1}, e, \eta) \\ &= 0 \quad \text{with prob. } 1 - \pi_j(h_{j+1} = 1 | h_j, f_j, f_{j-1}, e, \eta) \end{aligned}$$

Thus, during retirement, expectations are only made over future health realizations.

1.4 Estimation

1.4.1 Estimation Strategy

For the estimation of our model, we adopt a two-step strategy. In the first step, a set of parameters are set or estimated externally without using our model. Some of these, in particular the survival probabilities and the parameters governing the health transition probabilities are estimated directly from the SOEP data (waves 2004–2018). For the others, we set their values in line with the literature.

In the second step, we estimate the remaining set comprising 42 parameters using a moment matching estimator that minimizes the distance between model-implied moments and the corresponding empirical moments, taking as given the parameter values determined in the first step. Most importantly, we require the model to match the joint distribution of earnings and labor supply by age, health and education as well as the joint distribution of health efforts by age, health and education.[23] This results in 64 target empirical moments that are estimated from the data or taken from other sources, and summarized together with the parameters in Table 1.2.[24]

Formally, let Θ_0 be a vector of the 42 parameters to be estimated and $\hat{\Delta}$ be a vector of the 64 empirical moments that we want to match. Our structural model provides a mapping from a set of parameters Θ to the model-implied moments, denoted by a function $h(\Theta)$. The

[23] We do not explicitly target the joint distribution of wealth and health over age. This is because one of our key quantitative exercises is to investigate how much of the observed positive wealth-health association can be generated through the forces present in our model.

[24] Table 1.E.1 in the Appendix provides the full list of target statistics.

method of simulated moments estimator of Θ_0 is then given by

$$\hat{\Theta} = \arg\min_{\Theta}(\hat{\Delta} - h(\Theta))' W(\hat{\Delta} - h(\Theta)), \tag{1.11}$$

where W is a 64-by-64 weighting matrix. The standard errors of each individual component of $\hat{\Delta}$ (i.e., $\hat{\delta}_1, ..., \hat{\delta}_{64}$) are estimable in our case, although the full variance-covariance matrix of $\hat{\Delta}$ is unknown. For that reason, we follow the algorithm proposed by Cocci and Plagborg-Møller (2021) to estimate the standard errors of our estimates $\hat{\Theta}$. Their strategy first obtains the *worst-case* standard errors by assuming that all elements of $\hat{\Delta}$ are perfectly correlated with each other, which bounds the variance of any linear combination of its elements and therefore the variance of the estimator $\hat{\Theta}$. They then show that one can use an efficient selection of moments for every parameter that minimizes the worst-case estimator variance when the model is over-identified. We describe the algorithm to compute the standard errors and justify its use in detail in Appendix 1.E. The resulting estimates and standard errors are reported in the second and third columns of Table 1.2.

1.4.2 Model Parameters

As is well known for the application of the method of simulated moments, some moments are more informative for particular parameters although there is no one-to-one mapping between them. We now explain these links intuitively along with the description of the parameters belonging to the first step.

Demographics

We estimate the model at a biannual frequency so as to align with the frequency of health effort variables in our micro data. The first model period ($j = 1$) corresponds to age 25, so that agents enter the model after having obtained an education level. We assume that agents live at most until age 99, so that $J = 38$ with a model period of two years. Retirement age is set at 65 ($j_R = 21$).

Preference: Consumption/Saving and Labor Supply

We set the inverse of the elasticity of intertemporal substitution to $\sigma = 2$, a commonly-used value in the literature. The effect of poor health on the marginal utility of consumption, $\tilde{\kappa}$, is estimated internally to match the consumption differences between healthy and unhealthy 25-64 year-olds in the data (1.16). Note that in the model, a certain degree of consumption differences across health types is also endogenously generated. We estimate $\tilde{\kappa} = 0.872$, which implies a 13% loss for the unhealthy.

Next, we specify the disutility of working $\phi_j(n_j; h_j, e)$ as a combination of an age-, education-, and health-dependent shifter and a standard constant-Frisch-elasticity function:

$$\phi_j(n_j; h_j, e) = \nu_j^{h_j} \exp(\nu_e \mathbb{I}\{e = 0\}) \frac{n_j^{1+1/\gamma}}{1 + 1/\gamma}. \tag{1.12}$$

Thus, the labor supply disutility shifter is a combination of age- and health-specific coefficients—$\nu_j^{h_j}$—and ν_e, which determines extra disutility for those with a lower education level. Several labor supply patterns in the data motivate our parametric assumptions. As shown in the left panel of Figure 1.5, employment rates over age are hump-shaped with substantial gaps across health status. Moreover, there is a robust gap in employment rates between the education groups, as shown in the right panel of Figure 1.5. We estimate the above parameters internally to match two sets of moments that capture these patterns. These are the average employment shares among the healthy and unhealthy, by the age groups 25-34, 35-44, 45-54, and 55-64 and the average ratio of the employment rate of the college-educated to that of non-college educated (1.24). Given these nine target moments, we estimate nine parameters—ν_j^h for $j \in \{1, 8, 13, 20\}$ and $h \in \{0, 1\}$ as well as ν_e—while interpolating ν_j^h using piece-wise cubic splines for each h to obtain its value for all j.

The parameter γ is the Frisch elasticity of both intensive and extensive labor supply and is set to $\gamma = 1$, as is standard in the literature. We set $n_p = 0.5$, $n_f = 1$, and $\bar{n} = 3$ so that full-time work is one third of the total time endowment.

Preference: Lifestyle Behaviors

Health effort is a key and novel endogenous variable in our model. Its dynamics at the individual level are influenced by two kinds of utility costs in the model. Our aim is to parameterize such costs parsimoniously while being empirically consistent with the effort evolution across agents and over age.

We first specify the contemporaneous disutility incurring from exerting health effort level f_j as a combination of age-, education-, and health-dependent effort cost shifters, and a power function that increases with efforts, with the curvature parameter ψ shaping the degree of responsiveness in efforts:

$$\varphi_j(f_j; h_j, e) = \iota_j^{h_j, e} \frac{f_j^{1+1/\psi}}{1 + 1/\psi}. \tag{1.13}$$

To reproduce the education and health gradients in efforts presented in Figure 1.1 in

Section 1.2.1, we adopt age-specific coefficients $\iota_j^{h_j,e}$ for each health status h_j and education e. These empirical patterns are well summarized in the target moments, which consist of the mean health effort observed in the data by the age groups 25-34, 35-44, 45-54, 55-64, 65-74 and 75-84, separately for each health status and education. To match these 24 moments, we estimate 16 parameters—$\iota_j^{h_j,e}$ for $j \in \{1, 12, 20, 31\}$, $h \in \{0, 1\}$ and $e \in \{0, 1\}$—while interpolating $\iota_j^{h_j,e}$ using piece-wise cubic splines for each h and e. Next, we internally estimate the curvature parameter $\psi = 1.115$ to match the empirical dispersion of efforts (standard deviation of 0.16).

The other kind of the utility cost concerns the distribution of the stochastic effort adjustment costs. This dynamic adjustment cost is crucial in governing the proportion of agents who choose not to adjust their efforts. In the data, this share increases with age, as reported in Table 1.2. To replicate this pattern, we parameterize the age-dependent upper bound of $\mathcal{U}[0, B_j]$ as

$$B_j = \varsigma_0 \exp(\varsigma_1(j-1)). \tag{1.14}$$

and estimate the two parameters—ς_0 and ς_1—to match the share of individuals not adjusting efforts for three age groups: 25-44, 45-64, and 65-84.

Next, we internally estimate the utility constant to $b = 13.1$, such that the model-implied value of a statistical life year (VSLY) is equal to 8.49 times average annual per capita consumption. The VSLY describes the average utility-equivalent value that individuals in our model would attach to one extra year of life. In quantitative models with endogenous survival, the VSLY can be defined by equalizing the average flow utility of a life year across individuals with average consumption, multiplied by average marginal utility of consumption so as to transform this into utility units, as in Glover et al. (2023):[25]

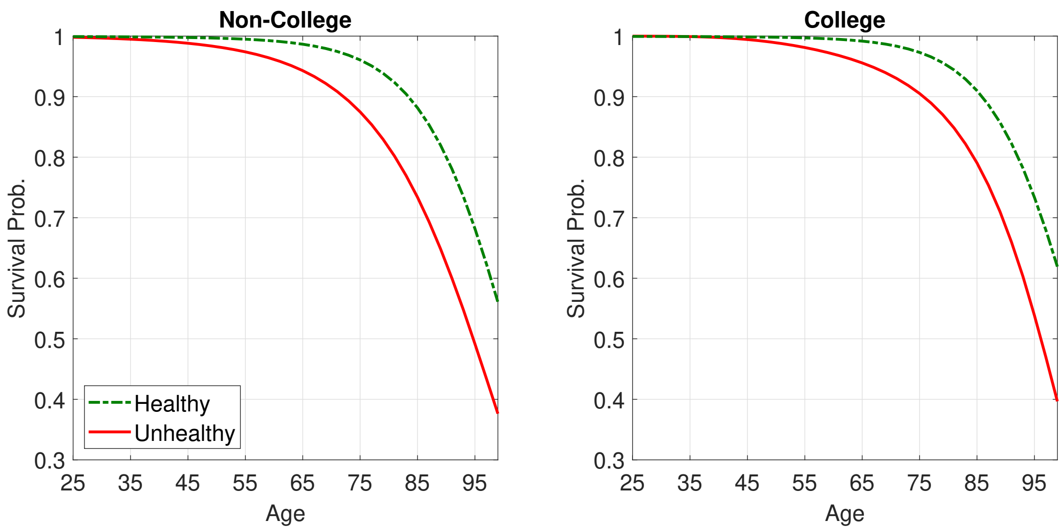
$$\bar{u}(c_j, n_j, f_j; h_j, e) + b = \frac{\partial \bar{u}}{\partial c} \times \underbrace{8.49\bar{c}}_{\text{VSLY}}. \tag{1.15}$$

We take the empirical target for the VSLY from a meta-analysis of value of a statistical life estimates in OECD (2012), who report a value of around 4.7 million 2005-USD among a sample of EU countries.[26] We transform this value into a VSLY of around 140 thousand 2018-EUR using the average age (44.4 years) and average life expectancy at that age (34.8

[25] Since our model frequency is two life years, we are technically comparing the value of two extra life years to average consumption over two years when estimating b. Thus, we can still use the ratio of 8.49 as our target statistic.

[26] The estimates are obtained from surveys, where participants are asked about their willingness to pay for small reduction in mortality risks. The results are in Table 6.1 in OECD (2012). In comparison to other estimates in the literature (such as Glover et al. (2023)), this is a rather conservative estimate.

Figure 1.4: Estimated Conditional Survival Probabilities by Education and Health



Notes: Probability of survival for two years conditional on being alive at a given age, coming from a probit model of survival on a cubic polynomial in age estimated on SOEP data. Survival probabilities are estimated separately for non-college and college individuals and by health status.

years) in Germany in 2018 and under the assumption of a 1% annual discount factor (Glover et al., 2023).

Survival Probability

We estimate the two-year survival rates $S_j(h_j, e)$ directly from the data using information on deaths of survey respondents contained in the SOEP. Specifically, we fit a probit model of survival up to age $j+1$ on a cubic polynomial in age by health status at age j and education. The resulting estimated conditional two-year survival probabilities are plotted in Figure 1.4.[27] Conditional on being alive at a given age, healthy people are more likely to survive the next two years than unhealthy people. This difference increases with age. Moreover, at all ages, higher educated people have higher chances of survival than lower educated people although the differences driven by education are relatively small (Pijoan-Mas and Ríos-Rull, 2014).

Health Evolution and Fixed Health Types

The probability of being healthy in the next period is a function of an individual's age, education, current health, and past and present health efforts. On top of that, we allow the health evolution to depend on unobserved fixed type, which we consider arising primarily

[27] To check that the estimated survival rates are reasonable and do not suffer from a lack of tracking the reasons respondents exited the SOEP survey, we compare the results in Figure 1.4 with the German Statistical Office's mortality risk tables. Doing so largely confirms our estimates.

from different initial conditions before the age of 25 when agents enter the model. As such, they can originate from inherent genetic predispositions but also from differences in family environments and lifestyles during childhood and adolescence. Given the inclusion of the unobserved heterogeneity, we employ a two-step group fixed effects estimator (Bonhomme et al., 2022) to estimate the health evolution process.

The first step in the estimation involves classifying individuals into a small number of discrete fixed health types η, based on the *kmeans* clustering algorithm. The goal is to group individuals together that are most similar in terms of a latent type which influences their health evolution net of observable characteristics. To that end we define a vector of individual-specific moments that are likely informative about an underlying latent health type. These moments include the number of doctor visits, self-rated health status (5-point scale), inpatient nights in a hospital, both physical and mental health summary scores (PCS and MCS) and the body mass index. Details on these moments, as well as the clustering procedure are given in Appendix 1.F. We run the classification repeatedly while increasing the number of clusters and randomizing the initial group centers. We then compare the total within-cluster sum of squares of each cluster solution to find a suitable number of clusters. We end up with two fixed health types $\eta \in \{0, 1\}$, where around 2/3 of individuals in our data are of the high health type $\eta = 1$ and the rest are of the low health type $\eta = 0$.[28]

In the second step, we estimate the probability of being healthy in the next period conditional on current and past health effort, education, current health and these fixed health type groups, $\pi_j(h_{j+1} = 1|h_j, f_j, f_{j-1}, e, \eta)$, directly from the data with the following logistic model:

$$\begin{aligned}
&\pi_j(h_{i,j+1} = 1|h_{i,j}, f_{i,j}, f_{i,j-1}, e_i, \eta_i) = \\
&\Big(1 + \exp\left(-(\pi^0_{i,j} + \lambda_1 f_{i,j} + \lambda_2 f_{i,j-1} + \delta h_{i,j} + \gamma_1 e_i + \gamma_2 \eta_i + \gamma_3 \mathbf{A}_i)\right)\Big)^{-1},
\end{aligned} \tag{1.16}$$

where $h_{i,j}$ is a dummy variable that equals 1 if person i is healthy at age j, $f_{i,j}$ is our compound health effort measure, e_i is a dummy variable equal to 1 if person i has college education, η_i is a dummy variable that equals 1 when individual i's health type is high, and $\mathbf{A}_i$ is a vector of dummies that are equal to 1 when individual i is a member of a 10-year age group.

We present the exact logistic estimates from (1.16) that we use in the model in Table 1.G.1 along with detailed discussions in Appendix 1.G. Notably, the estimated effects of current

[28] When comparing the total within-cluster sum of squares as a measure for cluster homogeneity, a kink appears most noticeably at two and three clusters. We opted for two health type groups, which offers a compromise between maintaining computational feasibility and accounting for a sufficient degree of heterogeneity.

and past health effort are positive and quantitatively meaningful. The estimates imply that, for example, a 75-year-old college-educated individual of the high health type can increase her probability of being healthy by almost 2% if she is currently healthy and increases just her contemporaneous health effort by one standard deviation above the average. If she is currently unhealthy this effort improvement will raise her probability of being healthy next period by over 7%. Moreover, by increasing effort for two consecutive periods to one standard deviation above the mean, the probability will be increased by 15% if she is currently unhealthy and over 3% if she is currently unhealthy. Generally speaking, past health effort is, on average, slightly more productive in increasing the healthy probability, which underlines the importance of considering the dependence of good health outcomes on a longer history of healthy lifestyles.

We gauge the empirical realism of our health transition parameter estimates in detail in Appendix 1.G and discuss their implications for disease prevalence and mortality in comparison to existing estimates in the medical literature. Relative to the latter, we conclude that our estimated effectiveness of past and present health effort in improving health outcomes is rather conservative.

Wage and Fixed Productivity Types

For estimation, we augment the wage equations (1.3) with the specification of the idiosyncratic risk z_j and statistical error terms:

$$\begin{aligned} \ln w_j &= \lambda_j(h_j, e) + \theta + z_j + \varepsilon_j \\ z_j &= \rho z_{j-1} + \upsilon_j, \end{aligned} \tag{1.17}$$

where $\theta \sim \mathcal{N}(0, \sigma_\theta^2)$, $\varepsilon_j \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, and $\upsilon_j \sim \mathcal{N}(0, \sigma_\upsilon^2)$. Thus, log wages are a combination of an observed, deterministic component $\lambda_j(h_j, e)$ that is dependent on education and health, as well as an idiosyncratic component that consists of unobserved fixed productivity heterogeneity θ and persistent shocks z_j.[29]

We estimate the deterministic component $\lambda_j(h_j, e)$ internally within our structural model to address selection bias that might arise due to the well-known issue that we do not observe wages for non-working individuals and it is likely that individuals select into employment based on their observable characteristics, including their health status (Low and Pistaferri,

[29] Although the wage equations (1.3) in the model do not include transitory shocks ε_j, the empirical equations do so in order to identify fixed productivity types θ (Storesletten et al., 2004). We abstract from correlations of fixed productivity or idiosyncratic shocks with observables, in particular health and education, as is common in the literature (e.g., Low and Pistaferri (2015)).

2015). Specifically, we parameterize it such that for each education group e,

$$\lambda_j(h_j, e) = \zeta_0^e \exp(\zeta_1^e(j-1) + \zeta_2^e(j-1)^2) \times (1 - w_p^e \mathbb{I}\{h_j = 0\}). \tag{1.18}$$

The coefficients ζ_0^e allow the two education groups to have a different intercept in their deterministic wage profile. The exponential term captures different trajectories of productivity over age by education group. The last term is a constant productivity penalty w_p^e that captures productivity losses due to poor health. In line with the literature (Hosseini et al., 2021), we allow these contemporaneous effects of poor health to differ by education, which might capture, for example, the fact that non-college workers are more likely to work in physically demanding jobs, where poor health might be more consequential in terms of productivity losses. We then estimate these eight parameters—ζ_0^e, ζ_1^e, ζ_2^e, and w_p^e for $e \in \{0, 1\}$—internally so that the model matches the mean (two-year) earnings by education and health status for the age groups 25-34, 35-44, 45-54, and 55-64 (16 moments in total) in the data.

Next, we estimate the distribution of fixed productivity types and the persistence of idiosyncratic shocks directly using individual-level wage data in the SOEP, using a standard procedure in the literature (De Nardi et al., 2023; French, 2005), as detailed in Appendix 1.F. This yields an estimated persistence of idiosyncratic productivity shocks of $\rho = 0.975$ and provides a distribution of empirical individual-specific productivity fixed effects estimates $\hat{\theta}_i$. To recover the fixed productivity types used in our model, we classify this distribution of $\hat{\theta}_i$ into two discrete types, similar to Low and Pistaferri (2015), corresponding to low productivity θ_l (the bottom 50%), and high productivity types θ_h (the top 50%).[30] We then set $\theta_l = -0.29$ and $\theta_h = 0.29$ symmetrically, such that the variance of the discrete types corresponds to the estimated variance $\sigma_\theta^2 = 0.084$. Given the estimates of the persistence of idiosyncratic shocks and the fixed productivity type distribution, the variance of the idiosyncratic productivity component σ_v^2 is estimated internally such that the model matches the observed variance of log earnings (0.59) in the data.

Initial Distribution

We construct the initial distribution of agents over the state space upon entry into the model directly from the data. We first describe the distribution over the fixed types. As before, education distinguishes college (31%) and non-college education (69%). As detailed above, the fixed productivity types are discretized into two equal-sized masses, and the fixed

[30] We also experimented with three discrete productivity types as in Low and Pistaferri (2015), which did not alter the results significantly. As in Low and Pistaferri (2015), we classify the individuals who never work in our sample and, hence, do not have an estimated productivity fixed effect into belonging to the low productivity type.

health types are estimated using the *kmeans* clustering algorithm, leading to 63% of the high health type ($\eta = 1$) and 37% of the low health type ($\eta = 0$). The remaining source of ex-ante heterogeneity in our model comes from differences in the discount factor β. We discretize the distribution of β into two equal-sized masses, β_l and β_h, using information about time preferences coming from an incentivized experiment conducted in the 2006-wave of the SOEP.[31] Since this information does not inform the levels of the discount factors in the model directly, we assume that $\beta_l = \mu_\beta - \delta_\beta$, and $\beta_h = \mu_\beta + \delta_\beta$, and estimate μ_β and δ_β internally, such that our simulated data matches the following seven relevant moments in the data: the median wealth for the age groups 25-34, 35-44, 45-54, 55-64, 65-74, and 75-84 (as shown in the left panel of Figure 1.8) and the Gini-index of wealth (0.746).

We require the joint distribution over education, unobserved health types, productivity types and discount factor types upon entry into the model to be the same as in the working-age population in the data.[32] This is important since the observed positive wealth-health association can be at least partly explained by the joint density of discount factor and other fixed types, in particular unobserved health types, as highlighted by De Nardi et al. (2023). In our sample, patience and fixed health types are indeed positively correlated (with the correlation being 0.1). Moreover, the health type is slightly positively correlated with the productivity type.

We also require the initial distribution to reflect differences in initial health and healthy lifestyles. Accounting for these initial differences is potentially important given the habitual nature of healthy lifestyles and path-dependence of health evolution, reflected in our estimated health technology (1.16). To that end, we use the conditional means of health and health effort at ages 25 to 30 as the initial states, where we condition on education and fixed health type. We report the resulting exogenous distribution across states including average health and health effort at the beginning of our model in Table 1.K.1. Finally, we assume that agents enter the model with zero wealth and set the real interest rate to $r = 0.082$, which corresponds to an annual rate of 4%.[33]

[31] Details on the experiment are given in Richter and Schupp (2014). The experiment consisted of the individual's decision whether to obtain money now or at a later point in time with increasing interest rates. From the implied interest rate each individual requires to be indifferent between the two options, we can extract information about their patience.

[32] In the data there remain small differences in the distributions over age, despite age typically being a control variable in the construction of the types. The only source that influences the distribution of fixed types over age in the model is endogenous survival. This may in particular be a concern, if agents of the low health type are more likely to exit the model due to death. However, given that exogenous survival rates during the working ages are very high (see Figure 1.4), we see this issue as negligible.

[33] In our data, we do not have sufficient information about wealth at age 25 and younger to justify a different assumption about initial wealth when entering the model.

Taxes and Transfers

We specify the progressive labor tax system using a commonly used parametric function (Heathcote et al., 2017):

$$\mathcal{T}(y_j, \bar{y}) = y_j - (1 - \tau_s) y_j^{1-\tau_p} \bar{y}^{\tau_p}. \tag{1.19}$$

In this formulation, τ_s captures the scale and τ_p captures the degree of progressivity of the tax system. $\bar{y}$ is the average income. In accordance with the estimates in Kindermann et al. (2020) for Germany, we set to $\tau_s = 0.321$ and $\tau_p = 0.128$.

In terms of pension benefits $P(e)$, we follow a similar approach as in Kindermann et al. (2020). We initially set these as equal to the earnings agents would have earned in the period prior to retirement if they had worked full-time with a median productivity shock value. We then scale them by a constant ω, which we estimate internally to match the average pension replacement rate of 47.7% in our data.

Finally, $\tilde{c}$ is the consumption floor given by the government to all agents, which is particularly relevant for those who do not work. We set this to 10% of average income.[34] Sickness benefits, captured by $\widetilde{T}$, paid to non-workers who are unhealthy are set to 11.5% of average income. Sickness benefits in Germany are, as a rule, based on 70% of the gross labor income and paid for a maximum duration of 78 weeks over three years for the same disease.[35] In the data, the average duration of payments due to sickness that are covered by the benefits ranges from 5 to 120 days per year, depending on the disease (Knieps and Pfaff, 2019). We choose an average duration of 60 days per year, which results in our chosen value for $\widetilde{T}$.

1.4.3 Estimation Results

Table 1.2 summarizes the internally estimated parameters (both point estimates and their standard errors), their target statistics, as well as the match between the empirical and model-implied data moments. We now discuss the fit of the model in greater detail along the dimensions relevant for the quantitative exercises in the following section.

The left panel of Figure 1.5 displays the employment rate by health status over 10-year age groups, comparing our model results with their data counterparts. The right panel shows

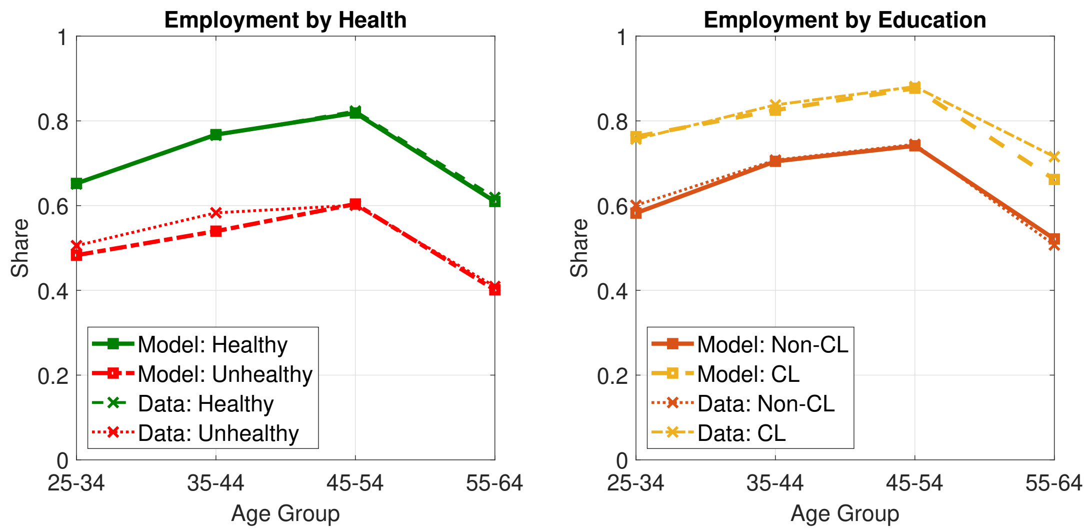
[34] In 2018, the calculated government transfer that is guaranteed as part of basic social security to secure the subsistence level was around 400 Euros per month for a single household (BAMS, 2018). This amounts to around 10% of average labor income in the same year.

[35] For the first up to 6 weeks after sickness, labor income is paid fully by their employer. After that, the health insurance company is mandated to pay. Eligibility of these sickness benefits depend on having worked for at least 4 weeks prior to sickness.

Table 1.2: Internally Estimated Parameters

Param -meter	Estimate	S.E.	Description	Target Statistics Model	Data	Description
Labor Supply and Wages						
$\nu_1^{h=1}$	2.634	0.399	Disutility of work	Figure 1.5		Age-Employment
$\nu_8^{h=1}$	1.666	0.081	parameters	(Left Panel)		Profiles by
$\nu_{13}^{h=1}$	1.278	0.027	(healthy)			Health
$\nu_{20}^{h=1}$	1.714	0.207				
$\nu_1^{h=0}$	2.412	0.445	Disutility of work			
$\nu_8^{h=0}$	1.813	0.126	parameters			
$\nu_{13}^{h=0}$	1.391	0.090	(unhealthy)			
$\nu_{20}^{h=0}$	2.415	0.377				
ν_e	0.807	0.005	Work Disutility for Non-CL	Figure 1.5 (Right Panel)		Employment by Education
$\zeta_0^{e=0}$	0.899	0.009	Deterministic wage	Figure 1.6		Age-Earnings
$\zeta_1^{e=0}$	0.0616	0.004	profiles			Profiles by Education
$\zeta_2^{e=0}$	-0.0025	0.0003	(non-college)			and Health
$\zeta_0^{e=1}$	1.165	0.026	Deterministic wage			
$\zeta_1^{e=1}$	0.0874	0.004	profiles			
$\zeta_2^{e=1}$	-0.0029	0.0002	(college)			
$w_p^{e=0}$	0.178	0.026	Wage loss for the unhealthy: Non-College			
$w_p^{e=1}$	0.145	0.051	Wage loss for the unhealthy: College			
Health Effort						
$\iota_1^{h=1,e=0}$	0.146	0.042	Disutility of effort	Figure 1.7		Age-Effort
$\iota_{12}^{h=1,e=0}$	0.560	0.066	parameters			Profiles by
$\iota_{20}^{h=1,e=0}$	1.048	0.086	(healthy +			Health and
$\iota_{31}^{h=1,e=0}$	1.603	0.081	non-college)			Education
$\iota_1^{h=0,e=0}$	0.628	0.186	Disutility of effort			
$\iota_{12}^{h=0,e=0}$	1.366	0.155	parameters			
$\iota_{20}^{h=0,e=0}$	1.650	0.129	(unhealthy +			
$\iota_{31}^{h=0,e=0}$	0.735	0.070	non-college)			
$\iota_1^{h=1,e=1}$	0.0913	0.024	Disutility of effort			
$\iota_{12}^{h=1,e=1}$	0.302	0.042	parameters			
$\iota_{20}^{h=1,e=1}$	0.740	0.065	(healthy +			
$\iota_{31}^{h=1,e=1}$	1.366	0.088	college)			
$\iota_1^{h=0,e=1}$	0.469	0.151	Disutility of effort			
$\iota_{12}^{h=0,e=1}$	0.997	0.143	parameters			
$\iota_{20}^{h=0,e=1}$	1.654	0.136	(unhealthy +			
$\iota_{31}^{h=0,e=1}$	1.089	0.084	college)			
ψ	1.115	0.067	f cost elasticity	0.163	0.161	Std.Dev.(f)
ς^0	0.00012	0.0001	Adjustment costs	0.256	0.267	Share of
ς^1	0.145	0.015		0.355	0.328	Non-Adjusters
				0.389	0.404	by Age Group
Remaining Parameters						
$\tilde{\kappa}$	0.872	0.038	Cons. Util. shifter	1.146	1.163	Cons. Ratio by Health
μ_β	0.943	0.003	Mean of β	Figure 1.8		Median Wealth Profiles
δ_β	0.0284	0.005	Dispersion of β	0.718	0.746	Wealth Gini
σ_x	0.0289	0.001	Produc. shock dispersion	0.585	0.595	Var(log income)
ω	0.359	0.011	Pension scale	0.473	0.477	Replacement rate
b	13.11	0.296	Utility constant	8.83	8.49	VSLY/$\bar{c}$

Figure 1.5: Model Fit of Employment by Health and by Education



Notes: Two-year employment rate by health status (left) and by education (right) over 10-year age groups in the model and data.

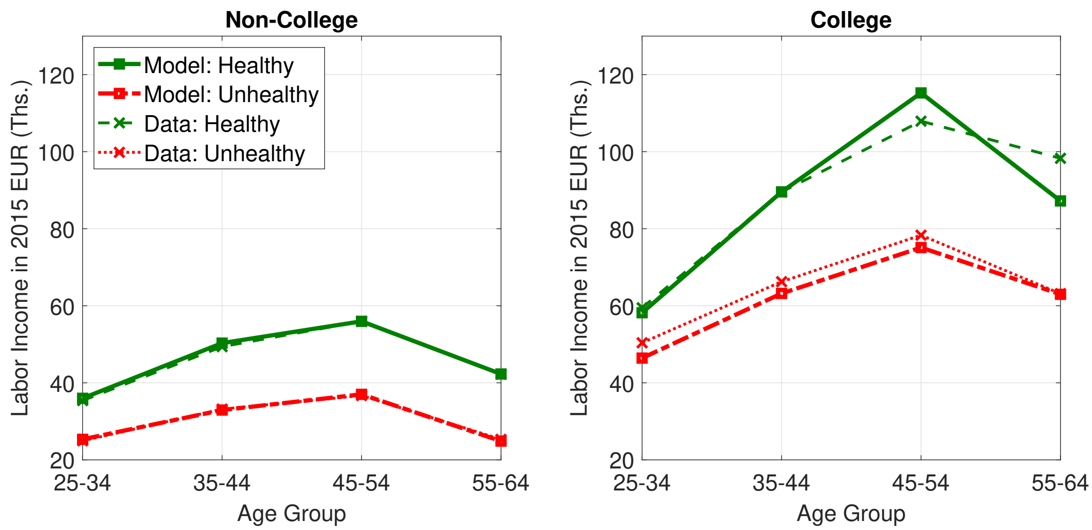
employment by education.[36] Similarly to what we observe in the data, the model generates a gap in the working population fraction by health. For example, at ages 25-34, the employment rate among healthy individuals is around 72%, whereas it is only 53% among the unhealthy. This gap in employment remains relatively constant over the working career. Similarly, our model replicates well the employment patterns by education, where non-college individuals work less than college individuals over all age groups. Notably, a constant additional work disutility for non-college workers suffices to generate the age pattern despite only targeting the average difference by education.

Figure 1.6 compares the life-cycle profiles of average labor income from our model-generated data with the SOEP data. We distinguish between the non-college (left panel) and college-educated (right) and plot the average earnings for healthy (green) and unhealthy (red) individuals. For both education groups, healthy individuals earn substantially more compared to unhealthy ones. Our model captures this difference conditional on education well. The productivity loss when working due to poor health is estimated to be 18% for non-college workers and 14% for college workers.

Figure 1.7 displays the evolution of average health effort over the life cycle by health status, again separating between the two education states. In the data, average health effort increases slightly for the non-college educated individuals over age and tends to be relatively

[36] In the data, we define two-year employment to be 1, if an individual is recorded as employed part- or full-time, or has labor income larger than 5,400 EUR in two consecutive years. If she is only recorded as employed for one year, we set 2-year employment to 0.5 and set it to 0 otherwise.

Figure 1.6: Model Fit of Labor Income by Health and Education



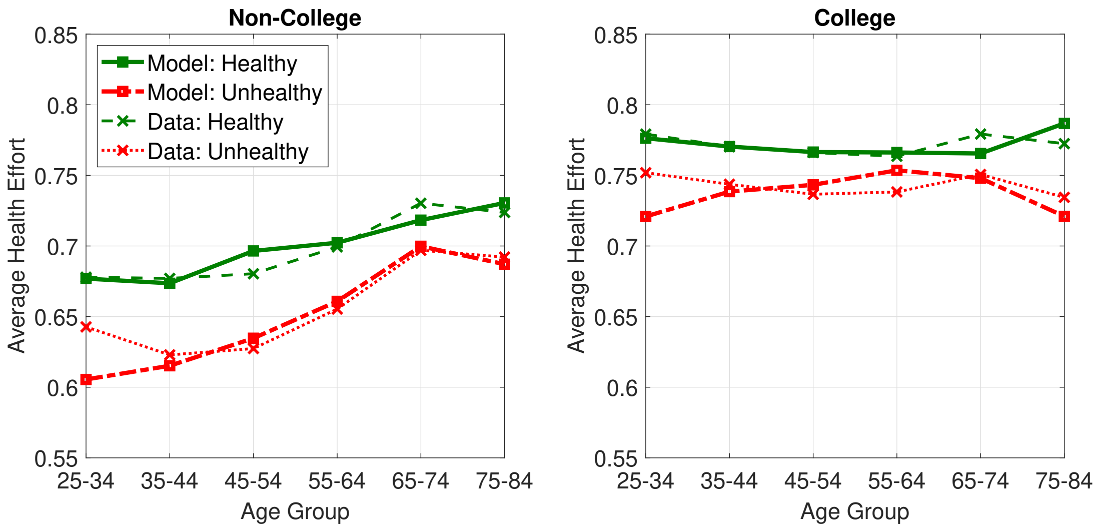
Notes: Average two-year labor income by 10-year age groups, distinguishing between healthy individuals (green) and unhealthy ones (red) in the data and the model. Left panel: Non-college educated individuals. Right panel: College educated individuals.

stable, albeit at a higher level, for the college-educated cones. Healthy individuals always exert more health effort compared to unhealthy ones. Our estimated model produces a similarly consistent difference between health groups, conditional on education.

Our estimation strategy is designed to discipline effort dynamics to be empirically reasonable along various dimensions. One such feature is the sizeable share of individuals who do not adjust their efforts, which in fact increases with age. Specifically, around 24% of young individuals (age 25-44) do not adjust their efforts, compared to a much higher share of 39% among the retired. Due to the adjustment costs that become more sizeable with age, our model replicates this pattern quite successfully.

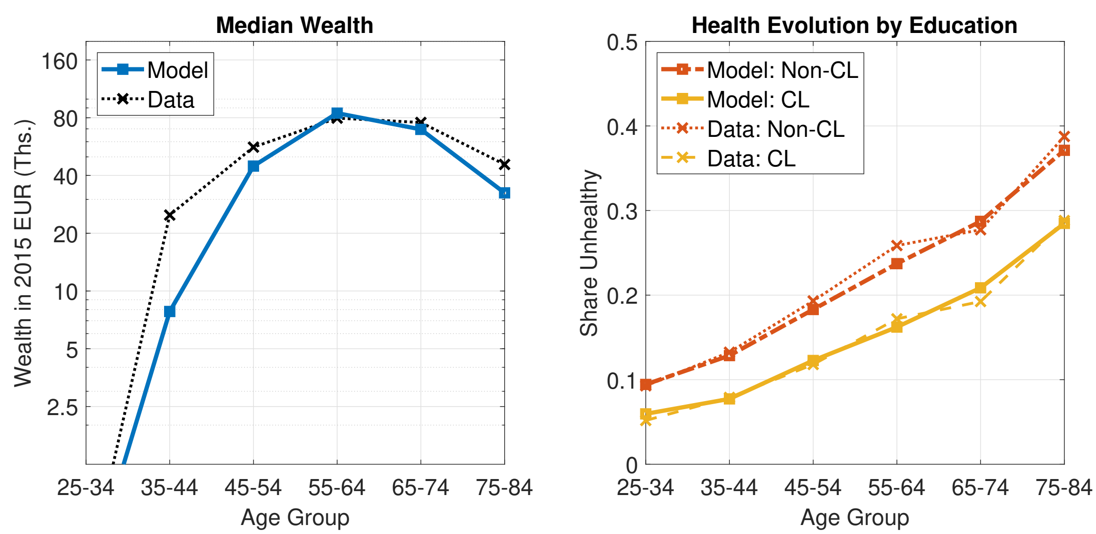
Finally, the left panel of Figure 1.8 shows median wealth profiles over age in the model and data, as before on a log/ratio scale. While we match the peak wealth in age group 55-64, the model produces slightly lower wealth levels at younger and older age groups. This is not surprising as in our model all agents start out with zero initial wealth and there are no bequests motives that would prompt individuals to maintain high wealth levels well into retirement. We estimate average β generating these profiles to be 0.943. Moreover, the differences in discount factors across β types is estimated to be 0.0284, which together with other forces in the model generates a Gini coefficient of wealth of around 0.72, slightly below its empirical counterpart.

Figure 1.7: Model Fit of Health Effort



Notes: Average health effort by 10-year age groups, distinguishing between individuals being healthy (green) and unhealthy (red) in the model and data. Left panel: Non-college educated individuals. Right panel: College educated individuals.

Figure 1.8: Model Fit of Wealth Evolution and Average Health



Notes: Left panel: Median wealth by 10-year age group in the model and data. Right panel: Average share of unhealthy individuals by education in the model and data.

1.4.4 Non-targeted Moments

We now turn to several relevant non-targeted moments generated by the model, as a validation check of our estimated model. First, our model successfully captures the evolution of health status in the data that we discussed in Section 1.2.1, as shown in the right panel of Figure 1.8.

In addition to the health-, and education-specific age profiles of health effort behavior, which we target in our estimation procedure, we also investigate how well our model captures the non-targeted adjustment patterns in individual lifestyle behaviors. To this end, the model produces an autocorrelation coefficient of health effort choices of 0.81, which is close to its data counterpart, 0.76. In light of the non-convex adjustment costs to health efforts, we further compare the model-generated shares of individuals that change their health effort levels by more than 10% or 20% to their empirical counterparts. Table 1.3 displays theses shares, separately for increases (positive changes) and decreases (negative changes) in health effort, by three different age groups.

We find that our model is successful in reproducing these micro-level adjustment distributions observed in the data. Overall, the model generates relatively large adjustments of around 20%, and their shares align quantitatively well with the data. Moreover, the model successfully generates asymmetry: for the same size changes, there is a higher fraction of agents making a positive adjustment compared to a negative adjustment for the young and prime-age groups. This is a salient feature in the data, which our model captures despite the fact that the estimation does not directly target these moments.

Table 1.3: Health Effort Adjustment at the Individual Level in Model and Data

Age Group	Shares with positive changes				Shares with negative changes			
	10%		20%		10%		20%	
	Model	Data	Model	Data	Model	Data	Model	Data
24-44	0.18	0.29	0.09	0.14	0.18	0.22	0.04	0.08
45-64	0.20	0.27	0.08	0.12	0.13	0.21	0.04	0.07
65-84	0.22	0.25	0.10	0.10	0.16	0.22	0.11	0.06

Notes: Average shares of individuals adjusting health effort in the model and data by age groups. Positive (Negative) Change: $\frac{f_j - f_{j-1}}{f_{j-1}} > (<) 10\%$ or 20%.

Finally, in line with the empirical observations outlined in Section 1.2.2, our model features a pronounced wealth gradient of lifestyle behaviors. To quantify this, we compute a wealth elasticity of health effort defined as the estimated coefficient on the logarithm of average wealth per age-group specific wealth quartiles from a linear regression of the logarithm of health effort on a constant, age group dummies and logarithm of average wealth per age-group specific wealth quartiles. We find that our model features a wealth elasticity of health effort of 2.4, which is very close to the one we obtain in the data at 2.5.

1.5 Quantitative Results

1.5.1 Wealth-Health Gaps and Channels

In this section, we use our estimated model to investigate the joint evolution of wealth and health and its underlying drivers. We begin by presenting how much of the wealth-health gaps are generated endogenously by our baseline model. The life cycle profiles of median wealth of healthy and unhealthy people are plotted in the left panel of Figure 1.9, as before on a log/ratio scale.[37]

We see that the relative gap in median wealth between the healthy (dashed green line) and the unhealthy (dotted red line) in the data is already present at young ages and persists throughout the life cycle. Our estimated model is able to endogenously generate a wealth-health gap that amounts to around three quarters of that observed in the data at younger ages, and that is as large as the one in the data for individuals between 65 and 74 years-old.[38] Given that our model agents differ in various characteristics, including rich ex-ante heterogeneity, one might wonder whether we should be surprised by this quantitative success.

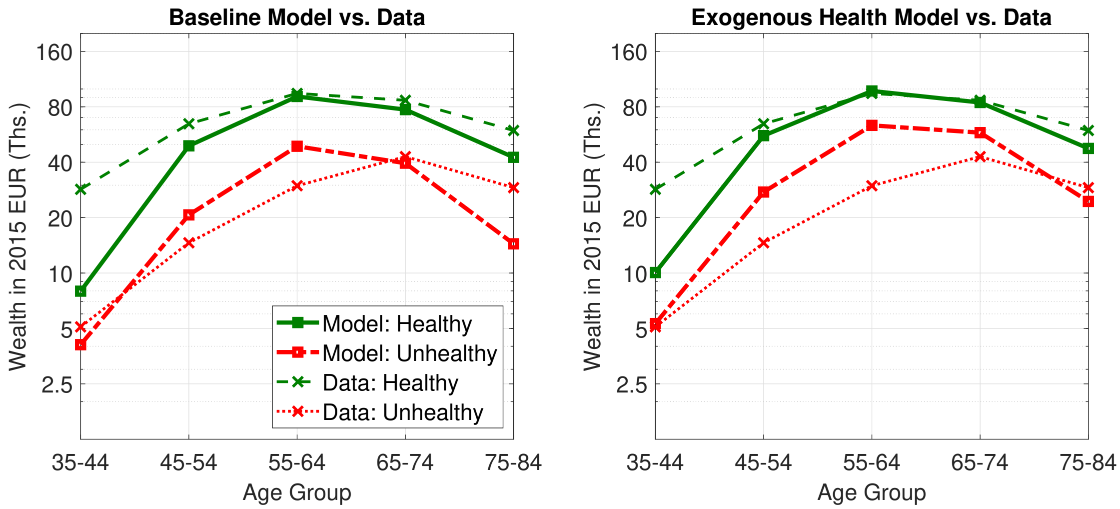
For that reason, we consider a variant of our model, where health transitions are no longer affected by health efforts, removing the need for the individual agents to decide on optimal health efforts. We estimate this *exogenous health model*, which still maintains the same rich ex-ante heterogeneity as in our baseline model, using a parallel estimation strategy and find that the model fits the target moments equally well.[39] However, as shown in the right panel of Figure 1.9, this exogenous health model can only account for less than two thirds of the non-targeted wealth-health gaps observed in the data, performing considerably worse than

[37] Since wealth levels are (almost) zero in the youngest age group (25-34 year-olds) both in the data and in the model, we plot the gaps from age group 35-44. We report the age profiles of wealth by health status at different points of the wealth distribution (25th percentile, 50th percentile, and 75th percentile) in Appendix Figure 1.K.2. At all wealth quartiles, the model generates sizable wealth-health gaps, which grow over age and are comparable in size as those in the data among prime-age groups. We also report the wealth-health gaps for each education group in Figure 1.K.3, confirming that our model generates sizable wealth-health gaps even conditional on education.

[38] It is not surprising that the model-generated gaps tend to open up later than in the data, given that all our model agents start with zero initial wealth.

[39] Concretely, we re-estimate the health transition probabilities in (1.16) without current and past health efforts but keeping all other covariates (see Table 1.G.1). Naturally, the estimated parameters exclude those shaping health effort disutility and adjustment in (1.13)-(1.14) and the target moments exclude those concerning health efforts.

Figure 1.9: Median Wealth Profiles by Health: Model vs. Data



Notes: The left panel displays median wealth by 10-year age groups, distinguishing between healthy individuals (green) and unhealthy ones (red) in the baseline model relative to the data. A log scale is used for the vertical axis. The right panel plots the counterparts from the re-estimated exogenous health model that abstracts from health efforts.

our endogenous health model. This result indicates that individual lifestyle behaviors contain valuable information for rationalizing the observed wealth-health gaps.

We now investigate and quantify channels behind these wealth-health gaps using a series of counterfactual experiments. To develop an intuition behind the logic of these experiments, we begin by presenting a greatly simplified version of our full model that nevertheless contains the key forces that are chiefly responsible for the wealth-health gaps. To that end, consider an individual who maximizes utility solving the following two-period problem (using the same notation as before):

$$\begin{aligned} &\max_{c_0,c_1,f,n} \; u_0(c_0) - \varphi(f) - \phi(n,h_0) + \beta S(h_1) u_1(c_1,h_1) \\ &\text{subject to} \quad c_0 + c_1 = w(h_0)n \\ &\qquad\qquad\qquad h_1 = \pi(f), \end{aligned} \tag{1.20}$$

where choice variables include current consumption (c_0), future consumption (c_1), lifestyle behaviors (f), and labor supply (n).[40] The key assumptions, as in our quantitative model, are that (i) better health improves the survival probability ($S'(h) > 0$); (ii) better health improves productivity or the wage offer ($w'(h) > 0$); (iii) health status affects the disutility of

[40]In this simple model, we abstract from several mechanisms that are present in our quantitative model to focus on illustrating the key mechanisms we highlight below. See Appendix 1.I for details.

labor supply ($\phi(n,h)$); (iv) better lifestyle behaviors improve health ($\pi'(f) > 0$); and (v) the marginal utility from future consumption is higher with better health ($\partial^2 u_1(c,h)/(\partial c \partial h) > 0$).

Using this simple model, we first illustrate two broad channels through which the health status affects wealth accumulation. The first is an *earnings channel*, which can drive the wealth-health relationship as unhealthy individuals mechanically earn less even when supplying the same hours (as they are less productive), but also because labor supply itself is affected by health status. To see how, we can combine the first order conditions of consumption and labor supply resulting from (1.20) (given by equations (1.27) and (1.28) in Appendix 1.I), which yields the labor-leisure condition:

$$\frac{\partial \phi(n,h_0)}{\partial n} = u_0'(c_0) w(h_0). \tag{1.21}$$

The left-hand side shows the marginal cost of labor supply, which is expected to be larger for the unhealthy. Hence, this force would induce the unhealthy to work less keeping wages constant. The right-hand side shows the marginal benefit of labor supply, which primarily consists of the wage. Since this is lower for the unhealthy, the first-order effect could potentially reduce the incentives to work.[41]

The second broad channel is a *savings channel*, which results from unhealthy individuals having different incentives to accumulate wealth compared to healthy individuals. The Euler equation resulting from (1.20) is given by:

$$u_0'(c_0) = \beta S(h_1) \frac{\partial u_1(c_1,h_1)}{\partial c_1} \tag{1.22}$$

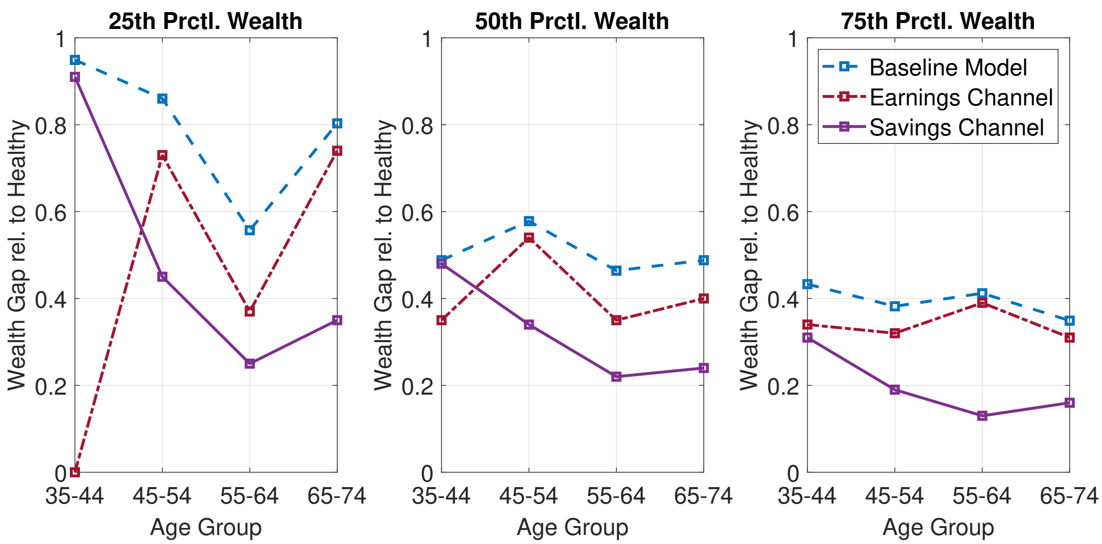
The right-hand side shows that savings can be higher with better health for two reasons; if one expects to live longer (i.e., a higher survival probability $S(h)$) or if one expects to have a higher quality of life (i.e., a higher marginal utility from consumption $\partial u(c_1,h)/\partial c_1$). We therefore expect this channel to contribute to the wealth-health gaps endogenously generated in the model.[42]

To quantify how important these channels working mostly from health to wealth are in our full quantitative life-cycle model, we perform two counterfactual experiments. First, to quantify the earnings channel, we assume that both the disutility from work and labor

[41] In practice, this effect depends on whether the substitution effect dominates the income effect as well as whether a health shock is permanent or not. See Appendix 1.I for further discussions.

[42] Note that our simplified model intentionally assumed that today's utility is independent of health to illustrate the savings channel clearly. Having the health-dependence in u_0 would affect the result, as discussed in Appendix 1.I.

Figure 1.10: Effects of the Earnings and Savings Channels on Wealth-Health Gaps



Notes: Differences in the wealth levels of those being healthy and unhealthy at the 25th (left), 50th (middle), and 75th (right) percentile of the wealth distribution in the baseline model (blue), and in the counterfactual scenarios without differences in labor supply disutility and labor productivity by health (red), and with average savings choices across health status (purple) across 10-year age groups. The counterfactual experiments are calculated using the baseline distribution of health.

productivity are no longer affected by health status (i.e., the disutility of labor supply is as if one was healthy for everyone and $w_p^e = 0$ for both education groups). This effectively shrinks the differences in labor incomes across health status. Second, to quantify the savings channel, we assume that the survival probability is as if one was healthy for everyone (i.e., $S_j(h_j = 1, e)\ \forall j, e$), and that the consumption utility and value of life is no longer diminished from being unhealthy (i.e., $\tilde{\kappa} = 1$). This reduces differences in the incentives to accumulate wealth between the healthy and unhealthy, conditional on other states. In both exercises above, we let agents behave optimally in terms of their labor supply, savings, and health effort choices. However, to isolate the effects going from health to wealth, we keep the baseline distribution of health when we simulate the counterfactual economy.[43]

Figure 1.10 summarizes the effects of these experiments on the wealth gap between healthy and unhealthy agents at the 25th percentile (left), the median (center) and the 75th percentile (right) and over age, expressed relative to the wealth of the healthy.[44] Both the red dash-

[43] That is, unbeknownst to the agents in the model, their health outcomes at the beginning of each period are set to be exactly the same as in the baseline economy. This also implies that survival realizations are the same as in the baseline.

[44] For the counterfactual exercises hereafter, we present wealth-health gaps in relative terms to ease interpretation. They are constructed as the difference between wealth owned by healthy and unhealthy individuals in a given age groups, divided by wealth of the healthy. Thus, a number of 0.6, for example, means that going from healthy to unhealthy amounts to a 60% drop in wealth or that unhealthy individuals own 40%

dotted line, illustrating the experiment of closing the earnings channel, and the purple solid line, which depicts the gaps after removing the savings channel as defined above, are below the baseline blue dotted line throughout the life cycle. This suggests that both channels contribute to the wealth-health gaps. Yet, their relative importance differs across age groups and wealth positions. The earnings channel is quantitatively more important for the younger, and particularly asset-poor agents, for whom wealth levels are relatively small such that differences in savings across health status are of little consequence. In contrast, differences in earnings across health status play a major role, as they provide almost the sole basis for wealth accumulation. In fact, at the 25th wealth percentile, minimizing such differences effectively closes the entire model-generated wealth-health gap in age group 35-44. At median wealth levels, the gaps between those being healthy and unhealthy are reduced by over 10 percentage points in that age group.

For all other age groups however, the effect of turning off the savings channel has quantitatively larger implications for the wealth-health gaps. The effect is particularly strong for asset-rich individuals, where the gaps are approximately halved, on average, and even reduced by almost 70% at age group 55-64. With the exception of the youngest age groups, the relative importance of the savings channel for driving the wealth-health gaps is quite constant across age. In sum, these results suggest that different savings incentives originating from differences in the length and quality of life across health status are an important reason why relative wealth-health gaps are persistent over the life-cycle.

Against the backdrop of the illustration in the simple model above, we use our model to further decompose the contributions of the earnings channel into effects that work through health-dependent labor productivity and disutility from work separately. As shown in Table 1.J.1, we find that the former is quantitatively much more important and that these two sub-channels are complementary to each other in generating the total effects of the earnings channel. Similarly, we further decompose the contributions of the savings channel into effects that come from the quality of life (i.e. through differences in κ) and effects that work through the length of life (survival rates) across healthy and unhealthy agents. We find that the survival channel is quantitatively more relevant in delivering the total effects of the savings channel, especially for the relatively older individuals, as shown in Table 1.J.1.

1.5.2 Heterogeneity in Lifestyle Behaviors and Wealth-Health Gaps

In Section 1.2.2, we presented suggestive evidence that lifestyle behaviors could contribute to the positive association between wealth and health observed in the data. In contrast to the

of the wealth of healthy ones at that point in the distribution.

channels investigated in Section 1.5.1 that run from health to economic outcomes, endogenous lifestyle choices have the potential to capture effects running in the other direction. By doing so, they can potentially *amplify* the wealth-health relationship over the life-cycle if good economic outcomes and higher wealth lead to higher effort choices, which in turn improve the probability of good health outcomes, feeding back into the channels in Section 1.5.1.[45] We investigate these effects in our model in two ways: First, we quantify the extent to which differing lifestyle behaviors across individuals explain the large wealth-health gaps in the model. Second, we illustrate how wealth impacts lifestyle choices, net of other factors.

Regarding the first way, we perform a counterfactual experiment, in which we force all agents to choose the age-specific average health effort level at the baseline model.[46] The rest of the model remains unchanged and we let the agents optimize given this constraint. In particular, the earnings and savings channels of health we investigated in Section 1.5.1 are operative in generating wealth-health gaps.

Figure 1.11 summarizes the wealth-health gaps in the data, the baseline model and the counterfactual model with equalized health effort choices at three different points along the wealth distribution. Equalizing health efforts throughout the life span reduces the wealth-health gaps across the wealth distribution relative to the baseline economy. For example, the maximum percentage difference in median wealth of the unhealthy relative to the healthy is reduced to around 33% from around 46% in the baseline at ages 55-64. Across the life cycle, equalizing health efforts reduces the relative wealth-health gap on average by 12% at the 25th percentile, by 23% at the median, and by 29% at the 75th percentile relative to the baseline model. These findings obtained in the presence of the earnings and savings channels yet in the absence of effort heterogeneity suggest that individual health behaviors are an important amplification mechanism for wealth-health gaps.[47]

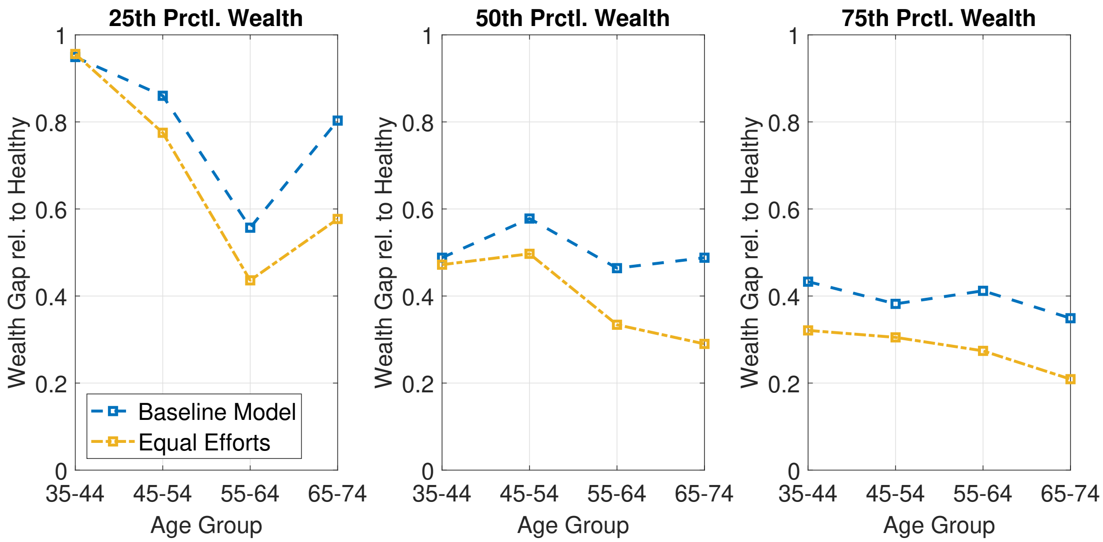
When we force everyone to choose the same average lifestyles, we remove heterogeneity in health outcomes that arises solely from differences in lifestyle behaviors. Since our model features a realistic positive wealth gradient of health efforts, this on average reduces the

[45] Such an amplification mechanism could therefore be especially powerful if the wealth-gradient in health efforts observed in both model and data is driven by higher wealth itself, on top of third factors such as education.

[46] In this exercise, we therefore maintain the estimated effects of other characteristics such as education on health transitions when removing effort heterogeneity, whereas the re-estimated exogenous health model does not (as can be seen in Table 1.G.1). Moreover, the current exercise allows us to flexibly explore the role of differences in lifestyle behaviors at different points in the life cycle.

[47] If we close both savings channel and earnings channel in the model, there are no incentives left to exert health efforts as being healthy has no benefits. Yet, some of the wealth-health gaps remain, as shown in Figure 1.J.1. This is because there remain other factors in the model that drive the evolution of both health and wealth. In particular, education affects the probability of being healthy even without any efforts, while at the same time generating higher wages.

Figure 1.11: Effect of Equalizing Health Efforts on Wealth-Health Gaps



Notes: Differences in the wealth levels of those being healthy and unhealthy at the 25th (left), 50th (middle), and 75th (right) percentile of the wealth distribution in the baseline model (blue) and in the counterfactual scenario with constant health effort choices (yellow). Differences are expressed relative to the wealth levels of the healthy.

share of good health outcomes among rich individuals and increases the share of good health outcomes among poorer individuals, keeping the distribution of wealth fixed, which decreases the wealth-health gap.[48] At the same time, the counterfactual of equalizing efforts could in principle also affect other choices that drive the gap, even without its effect on health.[49] In Appendix 1.J, we quantify these two effects and find that total effects of equalizing efforts works primarily through its direct effect on the health distribution.

In addition, given the habitual character of lifestyle behaviors both in the data and in the model, it is conceivable that behavior differences at younger ages matter relatively more for the whole life cycle than those at older ages. In Figure 1.J.2 in Appendix 1.J, we investigate the extent to which the wealth-health gaps are differently affected according to the timing of equalizing health behaviors. The results suggest that eliminating effort variation during earlier life years, especially in prime ages, has prominent lasting effects in terms of reducing the wealth-health gaps in later years.

The second important question is then what drives heterogeneity in lifestyle behaviors, in

[48]Therefore, by construction, *averages* of key variables such as life expectancy, health, earnings and health barely change.

[49]For example, an agent choosing lower health efforts relative to the baseline may find it optimal to also save less in anticipation of worse health outcomes in the future, which will make consumption less enjoyable. For the same reason, however, she might also save more to insure against the risk of not being able to work because of poor health outcomes. Overall, these indirect effects of the effort equalization counterfactual on the relationship between wealth and health are therefore ambiguous.

particular, along the wealth dimension. Although the wealth-gradient in lifestyles is likely in part driven by ex-ante heterogeneity, more wealth raises the incentive to exert better lifestyle behaviors even conditional on these fixed types. To see this, we resort again to the simple model (1.20), this time considering the optimality condition for efforts derived in Section 1.I:

$$\varphi'(f) = \beta S'(h_1)\pi'(f)u_1(c_1, h_1) + \beta S(h_1)\frac{\partial u_1(c_1, h_1)}{\partial h_1}\pi'(f). \tag{1.23}$$

The right-hand side determines the benefit of exerting more efforts. Its first term shows that improvement in the survival probability driven by better health is multiplied by the *level of utility*, a feature that is common to models with endogenous survival. Since utility levels are increasing in (future) wealth, richer individuals, or those expecting to be rich in the future should, all things equal, thus have a stronger incentive to exert health efforts. This also means that the (anticipation of) redistribution of future consumption has the potential to reduce current disparities in lifestyle behaviors. This, in turn, could reduce inequalities in future health outcomes and consequently narrow wealth-health gaps.

We illustrate the importance of these dynamic effects working through endogenous lifestyle behaviors using the following experiment in our quantitative model. We solve for optimal effort choices in a counterfactual economy where all agents think that when entering retirement, all assets and pensions will be taxed at 100% and everyone instead receives transfers that equal exactly the average retirement wealth in the baseline economy. In the simulation of the distribution, however, we maintain the savings and labor supply levels of the baseline model for every agent. Thus, only effort choices and their consequences for the health distribution are changed.

We report the results of this experiment in Table 1.4. Panel A shows the percentage changes in average health effort, conditional on wealth quartiles and age groups. For almost every age group and wealth quartile, agents increase their efforts relative to the baseline case.[50] This rise in healthy behaviors is accelerated with age. Moreover, there is a clear negative trend in the change in efforts going from the first wealth quartile to the fourth one at every age group. This is precisely because for rich individuals, this counterfactual scenario does not lead to significantly different expectations in wealth levels during retirement. For that reason, they do not need to change their lifestyles (which were already at a high level). Poor individuals, on the other hand, have much stronger incentives to survive and be healthy in later years, anticipating increased wealth that will allow them to enjoy a larger utility from consumption.

[50] This is sensible given that the size of the average uniform transfers is quite generous for a large fraction of agents given the skewed wealth distribution.

The adjustments in health efforts translate into changes in health outcomes, as shown in changes in the share of individuals in bad health in Panel B of Table 1.4. As expected the share drops in particular among poorer individuals. Mirroring the lifestyle changes, the improvements in average health again become stronger with age, but are already visible even before retirement. Taken together, the disparities in health outcomes of poor and rich individuals therefore become smaller, which eventually narrows the wealth-health gap (as presented in Panel C of Table 1.4), even in the absence of the earnings and savings channels defined in Section 1.5.1.

These results also indicate that changes in economic conditions during the life course can lead to meaningful changes in the distribution of health outcomes. A natural question is then to ask how much of inequality in health outcomes is pre-determined at the initial period (age 25). Using a decomposition exercise following Huggett et al. (2011) (as discussed in details in Appendix 1.H), our model shows that although initial conditions at age 25 play a substantial role in shaping the variation in economic outcomes, such as lifetime earnings, they are less important for explaining lifetime inequality in health-related outcomes. For example, approximately one-third of the variation in the share of healthy life years is predetermined by the conditions at age 25, in contrast to nearly 80% for lifetime earnings. In sum, these results add support to the idea that lifestyle behaviors, which allow individuals to react to changing economic circumstances, can act as an amplification mechanism between economic outcomes and health over the life cycle.[51]

1.6 Conclusion

We document a strong association between individual wealth and health over the life cycle in Germany. We then build a structural life-cycle model of endogenous wealth and health evolution as individual lifestyle behaviors shape future health outcomes. These, in turn, affect wealth accumulation through differences in earnings and savings behaviors across health status. Our estimated model accounts for the great majority of the empirical wealth-health gaps, rationalizing that large and persistent wealth-health gaps can occur even in countries where the healthcare system does not frequently entail large out-of-pocket expenses. Through

[51]The fact that health efforts react to changes driven by future wealth leaves untouched other reasons that drive effort choices that also work through the utility level channel, and could potentially also affect the wealth-health relationship. For example, the return to efforts is higher when the future is expected to be more enjoyable, which is the case not only when one is rich but also healthy. Moreover, during working years, the return to effort includes an effect coming through higher expected future wages. Interestingly, this last motive can be decreasing in wealth, as we show in Appendix 1.I. Generally, the direction in which such forces could affect the wealth-health relationship is often not clear, and a quantitative exploration goes beyond the scope of this paper.

Table 1.4: Results of Equalizing Wealth during Retirement Periods

Unit: %	Panel A				Panel B				Panel C		
	Average Effort by Wealth Quartile				**Share Bad Health** by Wealth Quartile				**Wealth-Health Gaps** at Percentile		
Age Group	1st	2nd	3rd	4th	1st	2nd	3rd	4th	25th	50th	75th
35-44	0.4	0.7	0.1	-0.3	0.0	-0.8	-1.2	0.0	0.0	0.0	-1.1
45-54	2.0	1.4	0.7	0.0	-1.5	-0.7	0.0	0.0	-0.1	-0.8	-0.7
55-64	7.7	4.0	1.8	0.1	-4.1	-2.3	-1.1	0.0	-4.3	-5.5	-5.6
64-75	11.2	10.2	4.8	1.1	-8.9	-7.4	-4.3	-1.1	-7.0	-16.4	-17.6

Notes: Reported numbers are percentage changes relative to the benchmark case without the counterfactual experiment. The counterfactual experiment assumes that effort choices are based on the belief of a uniform 100% tax on wealth and retirement benefits during retirement years along with transfers equal to the average retirement wealth in the baseline economy.

a series of decomposition exercises, we find that, quantitatively, while the earnings channel is important for the young and asset poor, the savings channel drives the wealth-health gaps at most ages, and especially for asset-rich individuals. We demonstrate that lifestyle behaviors can act as an amplification mechanism behind the dynamic relationship between wealth and health since good economic outcomes lead to higher health effort choices in our model.

While our model is relatively rich, we abstract from several potentially relevant mechanisms, in particular those through which money itself could influence future health. These include private medical expenditures, preventive monetary investments in health, and higher-quality but costly private insurance options. While we believe that these channels are likely less important in the German context, as we discuss in Appendix 1.A, they could nonetheless help the model to match the wealth-health gaps more closely. Moreover, these channels are crucial to consider when analyzing other countries where out-of-pocket medical expenses are more prevalent and private insurance frequently consists of better healthcare relative to the public option.

Our results imply that policies aimed at improving individual health behaviors (e.g., conditional cash transfers when joining a gym Charness and Gneezy, 2009), can result not only in lasting benefits in terms of improving health inequality over the life course but may also extend into dimensions of economic inequality. Conversely, our findings also suggest that rising wealth inequality may, by exacerbating heterogeneity in lifestyles, contribute to consolidating the pronounced positive association between economic- and health-related well-being, and could underlie the increasing divergence in health-related behaviors observed in recent years (Lampert et al., 2018). We leave this interesting empirical question for future work.

Bibliography

Attanasio, O., S. Kitao, and G. L. Violante (2010): "Financing Medicare: A general equilibrium analysis," in *Demography and the Economy*, University of Chicago Press, 333–366.

BAMS (2018): "Federal Ministry for Labor and Social Issues," https://www.bmas.de/DE/Service/Presse/Pressemitteilungen/2017/hoehere-regelbedarfe-in-der-grundsicherung-und-sozialhilfe.html.

Barbaresko, J., J. Rienks, and U. Nöthlings (2018): "Lifestyle indices and cardiovascular disease risk: a meta-analysis," *American Journal of Preventive Medicine*, 55, 555–564.

Becker, G. S. (2007): "Health as human capital: synthesis and extensions," *Oxford Economic Papers*, 59, 379–410.

Blundell, R., M. Borella, and J. Commault (2023a): "Old Age Risks, Consumption, and Insurance," Tech. rep., CEPR Discussion Papers.

Blundell, R., M. C. Dias, J. Britton, and E. French (2023b): "The Impact of Health on Labor Supply near Retirement," *Journal of Human Resources*, 58, 282–334.

Bonhomme, S., T. Lamadon, and E. Manresa (2022): "Discretizing unobserved heterogeneity," *Econometrica*, 90, 625–643.

Capatina, E. (2015): "Life-cycle effects of health risk," *Journal of Monetary Economics*, 74, 67–88.

Capatina, E., M. Keane, and S. Maruyama (2020): "Health Shocks and the Evolution of Earnings over the Life-Cycle," Tech. rep., School of Economics, The University of New South Wales.

Cawley, J. and C. J. Ruhm (2011): "The economics of risky health behaviors," in *Handbook of Health Economics*, Elsevier, vol. 2, 95–199.

Cena, H. and P. C. Calder (2020): "Defining a healthy diet: evidence for the role of contemporary dietary patterns in health and disease," *Nutrients*, 12, 334.

Cesarini, D., E. Lindqvist, R. Östling, and B. Wallace (2016): “Wealth, health, and child development: Evidence from administrative data on Swedish lottery players,” *The Quarterly Journal of Economics*, 131, 687–738.

Charness, G. and U. Gneezy (2009): “Incentives to exercise,” *Econometrica*, 77, 909–931.

Chen, C., Z. Feng, and J. Gu (2022): “Health, Health Insurance, and Inequality,” Tech. rep., University of Toronto, Department of Economics.

Cocci, M. D. and M. Plagborg-Møller (2021): “Standard errors for calibrated parameters,” *arXiv preprint arXiv:2109.08109*.

Cole, H. L., S. Kim, and D. Krueger (2019): “Analysing the Effects of Insuring Health Risks: On the Trade-off between Short-Run Insurance Benefits versus Long-Run Incentive Costs,” *The Review of Economic Studies*, 86, 1123–1169.

Colman, G. J. and D. M. Dave (2013): “Physical activity and health,” Tech. rep., National Bureau of Economic Research.

Cutler, D. M. and A. Lleras-Muney (2010): “Understanding differences in health behaviors by education,” *Journal of Health Economics*, 29, 1–28.

Cutler, D. M., A. Lleras-Muney, and T. Vogl (2011): “Socioeconomic Status and Health: Dimensions and Mechanisms,” in *The Oxford Handbook of Health Economics*, Oxford University Press, 124–163.

Darden, M., D. B. Gilleskie, and K. Strumpf (2018): “Smoking and mortality: New evidence from a long panel,” *International Economic Review*, 59, 1571–1619.

De Nardi, M., E. French, and J. B. Jones (2010): “Why do the elderly save? The role of medical expenses,” *Journal of Political Economy*, 118, 39–75.

De Nardi, M., S. Pashchenko, and P. Porapakkarm (2023): “The Lifetime Costs of Bad Health,” Working Paper 23963, National Bureau of Economic Research.

Finkelstein, A., E. F. Luttmer, and M. J. Notowidigdo (2013): “What good is wealth without health? The effect of health on the marginal utility of consumption,” *Journal of the European Economic Association*, 11, 221–258.

French, E. (2005): “The effects of health, wealth, and wages on labour supply and retirement behaviour,” *The Review of Economic Studies*, 72, 395–427.

French, E. and J. B. Jones (2011): “The effects of health insurance and self-insurance on retirement behavior,” *Econometrica*, 79, 693–732.

Frick, J. R., M. M. Grabka, and J. Marcus (2007): “Editing and multiple imputation of item-non-response in the 2002 wealth module of the German Socio-Economic Panel (SOEP),” Tech. rep., SOEPpapers on Multidisciplinary Panel Data Research.

Glover, A., J. Heathcote, D. Krueger, and J.-V. Ríos-Rull (2023): "Health versus wealth: On the distributional effects of controlling a pandemic," *Journal of Monetary Economics*, 140, 34–59.

Hai, R. and J. J. Heckman (2022): "The causal effects of youth cigarette addiction and education," Tech. rep., National Bureau of Economic Research.

Hall, R. E. and C. I. Jones (2007): "The value of life and the rise in health spending," *The Quarterly Journal of Economics*, 122, 39–72.

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

Hosseini, R., K. A. Kopecky, and K. Zhao (2021): "How Important Is Health Inequality for Lifetime Earnings Inequality?" FRB Atlanta Working Paper.

——— (2022): "The evolution of health over the life cycle," *Review of Economic Dynamics*, 45, 237–263.

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

Jang, Y. (2023): "Credit, default, and optimal health insurance," *International Economic Review*, 64, 943–977.

Jung, J. and C. Tran (2016): "Market inefficiency, insurance mandate and welfare: US health care reform 2010," *Review of Economic Dynamics*, 20, 132–159.

Karlsson, M., T. J. Klein, and N. R. Ziebarth (2016): "Skewed, persistent and high before death: Medical spending in Germany," *Fiscal Studies*, 37, 527–559.

Khan, A. and J. K. Thomas (2008): "Adjustment costs," *The New Palgrave Dictionary of Economics. Palgrave Macmillan, Basingstoke.*

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

Kitao, S. (2014): "A life-cycle model of unemployment and disability insurance," *Journal of Monetary Economics*, 68, 1–18.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007): "Experimental analysis of neighborhood effects," *Econometrica*, 75, 83–119.

Knieps, F. and H. Pfaff (2019): *BKK Gesundheitsreport 2019: Psychische Gesundheit und Arbeit Zahlen, Daten, Fakten*, MWV.

Knoops, K. T. B., L. C. de Groot, D. Kromhout, A.-E. Perrin, O. Moreiras-Varela, A. Menotti, and W. A. Van Staveren (2004): "Mediterranean diet, lifestyle factors, and 10-year mortality in elderly European men and women: the HALE project," *The Journal of the American Medical Association*, 292, 1433–1439.

Kopecky, K. and T. Koreshkova (2014): "The impact of medical and nursing home expenses on savings," *American Economic Journal: Macroeconomics*, 6, 29–72.

Kvasnicka, M., T. Siedler, and N. R. Ziebarth (2018): "The health effects of smoking bans: Evidence from German hospitalization data," *Health Economics*, 27, 1738–1753.

LaCroix, A. Z., J. Lang, P. Scherr, R. B. Wallace, J. Cornoni-Huntley, L. Berkman, J. D. Curb, D. Evans, and C. H. Hennekens (1991): "Smoking and mortality among older men and women in three communities," *New England Journal of Medicine*, 324, 1619–1625.

Lampert, T., L. E. Kroll, B. Kuntz, and J. Hoebel (2018): "Health inequalities in Germany and in international comparison: trends and developments over time," *Journal of Health Monitoring*, 3.

Lee, I.-M. (2003): "Physical activity and cancer prevention–data from epidemiologic studies." *Medicine and Science in Sports and Exercise*, 35, 1823–1827.

Low, H. and L. Pistaferri (2015): "Disability insurance and the dynamics of the incentive insurance trade-off," *American Economic Review*, 105, 2986–3029.

Margaris, P. and J. Wallenius (2023): "Can Wealth Buy Health? A Model of Pecuniary and Non-Pecuniary Investments in Health," *Journal of the European Economic Association*, jvad044.

OECD (2012): *Mortality Risk Valuation in Environment, Health and Transport Policies*, OECD Publishing.

——— (2019): "Germany: Country Health Profile 2019, State of Health in the EU," *OECD/Eurpean Observatory on Health Systems and Policies*.

Ozkan, S. (2017): "Preventive vs. curative medicine: A macroeconomic analysis of health care over the life cycle," Tech. rep., University of Toronto, Department of Economics.

O'Donnell, O., E. Van Doorslaer, and T. Van Ourti (2015): "Health and inequality," in *Handbook of Income Distribution*, Elsevier, vol. 2, 1419–1533.

Pashchenko, S. and P. Porapakkarm (2017): "Work incentives of Medicaid beneficiaries and the role of asset testing," *International Economic Review*, 58, 1117–1154.

Pijoan-Mas, J. and J.-V. Ríos-Rull (2014): "Heterogeneity in expected longevities," *Demography*, 51, 2075–2102.

Poterba, J. M., S. F. Venti, and D. A. Wise (2017): "The asset cost of poor health," *The Journal of the Economics of Ageing*, 9, 172–184.

Prados, M. J. (2018): "Health and earnings inequality over the life cycle: The redistributive potential of health policies," *Manuscript, Columbia University, New York, NY.*

Richter, D. and J. Schupp (2014): "SOEP 2006-TIMEPREF: Dataset on the economic behavior experiment on time preferences in the 2006 SOEP Survey," Tech. rep., SOEP Survey Papers.

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

Schlesinger, S., M. Neuenschwander, A. Ballon, U. Nöthlings, and J. Barbaresko (2020): "Adherence to healthy lifestyles and incidence of diabetes and mortality among individuals with diabetes: a systematic review and meta-analysis of prospective studies," *Journal of Epidemiology and Community Health*, 74, 481–487.

Schnell-Inderst, P., T. Hunger, K. Hintringer, R. Schwarzer, V. Seifert-Klauss, H. Gothe, J. Wasem, and U. Siebert (2011): "Individuelle Gesundheitsleistungen," *Schriftenreihe Health Technology Assessment*, 113.

Schwandt, H. (2018): "Wealth shocks and health outcomes: Evidence from stock market fluctuations," *American Economic Journal: Applied Economics*, 10, 349–77.

Storesletten, K., C. I. Telmer, and A. Yaron (2004): "Consumption and risk sharing over the life cycle," *Journal of Monetary Economics*, 51, 609–633.

Van Oyen, H., N. Berger, W. Nusselder, R. Charafeddine, C. Jagger, E. Cambois, J.-M. Robine, and S. Demarest (2014): "The effect of smoking on the duration of life with and without disability, Belgium 1997–2011," *BMC Public Health*, 14, 1–12.

Verdun, Z. S. (2022): "Impact of a health shock on lifestyle behaviours," Tech. rep., European University Institute.

Zhao, K. (2014): "Social security and the rise in health spending," *Journal of Monetary Economics*, 64, 21–37.

Appendices to Chapter 1

1.A Medical Spending in Germany

The healthcare system in Germany is characterized by the co-existence of two insurance systems. Almost 90% of the population are covered by statutory health insurance (SHI), while the remaining share is covered by a substitutive private health insurance (PHI). Only individuals with an annual income above a certain opt-out threshold (currently around 64,000 EUR annually in 2022), the self-employed, or civil servants can choose to be covered by a PHI. A detailed discussion of the differences between the two insurance types and their funding and reimbursement schemes can be found in Karlsson et al. (2016). Notably, SHI coverage, as mandated by law, includes a very generous package of benefits, including all medically necessary treatments, prescription drugs, and, importantly for our purpose, preventive, and rehabilitation care. The PHI benefit packages are more heterogeneous but typically oriented towards the public package. They may include additional features, such as preferential treatment in hospitals, or dental and eye care. Given that PHI enrollees are generally wealthier, as they tend to be better educated and earn higher incomes (Karlsson et al., 2016), if these features materially improve individual health, they may be an important explanatory factor for the wealth-health relationship.

On top of that, there are numerous "individual health services", including non-standard screenings and therapies that are increasingly offered by physicians but are typically paid for directly by the patients and not covered by health insurance. Similarly, other potentially health-promoting expenses on nutritional supplements, physical treatments or even private psychological counselling could theoretically strengthen the wealth-health relationship if these are normal goods and significantly improve an individual's future health prospects.

However, the use of many of these health services is at least scientifically unclear, and they often comprise medically unnecessary cosmetic and luxury treatments or use methods whose benefits have not been sufficiently certified (Schnell-Inderst et al., 2011).[52] Moreover,

[52] This is not to say that in given circumstance, such services may be very sensible. However, consumer pro-

Table 1.A.1: Effect of Earnings and Wealth on Spending on Health Goods

	Cons. of Health Goods and Services$_i$	
Good Health$_i$	-108.7***	-107.9**
	(53.1)	(59.2)
Age$_i$	8.1***	6.7***
	(1.0)	(1.3)
College$_i$	104.3***	92.7***
	(32.7)	(28.3)
Earnings$_i$	0.7	
	(0.5)	
Wealth$_i$		0.07
		(0.05)
N	16,193	11,314
R^2	0.007	0.006

Notes: The dependent variable is annual household consumption spending on health goods and services. Coefficients and standard errors (in parentheses) of earnings and wealth are multiplied by 1,000. Stars denote statistical significance at the 10%, 5%, and 1% level.

using data on household consumption spending from the 2010 survey wave of the SOEP, we do not see a significant statistical correlation between spending on health-related goods and services and labor income (or wealth) after controlling for individual characteristics (that are also present in our model). Table 1.A.1 shows the results of a linear regression of annual consumption of health-related goods and services on a dummy for good health, age, college education, and labor income or wealth, respectively. In line with our expectations, the estimated coefficients indicate that individuals in good health spend significantly less on health-related consumption, while older and higher educated individuals tend to spend more.[53] Labor income or wealth, in contrast, are not statistically significantly associated with higher health-related consumption.

Notwithstanding this suggestive evidence, there can be alternative possibilities through which larger financial resources could affect health that go beyond direct medical goods and services. These include, for instance, access to better housing in less polluted, quieter neighborhoods, the possibilities of more frequent or costly recreational activities or vacations, and potential effects of wealth on psychological stress, which can also translate to physical health conditions (Schwandt, 2018). However, such effects are hard to detect statistically as

tection authorities frequently warn against using unsolicited health services without extensive information.

[53] Karlsson et al. (2016) investigate individual medical spending using data from a private health insurer and find that medical spending increases over age and is particularly concentrated in the last three years before death.

they likely take a long time horizon to realize and are dependent on individual circumstances. Perhaps unsurprisingly, the literature that tries to establish a causal link from resources to health among adults in developed countries remains debatable (Cutler et al., 2011).

In sum, the arguments provided in this discussion lead us to believe that a "money can buy health" channel is less relevant in Germany than it might be in other countries, such as the U.S. Thus, our paper focuses on another margin that is frequently pondered as an important mechanism behind the wealth-health relationship: lifestyle behaviors (Cutler et al., 2011; Cawley and Ruhm, 2011).

1.B Comparison of Different Health Measures

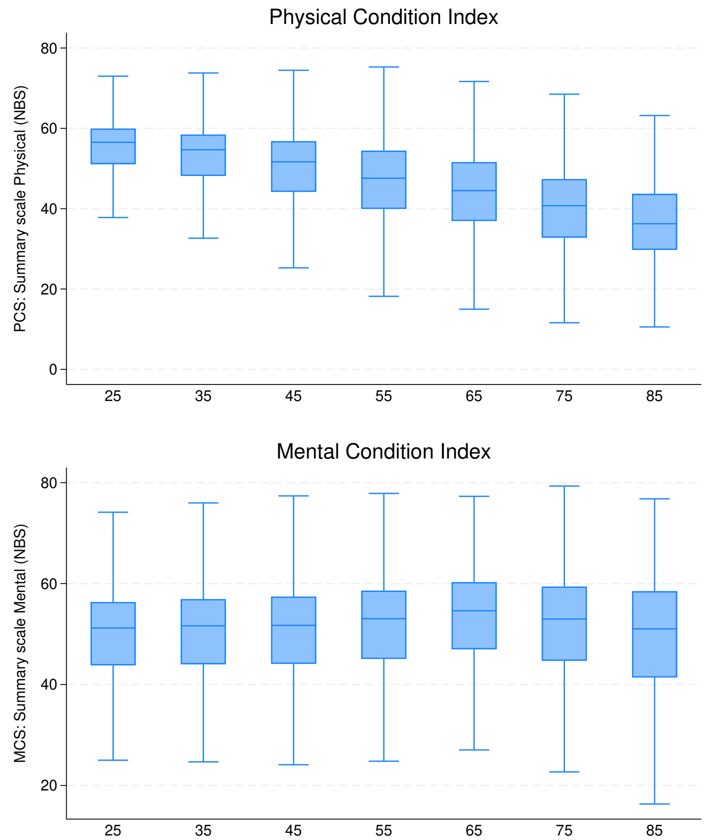
We compare our binary health measure to two alternative measures of health. First, beginning in 2002, the SOEP includes a series of questions on the health-related conditions of the respondents, which are repeated every second year. These are designed to mirror the second version of the 12-item Short Form Health Survey (SF-12 v2) questionnaire. The purpose of these questions is to provide generic indicators of perceived physical and mental health, called Physical and Mental Component Summary scores (PCS and MCS, respectively). For example, they ask about difficulty getting dressed, climbing stairs, or feeling alone. The scores are transformed into a 0-100 range and standardized to have a mean of 50 and standard deviation of 10. Figure 1.B.1 displays box plots of the evolution of these indicators by 10-year age group.

Second, we construct a *frailty* index of individuals' health history as in Hosseini et al. (2022). Beginning in 2011, the SOEP added questions regarding the diagnosis of specific health conditions by doctors, ranging from diabetes and asthma to depression and anxiety. We construct the index by adding a 1 whenever an individual has been diagnosed with one of these illnesses. Thus, the higher the frailty, the worse the health. The resulting average frailty by 10-year age groups is depicted in Figure 1.B.2.

Table 1.B.1 summarizes the correlation between our preferred binary health measure and these alternative, possibly more objective, health measures, as well as with the original 5-point self-reported health scale.[54] As expected, binary health is negatively correlated with frailty and positively correlated with the physical and mental health summary score (though the correlation with the mental health score is rather weak). Moreover, the correlations of the original 5-point self-reported health scale with these measures are only slightly higher than with the aggregated binary health measure, which suggests that we do not lose much

[54] All measures have been standardized. Note that PCS and MCS scores are orthogonal to each other by construction.

Figure 1.B.1: Physical and Mental Health Summary Scores over the Life Cycle



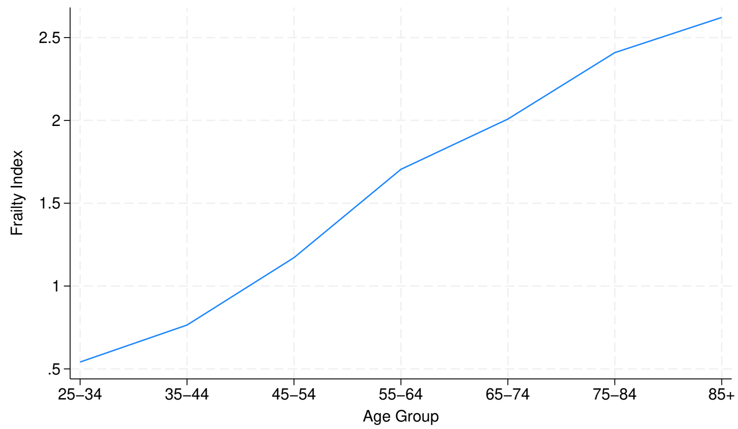
Notes: Box plots of Physical and Mental Health Summary Scores over 10-year age groups in the SOEP. The scores are calculated based on the SF-12 v2 series of questions on health-related quality of life. They are normalized to a mean of 50 and a standard deviation of 10 for 2004. A higher score indicates better health.

variation by focusing on the latter.

Table 1.B.1: Correlations across Different Health Measures

	Binary Health	5-point SRHS	Frailty	PCS	MCS
Binary Health	1	0.77	-0.41	0.62	0.26
5-point SRHS		1	-0.50	0.76	0.29
Frailty			1	-0.55	-0.16
PCS				1	-0.02
MCS					1

Figure 1.B.2: Evolution of Frailty over the Life Cycle



Notes: Average frailty by 10-year age group. The frailty index is calculated by adding a 1 each time an individual is diagnosed with a specific health condition (Hosseini et al., 2022).

1.C Construction of Health Effort

We use information on three individual health-related behaviors in constructing our health effort measure, following Cole et al. (2019). First, the frequency of practicing a sport or exercising is given by never or almost never, several times a year, at least once a month, and at least once a week. Second, survey respondents are asked how strongly they take health considerations into account in their nutrition. The answers range from very strongly to not at all.[55] Third, we use information on the number of cigarettes smoked in a day, which we cap at 50 as in Cole et al. (2019). We standardize each measure to have mean zero and standard deviation one (Kling et al., 2007) and use the negative of cigarettes smoked as a measure of healthy behaviors. The correlation of the three behaviors is reported in Table 1.C.1.

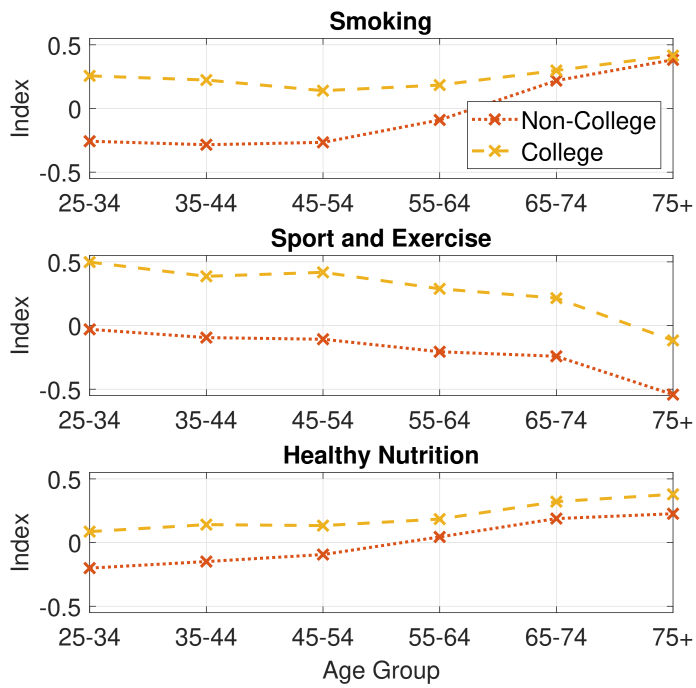
Table 1.C.1: Health Effort Components and Weights

Health Behavior	Physical Exercise	Healthy Nutrition	Abstention from Smoking	Loading
Physical Exercise	1	0.17	0.15	0.5918
Healthy Nutrition		1	0.21	0.5865
Abstention from Smoking			1	0.5530

All of these behaviors are likely also correlated with other observable characteristics. For example, Figure 1.C.1 shows the average evolution over age of the three components of

[55] Information about amounts and frequencies of alcohol consumption are only infrequently included in our data, which is why we rely on more general health-conscious nutrition.

Figure 1.C.1: Evolution of Each Standardized Lifestyle Behavior



Notes: Average of each standardized component of health effort by 10-year age group: Abstention from smoking, sport or exercise, and health-conscious nutrition.

health effort, separately for the college and non-college educated. While smoking becomes less frequent with age, and nutrition becomes healthier, physical exercise declines. For each component, a clear positive educational gradient is observed. Similarly, each behavior, in particular the frequency of sports and exercises, is positively correlated with wealth. Given that the weight on each behavior should reflect its relative importance in explaining lifestyle variations net of potentially confounding factors, we purge each behavior from variation coming from such factors by regressing them on age, age squared, years of schooling, marital status, work status, insurance type, labor income, and wealth.

Using the residualized effort measures, we perform a principal component analysis, where we take as the first principal component the measure that most closely resembles the notion of individual lifestyle behaviors. The first principal component explains around 45% of all variance in the residualized physical exercise, nutrition, and abstention from smoking. We then calculate the weights as the relative loadings of each behavior, which are relatively equal as summarized in the last column of Table 1.C.1. Finally, we normalize the aggregated effort variable to be in the unit interval.

1.D The Effects of Health on Employment and Labor Income

In our baseline model in the main text, we introduce a productivity (wage) penalty and differences in disutility of work for unhealthy individuals. In this section, we provide empirical evidence that supports our modeling approach. Specifically, we estimate how contemporaneous health affects the probability of working, as well as labor income and hours worked conditional on working, using the SOEP data and the following model:

$$y_{i,t} = \alpha_h Health_{i,t} + \delta_1 y_{i,t-1} + \delta_2 y_{i,t-2} + \gamma \mathbf{X_{i,t}} + \gamma_i + u_{i,t}, \tag{1.24}$$

where $y_{i,t}$ denotes either a dummy that equals 1 if individual i is working at time t and 0 otherwise, log labor income conditional on employment, or log hours worked conditional on employment. $\mathbf{X_{i,t}}$ includes a constant, age, age^2, marital status, type of health insurance (private or public), survey year, the number of children in the household, and dummies for the occupation in case of work. We also include individual fixed effects γ_i. We are interested in α_h, the contemporaneous effect of health on wage or hours worked.[56] In estimating such an effect, one concern might be simultaneity bias, which arises if labor income or hours worked themselves affect health status. We consequently instrument health status in year t by the number of doctor visits and the nights spent in the hospital in that same year. Given generous health insurance coverage benefits and sick-day regulations in Germany, the effect of the number of doctor visits or nights spent in the hospital on earnings and hours should work largely through health.

The results of estimating (1.24) using GMM are reported in Table 1.D.1. Column (i) gives the estimated effect of health in year t on the probability that individual i works in the same year, estimated across the whole population. Going from being unhealthy to healthy increases this probability by an estimated 15.2%, even conditional on employment in the past two periods. We find a similar role of health in affecting labor supply along the extensive margin as that observed in other countries.

Columns (ii) and (iii) report the effect of being healthy on income and hours worked, restricting the sample to those working in t. Good health increases labor income conditional on working by around 7%. The majority of this increase is due to longer working hours, which increase by over 6%. This suggests that, even conditional on working, healthy individuals

[56] It would also be reasonable to assume that health has only lagged effects on labor income and supply. Moreover, we could also highlight heterogeneous effects of health on particular demographic subgroups, as in Hosseini et al. (2021). However, our goal here is simply to quantify the contemporaneous effects of health on labor market outcomes, net of other confounding effects.

Table 1.D.1: Effect of Health on Work Status, Labor Income, and Hours Worked

	(i) $work_{i,t}$	(ii) $\log(income_{i,t} \mid work_{i,t} = 1)$	(iii) $\log(hours_{i,t} \mid work_{i,t} = 1)$
$Health_{i,t}$	0.152	0.072	0.068
	(0.016)	(0.017)	(0.017)
N	104,085	61,185	61,185

Notes: Estimated coefficient $\hat{\alpha}_h$ from equation (1.24). $Health_{i,t}$ is instrumented by number of doctors visits and nights spent in the hospital in t. Column (i) reports results from the estimation on the whole sample of 25-64 year-olds, column (ii) and (iii) only on the sample of employed individuals. First-stage tests confirm relevance assumption of these instruments.

increase their labor supply, possibly through switching from part-time to full-time work. The results furthermore indicate that good health could be accompanied by an increase in productivity that manifests in higher wages per hour, and thus larger labor income gains from being healthy.

1.E Details on the Estimation of Standard Errors

We estimate 42 parameters Θ_0 to match 64 empirical moments $\hat{\Delta}$ using the method of simulated moments. To conduct standard inference on our estimates using this estimator, we would need know a consistent estimate of the full variance-covariance matrix of the empirical moments $\hat{V}$. Alternatively, a bootstrap method can be used to construct standard error estimates. In our case both of these options are infeasible. While most of our empirical moments are computed from the SOEP data, they often use specific subsets of the data. In particular, wealth information is only available every 5 years. On top of that, the estimate for the values of a statistical life year (VSL) are taken from a meta-analysis of VSL estimates in OECD countries (OECD, 2012), which prevents us from computing the correlation between the elements of $\hat{\Delta}$. Moreover, the application of a bootstrap method would be computationally expensive given that our parameter and moment space is relatively large.

For that reason, we use the strategy of Cocci and Plagborg-Møller (2021), who show that the standard errors of the method of moment estimates $\hat{\Theta}$ can be bounded when assuming that the elements of $\hat{\Delta}$ are perfectly correlated with each other. They are computed as the weighted sum of the standard errors of individual empirical moments. They show that these worst-case standard errors can further be minimized for over-identified models by selecting only those moments which are most-informative about the parameter at question. To construct the weights, we compute the Jacobian matrix that contains the derivatives of the model-implied moments with respect to the standard errors using first differences. The main assumption behind this method is a joint normality assumption of all empirical moments. We view this as reasonable in our context as all moments with the exception come from the same data set.

The algorithm to compute the efficient worst-case standard error for each component of $\hat{\Theta}$ then comprises the following steps (see Cocci and Plagborg-Møller (2021), page 11-12): First, we construct an efficient estimator $\hat{\Theta}$ using the weight matrix that has the inverse of each empirical moment's standard error on its diagonal, and zeros on the off-diagonals. Next, we construct the Jacobian matrix using first differences. Finally, we solve the median regression (eq. 6 in Cocci and Plagborg-Møller (2021)) that allows us to perform the efficient moment selection procedure for each parameter, which yields the standard error estimates as reported in Table 1.2.

1.F Further Details on Structural Model Estimation

Classification of Fixed Health Types

As explained in Section 1.4, the first step of estimating the probability of being in good health in the next period involves the classification of individuals in our data into fixed unobservable health type groups η using the *kmeans* algorithm. We construct the data moments used for the classification in the following way: First, we take all direct measures of health and health-related status that are available in our data for at least half of the sample period. These are (i) the number of annual doctor visits, (ii) self-rated health status on a 5-point scale, (iii) inpatient nights in a hospital, (iv) and (v) the Physical and Mental Component Summary scores (see Appendix 1.B), and (vi) the body-mass index.[57]

Second, we residualize these variables against age, age squared, a college education dummy, gender, health insurance type status, and cohort dummies. We do this because the individual health type should be informative about variation in health and health-related status *net of* variation that arises from other time-constant observable characteristics. Moreover, we strip the health moments from variation coming from mere satisfaction with own health (on a 10-point scale). This is to make sure that the classification into unobserved health types is based on fundamental factors that are no changed as a result from noisy reporting and measurement issues. Third we standardize the resulting residuals to give every variable the chance to be equally important for the health type classification. Since the health type is fixed over time, we take one average standardized residual per individual.

[57] We experimented with including individual fixed effects from a regression of future health on current and past health, effort and age as additional moments. However, this restricted our sample too much.

Table 1.E.1: Empirical Moments and Standard Errors

Description	**Value**	**S.E.**	**Description**	**Value**	**S.E.**
Employment Share	0.651	0.002	Median Wealth	0.062	0.003
among healthy	0.766	0.002	divided by average	0.516	0.015
by 10-year age group	0.823	0.002	2-year labor income	1.166	0.024
	0.619	0.002	by 10-year age group	1.651	0.037
Employment Share	0.506	0.008		1.567	0.043
among unhealthy	0.583	0.005		1.006	0.047
by 10-year age group	0.601	0.005	Education Gradient in Employment	1.237	0.003
	0.409	0.005	Non-Adjuster Shares	0.267	0.004
Average Effort among	0.678	0.002	by Long Age Group	0.328	0.003
non-college and healthy	0.677	0.002		0.404	0.004
by 10-year age group	0.680	0.002	VSL multiple	8.493	0.595
	0.699	0.002	Standard Deviation of Effort	0.161	0.000
	0.730	0.002	Consumption Ratio of Healthy/Unhealthy	1.163	0.022
	0.724	0.002	Average Labor Income	35.393	0.196
Average Effort among	0.643	0.007	in Ths for non-college	49.379	0.232
non-college and unhealthy	0.623	0.005	and healthy by 10-year age group	55.955	0.266
by 10-year age group	0.627	0.004		42.219	0.353
	0.655	0.003	Average Labor Income	24.948	0.563
	0.697	0.003	in Ths for non-college	33.166	0.519
	0.692	0.003	and unhealthy by 10-year age group	36.691	0.499
Average Effort among	0.779	0.002		25.311	0.499
college and healthy	0.770	0.002	Average Labor Income	59.483	0.488
by 10-year age group	0.766	0.002	in Ths for college	89.538	0.632
	0.763	0.002	and healthy by 10-year age group	107.928	0.761
	0.779	0.002		98.277	1.108
	0.769	0.004	Average Labor Income	50.388	1.849
Average Effort among	0.752	0.011	in Ths for college	66.253	1.656
college and unhealthy	0.744	0.008	and unhealthy by 10-year age group	78.318	1.688
by 10-year age group	0.737	0.006		63.133	1.786
	0.738	0.005	Variance of Log Labor Income	0.595	0.002
	0.751	0.005	Pension Replacement Rate	0.477	0.002
	0.734	0.006	Wealth Gini Coefficient	0.746	0.004

The fourth step comprises the clustering of individuals using the *kmeans* algorithm that assigns observations to the cluster with the smallest Euclidean distance. We repeat the clustering for randomly chosen initial group centers and for up to 5 clusters. We then calculate the within-cluster sum of squares for each cluster number. Our goal in selecting the number of clusters is to have intra-cluster variation that is as small as possible while maintaining computational feasibility in our model. Since the within-cluster sum of squares display a kink ("elbow") after 2 clusters, we opt to select two clusters.

Estimation of Wages and Productivity

Our estimation of the distribution of fixed productivity types and the persistence and variance of idiosyncratic shocks involves the following steps. First, we compute real hourly wages x_{ij} for individual i with age j in our data on the sample of workers that work for at least two consecutive years. We then recover combined residuals and individual fixed effects estimates from a regression of log wages on the full set of age and health dummies (D_{it}^{age} and D_{it}^{health}, respectively) according to:

$$\ln x_{ij} = \sum_{t=25}^{63} \sum_{h=\{0,1\}} d_t^h \times D_{it}^{age} \times D_{it}^{health} + \theta_i + u_{ij}, \tag{1.25}$$

as in De Nardi et al. (2023); French (2005). Here, the coefficients d_t^h capture the effect of the interaction of dummy variables for age and health status and θ_i captures unobserved fixed labor productivity. While we treat this fixed productivity continuous in the estimation, we follow Low and Pistaferri (2015) in assuming discrete productivity "types" in the model as detailed in Section 1.4.2.

Next, we regress the combined estimated (predicted) residuals ($\hat{\theta}_i+\hat{u}_{ij}$) on cohort dummies and education to strip them from variation coming from these sources that we capture through $\lambda_j(h_j, e)$. We then estimate the parameters of the idiosyncratic components using a standard generalized method of moments (GMM) procedure that minimizes the distance between the empirical age-profile of the variances of the combined residuals and the population analogue following Storesletten et al. (2004).[58] We obtain the estimated persistence of idiosyncratic productivity shocks $\rho = 0.975$.

1.G Discussion of Estimated Health Technology Parameters

Table 1.G.1 shows the results of estimation of (1.16) along with the estimates of the exogenous health model. All estimates are statistically significant at the 95% level. Table 1.G.2 reports average marginal effects calculated from the estimated parameters for the baseline model.

The estimates from the columns for the baseline model with endogenous health imply that the probability of being healthy in the next period, conditional on effort, current health, education and health type, decreases monotonically over age. Individuals with the high health

[58] Concretely, to distinguish the variance of the fixed effect from the variance of transitory shock, we again follow Storesletten et al. (2004) and references therein by computing the sum of three consecutive residuals for 25-year olds.

Table 1.G.1: Logit Estimation of Probability of being Healthy in 2 years

	Model:	**Endogenous Health**		**Exogenous Health**	
Variable	**Coef.**	**Estimate**	**Std.Error**	**Estimate**	**Std.Error**
Current Health Effort	λ_1	0.693	0.138		
Past Health Effort	λ_2	0.734	0.137		
Current Health	$h_t = 1$	2.311	0.029	2.340	0.029
Age Group Dummies					
35		-0.289	0.079	-0.301	0.078
45		-0.644	0.074	-0.655	0.074
55		-0.881	0.074	-0.871	0.074
65		-1.138	0.074	-1.074	0.073
75		-1.586	0.077	-1.527	0.077
Health Type	$\eta = 1$	0.632	0.028	0.654	0.028
College	$e = 1$	0.238	0.033	0.388	0.032
Constant		-0.906	0.095	0.013	0.072
Pseudo R^2		0.242		0.237	

Notes: $N = 43,336$. Standard Errors are heteroscedasticity robust.

type consistently have, ceteris paribus, a larger probability of being healthy than those with the low health type. The same, albeit to a smaller degree, is true for agents with college rather than non-college education. However, the largest differences in the probability of being healthy conditional on all other covariates, arise between individuals who are currently unhealthy and individuals who are currently healthy. For example, a healthy 75-year-old college-educated individual of the high health type has a 67% probability of being healthy in two years absent any effort (past and present) if she is currently healthy, while this probability is only 16% if she is currently unhealthy.

Much research, primarily medical, has aimed to causally identify the effect of different lifestyle components on good future health. For example, Lee (2003) review data from 50 epidemiological studies on the relationship between physical activity and cancer incidence. Similarly, Colman and Dave (2013) analyze the connection between physical activity and the prevalence of hypertension, diabetes, and heart disease. Other papers, such as those by LaCroix et al. (1991) and Van Oyen et al. (2014) highlight the impact of smoking on mortality and disability. More recently, Cena and Calder (2020) review evidence on the health-promoting effects of more plant-based diets. Generally speaking, there is a strong consensus in this literature on the beneficial effects of healthy lifestyle behaviors, such as

Table 1.G.2: Average Marginal Effects from Health Technology Estimates

	Low Health Type ($\eta = 0$)											
	No College ($e = 0$)						*College* ($e = 1$)					
	Unhealthy ($h_t = 0$)			Healthy ($h_t = 1$)			Unhealthy ($h_t = 0$)			Healthy ($h_t = 1$)		
Age	π^0	λ_1	λ_2	π^0	λ_1	λ_2	π^0	λ_1	λ_2	π^0	λ_1	λ_2
25-34	0.29	0.17	0.18	0.80	0.05	0.06	0.34	0.17	0.18	0.84	0.04	0.05
35-44	0.23	0.17	0.18	0.75	0.07	0.07	0.28	0.17	0.18	0.79	0.05	0.06
45-54	0.18	0.16	0.17	0.68	0.09	0.09	0.21	0.17	0.18	0.73	0.07	0.08
55-65	0.14	0.15	0.16	0.63	0.10	0.11	0.18	0.16	0.17	0.68	0.09	0.09
65-74	0.11	0.13	0.14	0.57	0.12	0.12	0.14	0.15	0.16	0.62	0.10	0.11
75+	0.08	0.11	0.11	0.45	0.15	0.15	0.10	0.12	0.13	0.51	0.13	0.14
	High Health Type ($\eta = 1$)											
	No College ($e = 0$)						*College* ($e = 1$)					
	Unhealthy ($h_t = 0$)			Healthy ($h_t = 1$)			Unhealthy ($h_t = 0$)			Healthy ($h_t = 1$)		
Age	π^0	λ_1	λ_2	π^0	λ_1	λ_2	π^0	λ_1	λ_2	π^0	λ_1	λ_2
25-34	0.43	0.15	0.16	0.88	0.03	0.03	0.49	0.14	0.15	0.91	0.02	0.03
35-44	0.36	0.16	0.17	0.85	0.04	0.04	0.42	0.15	0.16	0.88	0.03	0.03
45-54	0.29	0.17	0.18	0.80	0.05	0.06	0.34	0.17	0.18	0.84	0.04	0.05
55-65	0.24	0.17	0.18	0.76	0.06	0.07	0.29	0.17	0.18	0.80	0.05	0.06
65-74	0.20	0.17	0.17	0.71	0.08	0.08	0.24	0.17	0.18	0.76	0.07	0.07
75+	0.13	0.15	0.115	0.61	0.11	0.11	0.16	0.16	0.17	0.67	0.09	0.10

physical activity, a healthy diet, and abstention from smoking, on morbidity and mortality. However, since these studies typically focus on the effect of a specific lifestyle behavior on the onset of a specific disease, such as hypertension or diabetes, it is not possible to directly compare their estimates with our health transition technology parameters, which are estimated based on self-reported health status.

To facilitate a meaningful comparison, we accordingly employ three approaches. First, similar to Cole et al. (2019), we use the SOEP data to map health status to the prevalence of a specific health condition, conditional on age group and education (see Table 1.G.3). We use this information to construct the probability of the onset of a specific disease in the future, conditional on current health status, age group, fixed health type, as well as current and past health effort, which is implied by our estimated health technology parameters using the formula:

$$\begin{aligned} Pr(disease_{j+1}|h_j, f_j, f_{j-1}, e, \eta) =& \pi_j(h_{j+1} = 1|h_j, f_j, f_{j-1}, e, \eta) \times Pr(disease|h_{j+1} = 1, e) \\ &+ (1 - \pi_j(h_{j+1} = 1|h_t, f_t, f_{j-1}, e, \eta)) \times Pr(disease|h_{j+1} = 0, e) \end{aligned}$$

Table 1.G.3: Prevalence of Diseases in Population by Age Group and Health Status

		Health Condition Prevalence by Education							
		No CL	CL	No CL	CL	No CL	CL	No CL	CL
Age	Health	Diabetes		Cancer		Hypertension		Heart Condition	
25-34	Unhealthy	0.038	0.000	0.015	0.006	0.111	0.073	0.029	0.011
	Healthy	0.007	0.006	0.006	0.005	0.042	0.028	0.013	0.011
35-44	Unhealthy	0.055	0.034	0.035	0.029	0.201	0.118	0.062	0.044
	Healthy	0.018	0.011	0.015	0.011	0.104	0.067	0.015	0.012
45-54	Unhealthy	0.116	0.064	0.074	0.075	0.327	0.286	0.118	0.084
	Healthy	0.039	0.022	0.025	0.030	0.201	0.162	0.032	0.019
55-64	Unhealthy	0.200	0.177	0.094	0.113	0.525	0.462	0.213	0.172
	Healthy	0.089	0.063	0.051	0.047	0.342	0.328	0.075	0.058
65-74	Unhealthy	0.263	0.243	0.164	0.179	0.575	0.593	0.348	0.347
	Healthy	0.147	0.123	0.084	0.104	0.456	0.423	0.149	0.150
75+	Unhealthy	0.262	0.251	0.138	0.221	0.583	0.621	0.460	0.491
	Healthy	0.179	0.171	0.102	0.135	0.490	0.508	0.248	0.276

Finally, we average this implied probability of having a specific disease over individuals in the top, middle, and bottom terciles of the current health effort distribution and/or the past effort distribution, conditional on age group, current health and education but averaging over health type. To be comparable to Cole et al. (2019), we use only individuals between the age of 25 and 75. We then calculate the average percent deviation of the implied disease probabilities in each effort tercile relative to their within-status mean and compare the results to those in Colman and Dave (2013).

Table 1.G.4 shows the results. Overall, the effectiveness of health efforts in reducing the probability of disease onset implied by our estimated health technology parameters seems lower than that reported in Colman and Dave (2013) for the case of exercise. For example, while they find that exercise can reduce the prevalence of heart conditions by between 23-29%, our estimates imply that being in the top effort tercile for current and past health effort lessens the prevalence of heart conditions by around 5% compared to the mean.

Yet, the disadvantage of this approach is that it focuses on just one specific component of our compound health effort measure, namely exercise. We consequently implement a second approach, again in an effort to gauge our estimated health technology parameters against the literature, this time using a mapping between health status and survival in old age to benchmark our estimates against the results found in Knoops et al. (2004). Their study not only explores the effect of a comprehensive lifestyle measure, comprised of a Mediterranean diet, moderate alcohol use, physical activity, and nonsmoking, but also uses data on European men and women between ages 70 and 90 and is thus closer to our German data source.

Table 1.G.4: Implied Probability of Disease by Past and Current Effort Tercile

	Percent Change of Probability relative to the within-status Mean			
Effort Tercile	Diabetes	Cancer	Hypertension	Heart Condition
Current Effort				
Low	3.52	2.85	1.52	4.05
Middle	-0.52	-0.43	-0.21	-0.61
High	-3.35	-2.72	-1.45	-3.86
Past Effort				
Low	2.11	1.74	0.88	2.52
Middle	-0.26	-0.22	-0.10	-0.33
High	-2.12	-1.73	-0.90	-2.51
Both				
Low	4.26	3.5	1.81	5.06
Middle	-0.76	-0.62	-0.31	-0.923
High	-4.11	-3.36	-1.75	-4.87
Coleman & Dave	1.2-3% decrease		10-31% decrease	23-29% decrease

To compare their estimate of the impact of healthy lifestyles on mortality, we simulate the random health and survival evolution of 100,000 individuals between the ages of 70 and 84 that are equipped with our estimated health transition technology, as specified in Section 1.4.2.[59] As Table 1.G.5 summarizes, our parameter estimates paired with the empirical average lifestyle effort results in a 10-year mortality rate around 42% percent, which is slightly above the rate reported in Knoops et al. (2004). When restricting everyone to have a healthy lifestyle, which we assume to be the effort at the 90th percentile by age, the simulation-implied mortality rate drops to 40.6%. This drop is slightly smaller, yet comparable to that found in Knoops et al. (2004). Vice versa, if we assume everyone exerts efforts equal to the 10th percentile, mortality over 10 years is increased by half a percentage point. We take this as confirmation that our estimated health technology parameters, and especially the effectiveness of health efforts, are conservative but reasonable in light of the empirical medical literature.

Finally, several papers investigate the causal effect of compound measures of healthy lifestyles on specific disease prevalence. For example, Schlesinger et al. (2020) find, in a meta-analysis of the literature, that adherence to healthy lifestyle behaviors (i.e., a favourable diet, physical activity, nonsmoking, moderate alcohol intake, and normal weight) lowers the risk

[59] We choose 84 instead of 90 to have ample sample size to measure 10-year mortatiliy. We assume that initial age is drawn uniformly between 70 and 84.

Table 1.G.5: Mortality among Older-Age Individuals implied by Our Estimates

	Mortality Rates over 10 years (%)	
	Knoops et al.	Implied by Simulation
Average Lifestyle	39.9	42.3
Healthy Lifestyle	35	40.6
Unhealthy Lifestyle		42.8

of type 2 diabetes by almost 80%, which qualifies the numbers found in column 1 in Table 1.G.4. Similarly, Barbaresko et al. (2018) survey 22 research papers that analyze the effect of adhering to a healthy lifestyle on the onset of various serious conditions, and find a reduced risk of 66% for cardiovascular disease, 60% for stroke, and 69% for heart failure.

1.H Sources of Lifetime Inequality

To get a sense of the importance of initial conditions in shaping inequality in lifetime outcomes, we follow the strategy in Huggett et al. (2011) and calculate the share of (the present value of) lifetime earnings, of the variance in the wealth at retirement ages, of the number of healthy years, and of the the share of healthy years to overall life years that can be explained by variation in the individual states at age 25. Specifically, following Huggett et al. (2011), we compare the conditional variance in these outcomes, where we condition on all individual state variables at age 25, with the unconditional variance. The state variables are education, discount factor type, productivity type, and health type, as well as initial health and initial health effort habits. For the latter, we group individuals into three equally sized groups reflecting their initial health effort habits. If a significant share of wealth and health inequality can be explained by initial conditions, the positive association between wealth and health is more likely to be predetermined at age 25. On the other hand, if the explained share is small, this points to the significance of luck in terms of economic but also health shock realizations during life in determining inequalities.

Table 1.H.1 summarizes the results. We find that around 81% of the variation in lifetime earnings in our model is accounted for by differences in the initial conditions individuals face at age 25, similar to the 62% that Huggett et al. (2011) find for this outcome in the U.S. The corresponding statistic for wealth at the retirement age (i.e., age 65-66) is lower but is quite large at 53%. By contrast, the differences in initial conditions explain much smaller fractions of the variations in healthy years (25%), and the share of healthy years in life (37%), implying that events over the lifetime largely drive the health-related outcomes. Overall, our

Table 1.H.1: Contribution of Initial Conditions at Age 25 to Lifetime Inequality

Statistic	Model
Fraction of variance in lifetime earnings	81.3%
Fraction of variance in wealth at age 65-66	53.0%
Fraction of variance in healthy years	24.9%
Fraction of variance in the share of healthy years in life	36.5%

results indicate the the role of both initial conditions and lifecycle events (and choices made by agents) in accounting for health and wealth inequality over the lifecycle.

1.I Details about the Conceptual Two-Period Model

We presented a simple two-period model with endogenous health and wealth accumulations in Section 1.5.1 to build insights on key channels. Here we provide more details such as a full set of assumptions, derivations for the optimality conditions, and further results with different assumptions.

In addition to the key assumptions laid out in Section 1.5.1, we further assume that utility is positive ($u_t > 0$ for $t = 0, 1$) and that the survival probability is positive ($S(h_1) > 0$). For simplicity, we assume zero interest rate, which is not important for our results. Current health (h_0) is assumed to be a state variable, and future health (h_1) can be shaped by the effort choice through $\pi(f)$. Having endogeneity of current health is feasible, yet complicates the analytic results. Similarly, we abstract from several mechanisms that are present in our quantitative life-cycle model to focus on illustrating our key channels of interest. These include the effect of current health on effort cost disutility, the effect of current health on current consumption utility and the effect of current health on future health. We provide implications of incorporating these extra effects below.

We can rewrite the constrained optimization problem (1.20) as

$$\max_{c_0,f,n} \{u_0(c_0) - \varphi(f) - \phi(n, h_0) + \beta S(\pi(f))u_1(w(h_0)n - c_0, \pi(f))\} \tag{1.26}$$

which yields the following first-order conditions:

$$[c_0] \quad : \quad u_0'(c_0) = \beta S(h_1)\frac{\partial u_1(c_1, h_1)}{\partial c_1} \tag{1.27}$$

$$[n] \quad : \quad \frac{\partial \phi(n, h_0)}{\partial n} = \beta S(h_1)\frac{\partial u_1(c_1, h_1)}{\partial c_1}w(h_0) \tag{1.28}$$

$$[f] \quad : \quad \varphi'(f) = \beta S'(\pi(f))\pi'(f)u_1(c_1, \pi(f)). \tag{1.29}$$

The first equation 1.27 describes the optimal savings choice, as discussed in Section 1.5.1. As noted earlier, one could consider the health-dependence on u_0 as well. Then, the condition would read

$$\frac{\partial u_0(c_0, h_0)}{\partial c_0} = \beta S(h_1)\frac{\partial u_1(c_1, h_1)}{\partial c_1}. \tag{1.30}$$

Therefore, if we consider a health penalty in the form of a multiplicative constant $\kappa(h)$, one can see that the relative health status would shape the strength of the savings motive. For example, if $h_0 < h_1$, then it could reinforce the savings motive. On the other hand, for those with $h_0 = h_1$, the savings channel we discussed in Section 1.5.1 would only work through the length channel (i.e., $S(h_1)$).

Combining equations (1.27) and (1.28), we can obtain:

$$\frac{\partial \phi(n, h_0)}{\partial n} = u_0'(c_0)w(h_0), \tag{1.31}$$

which is the labor-leisure condition (1.21). As is standard in any labor-leisure condition, the effects of higher wages due to health on labor supply depends on whether the substitution effect is stronger than the income effect, which is shaped by the functional form on utility. In practice, it would also matter if the wage decline is temporary or not, since a temporary change would induce a stronger positive effect on labor supply than a permanent change.

Finally, (1.29) describes the optimality condition for the effort choice. As in the labor disutility, one could potentially introduce health-dependence on the disutility of efforts. The implication is going to be parallel: poor health would shift the left-hand side up, which would increase the marginal cost of efforts.

Moreover, we note that if we assume that health for the working period (i.e., h_0) can also be endogenously affected by the effort choice, the right-hand side would additionally include:

$$\beta S(\pi(f))\frac{\partial u_1(c_1, \pi(f))}{\partial c_1}w'(\pi(f))\pi'(f), \tag{1.32}$$

which captures an effect coming through higher expected future wages when healthy. Interestingly, this motive can be decreasing in wealth, as it is weighed by the marginal utility of future consumption, which decreases with wealth. In other words, the motive to exert efforts to be healthy in the future and therefore be more productive, is weaker with rising income, which we can interpret as an income effect of effort. This force would mitigate the earnings channel in generating wealth-health gaps.

Table 1.J.1: Contributions to Wealth-Health Gaps of the Baseline Model

	Earnings Channel								
	Total			Wage Loss Only			Disutil. Only		
Wealth	25th	50th	75th	25th	50th	75th	25th	50th	75th
Age Group									
35-44	100%	28%	21%	21%	0%	15%	-1%	11%	1%
45-54	15%	6%	15%	14%	5%	8%	-4%	0%	1%
55-64	34%	23%	6%	17%	12%	6%	5%	7%	1%
65-74	8%	19%	12%	7%	10%	11%	0%	7%	3%
	Savings Channel								
	Total			Length Only			Quality Only		
Wealth	25th	50th	75th	25th	50th	75th	25th	50th	75th
Age Group									
35-44	4%	1%	28%	-2%	0%	16%	-4%	-7%	11%
45-54	48%	42%	50%	16%	18%	36%	12%	2%	7%
55-64	55%	52%	69%	30%	19%	37%	17%	7%	9%
65-74	56%	51%	55%	32%	33%	40%	8%	5%	7%

Notes: This table reports the proportions of the baseline relative wealth-health gaps explained by different components of the earnings and savings channels. See the text for their definitions.

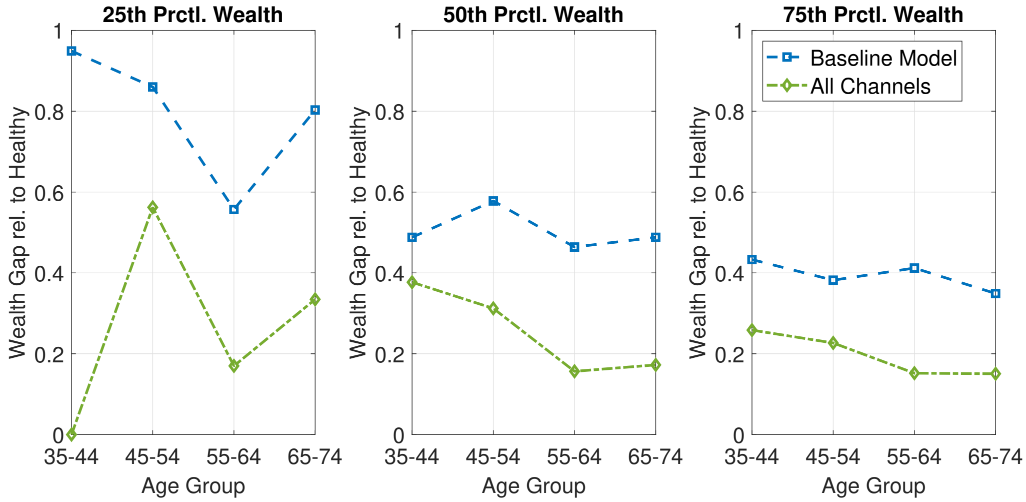
1.J Additional Quantitative Exercises

Savings and Earnings Channel

Table 1.J.1 reports the proportions of the baseline relative wealth-health gaps that are explained by different channels. With *Wage Loss Only*, we only impose $w_p^e = 0$ for both education groups. With *Disutil. Only*, we only impose that the disutility of labor supply is as if one was healthy for everyone. With *Length Only*, we only equalize the survival probability at the healthy level: $S_j(h_j = 1, e) \forall j, e)$. With *Quality Only*, we only impose that the consumption utility and value of life is not reduced from being unhealthy ($\tilde{\kappa} = 1$). In all exercises, we keep the distribution of health fixed at the baseline economy.

Figure 1.J.1 shows the results of a counterfactual experiment, in which we shut down both savings and earnings channel, and leave the distribution of health free to adjust to different health effort choices. This effectively takes away any incentive to exert efforts, as being unhealthy is no longer different from being healthy in terms of labor supply, wages, survival or consumption utility. This shrinks the wealth-health gaps considerably, by around 60%, on average. The remaining gaps in our model can be explained as individuals still differ in fixed characteristics that drive both wealth accumulation and the probability of being

Figure 1.J.1: Effect of Both Earnings and Savings Channels



Notes: Differences in the wealth levels of those being healthy and unhealthy at the 25th (left), 50th (middle), and 75th (right) percentile of the wealth distribution in the baseline model (blue) and in the counterfactual scenario when shutting down both earnings and savings channel together (green). Differences are expressed relative to the wealth levels of the healthy.

health, most notably education.

Equalizing Efforts

In this section, we first explain how to quantify the contributions of the two different (direct versus indirect) effects that we discussed in Section 1.5.2 to the wealth-health gaps of the baseline economy separately at different ages and points of the wealth distribution in Table 1.J.2.

Specifically, we quantify the contribution of *direct* effects of health effort equalization that work through the health distribution by simulating our baseline economy but, unexpectedly to the model agents, changing the health distribution to be the same as in the equal efforts counterfactual. That is, all decisions on savings, labor supply and health efforts are the same as in the baseline economy but the health evolution of every agent is as if she would have exerted the average effort level. Analogously, we quantify the contribution of the *indirect* effects of the equal efforts experiment that work through choices, by simulating the counterfactual economy, but keeping the health distribution of the baseline case. The results clearly suggest that the total effects of equalizing efforts works primarily through its direct effect on the health distribution rather than the indirect effects.

Next, in addition to equalizing efforts at all age groups, we perform a series of further counterfactual exercises, in which we separately equalize individual health efforts for the

Table 1.J.2: Contributions of Equal Efforts to Baseline Wealth-Health Gaps

	Equal Efforts								
	Total			Direct Effects Only			Indirect Effects Only		
Wealth	25th	50th	75th	25th	50th	75th	25th	50th	75th
Age Group									
35-44	-17%	3%	26%	0%	6%	24%	-2%	-4%	1%
45-54	10%	14%	20%	12%	14%	17%	-3%	-2%	2%
55-64	22%	28%	34%	21%	29%	33%	-3%	-1%	2%
65-74	28%	41%	40%	25%	40%	41%	2%	3%	2%

Notes: This table reports the proportions of the baseline relative wealth-health gaps that are explained by different effects. See the text for the definitions of direct and indirect effects.

following ages groups: 25-44-year-olds, 45-64-year-olds, and 65-and-older (i.e., retired individuals).

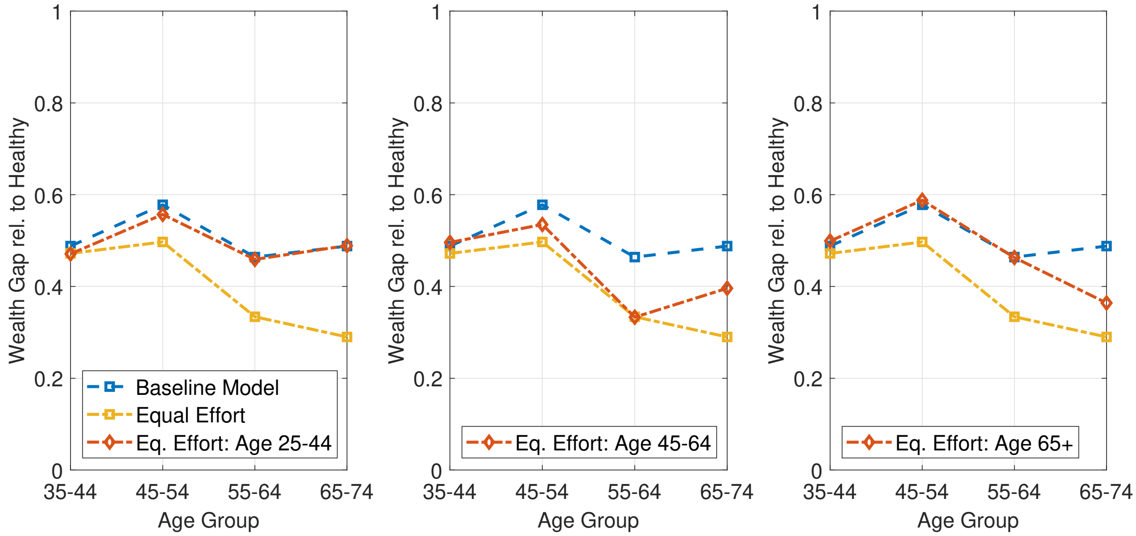
Figure 1.J.2 displays the resulting wealth-health gaps at the median for different scenarios. The left panel suggests that when equalizing health efforts among the young working-age agents only (ages 25-44), the wealth-health gaps are also slightly reduced in the 45-54-year-old age group. For older individuals, however, the gaps remain as large as in the baseline economy, meaning that eliminating effort variation early on has some moderately lasting effects in terms of closing the wealth-health gaps during the working ages. This is sensible given that the estimated adjustment costs are low when agents are young.

The lasting effect becomes more pronounced when equalizing efforts among prime-age workers (ages 45-64), who begin to face a more significant risk of becoming unhealthy. On the one hand, the gap at ages 45-54 is higher than in the counterfactual case with constant effort everywhere, as health behaviors are allowed to vary at young ages and this spills over into the age groups where efforts are held constant. On the other hand, the gap at ages 65-74 is diminished by almost 20% relative to the benchmark case even though health behaviors are allowed to vary.

1.K Additional Figures and Tables

Table 1.K.1 summarizes the initial distribution we estimate for our quantitative model. Several patterns are worth noting. Among college-educated individuals, 5% report being unhealthy between ages 25-30, while this number is over 8% among the non-college educated. Moreover, average initial health effort is almost two-thirds of a standard deviation higher for the college educated. The fixed health type is strongly correlated with initial health. Over

Figure 1.J.2: Effect of Timing of Health Efforts on Wealth-Health Gaps



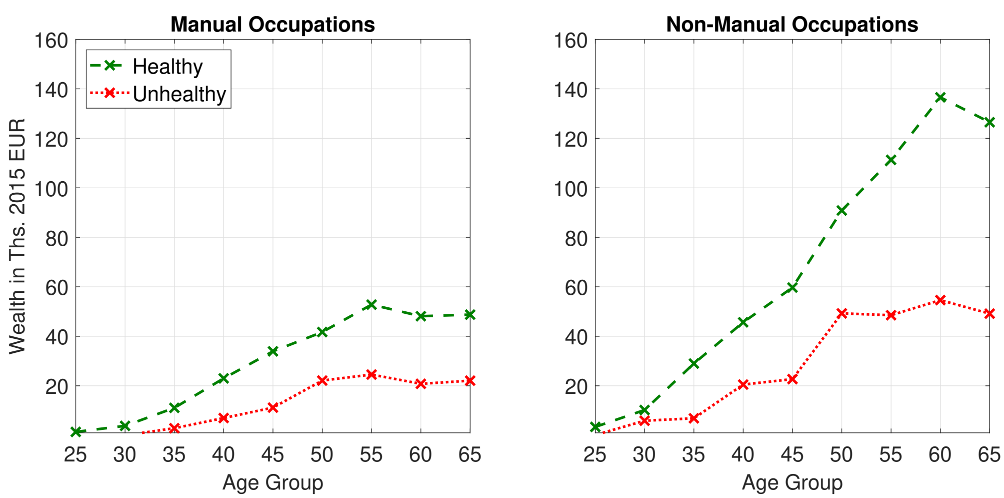
Notes: Differences in the wealth levels by health status at the median of the wealth distribution in the baseline model (blue), in the counterfactual scenario with constant health effort choices across all age groups (yellow), and in the counterfactual scenarios where health efforts are equalized separately for the 25-44-year-old (left), 45-64-year-old (middle), and 65+ (right) age groups.

11% of those with the low health type are on average unhealthy, while it is less than 6% for the high health type. In contrast, initial health effort levels differ only little across health types. Generally, differences in both initial health and initial health effort are only marginal across productivity and discount factor types in the data, which is why we do not report them here.

Table 1.K.1: Initial Distribution

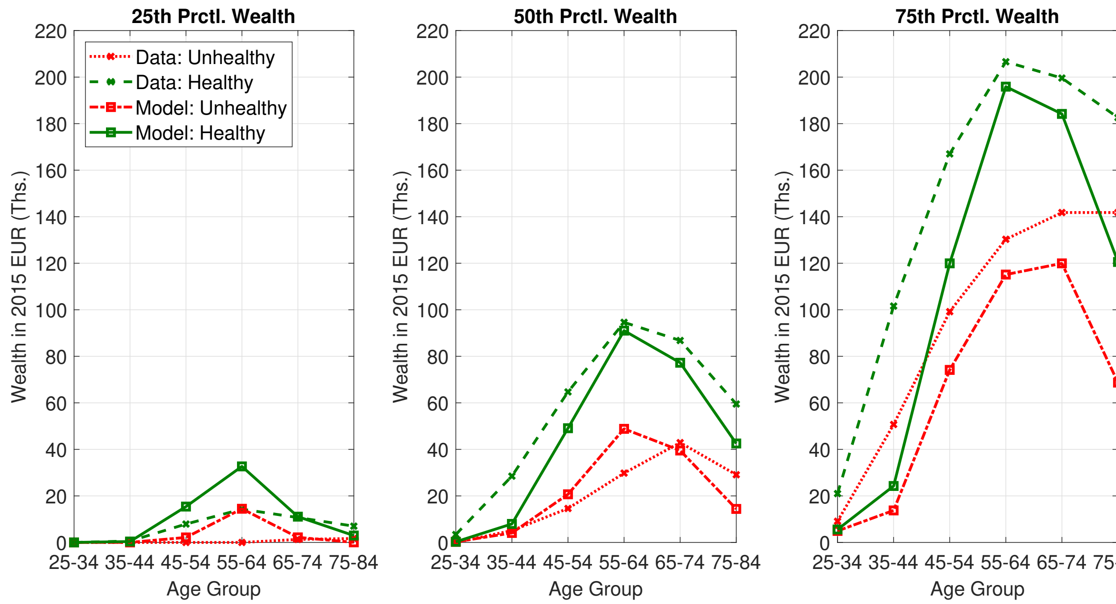
	No College ($e = 0$)							
	$\beta = \beta_l$				$\beta = \beta_h$			
$\theta =$	θ_l		θ_h		θ_l		θ_h	
$\eta =$	η_l	η_h	η_l	η_h	η_l	η_h	η_l	η_h
Prob. Mass	0.062	0.133	0.070	0.101	0.061	0.099	0.063	0.102
Avg. h	0.878	0.937	0.878	0.937	0.878	0.937	0.878	0.937
Avg. f	0.663	0.690	0.663	0.690	0.663	0.690	0.663	0.690
	College ($e = 1$)							
	$\beta = \beta_l$				$\beta = \beta_h$			
$\theta =$	θ_l		θ_h		θ_l		θ_h	
$\eta =$	η_l	η_h	η_l	η_h	η_l	η_h	η_l	η_h
Prob. Mass	0.034	0.033	0.024	0.045	0.029	0.047	0.025	0.072
Avg. h	0.926	0.960	0.926	0.960	0.926	0.960	0.926	0.960
Avg. f	0.773	0.785	0.773	0.785	0.773	0.785	0.773	0.785

Figure 1.K.1: Median Wealth Profiles by Health Status and Occupation



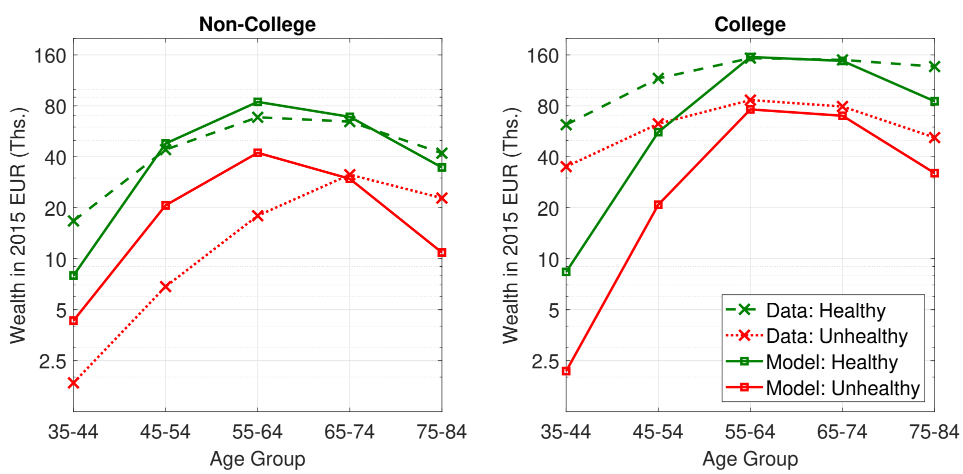
Notes: Median wealth per 5-year age group and health status for manual (left) and non-manual (right) occupations, separated by healthy (green) and unhealthy (red) status. Manual occupations include agricultural workers, craft and trades-persons, plant and machine operators, and other elementary occupations. The non-manual category includes all other occupations.

Figure 1.K.2: Wealth-Health Gaps at Different Distribution Points: Model vs. Data



Notes: Wealth by 10-year age groups, distinguishing between healthy individuals (green) and unhealthy ones (red) in the model relative to the data at different point of the wealth distribution. Left panel: 25th percentile. Central panel: 50th percentile. Right panel: 75th percentile.

Figure 1.K.3: Wealth Profiles by Health and Education: Model vs. Data



Notes: Wealth by 10-year age groups, distinguishing between healthy individuals (green) and unhealthy ones (red) in the model relative to the data. Left panel: Non-college educated individuals. Right panel: College educated individuals.

Chapter 2

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

with Suzanne Bellue[2]

Abstract: We study the long-run aggregate, distributional, and intergenerational 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 innovation in our model is the skill production technology during school years with tracking. School tracks endogenously differ in their pace of instruction and the students' average skills. We show analytically that this technology can rationalize reduced-form evidence on the effects of school tracking on the distribution of end-of-school skills. We then calibrate the model using representative data from Germany, a country with a very early school tracking policy by international standards. Our calibrated model shows that an education reform that postpones the tracking age from ten to fourteen generates improvements in intergenerational mobility but comes at the cost of modest losses in aggregate human capital and economic output, reducing aggregate welfare. This efficiency-mobility trade-off is rooted in the effects of longer comprehensive schooling on learning and depends crucially on the presence of general equilibrium effects in the labor market.

[1] We are very grateful to Antonio Ciccone, Michèle Tertilt, and Minchul Yum for their continued and invaluable support and guidance in this project. In addition, we thank Hans Peter Grüner, Andreas Gulyas, Yannick Reichlin, Luca Henkel, and numerous seminar and conference participants 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, CREST-ENSAE

2.1 Introduction

School tracking—the allocation of school children into different types of schools at some point during their school career—is a common feature of education policy across OECD countries. For example, in 2018, 27 out of 37 countries in the OECD had an education system with at least two distinct school programs available to 15-year-old students.[3] The argument behind tracking is typically one of efficiency: grouping children by ability and aspirations creates more homogeneous classrooms and allows for tailored instruction levels and curricula, which improves educational outcomes (Duflo et al., 2011). On the other hand, as the track decision is often related to the socioeconomic background of children, tracking may impair socioeconomic mobility across generations and increase inequality in education and income (Carlana et al., 2022; Meghir and Palme, 2005; Pekkarinen et al., 2009; Hanushek and Wößmann, 2006). This concern is particularly strong for countries that track at a relatively young age (Dustmann, 2004). For this reason, school tracking, and, in particular, its timing, is a recurrent issue in the public and academic debate about education reforms in countries with an early tracking regime, such as Germany (OECD, 2020a).[4]

This paper contributes to the debate by quantitatively assessing the long-run aggregate, distributional, and inter-generational effects of school tracking policies. Any such assessment needs to consider the effects of tracking on educational outcomes in school and college, the effects of the supply of different skills on labor market outcomes, and the intergenerational effects of parental skills differences. Quantitative macroeconomic models of overlapping generations have proven useful in analyzing these effects and the interplay between them, but have so far not incorporated how skill accumulation is affected by school tracking policies (Lee and Seshadri, 2019; Daruich, 2022; Restuccia and Urrutia, 2004; Yum, 2023). We aim to fill this gap by providing a macroeconomic model that features tracking in secondary school, allowing us to quantify the role that tracking plays for aggregate and distributional socioeconomic outcomes, within and across generations.

[3] An overview of school tracking policies in OECD countries is given in Chapter 3 in OECD (2020b). We differentiate school tracking, which refers to allocating students into physically distinct types of schools that differ in the curriculum taught, intensity, and length, from ability grouping within a school, where the curriculum and educational goals remain the same. School tracking is also common among non-OECD countries. Based on 2018 PISA data, only two out of 38 non-OECD countries with available information featured an education system with one comprehensive school program available to 15-year-old students.

[4] There is substantial variation in the timing of tracking across countries (see Figure V.3.9 in OECD, 2020b, based on 2018 PISA data). Germany and Austria are among the countries with the earliest track selection, at age 10. The Slovak Republic and the Czech Republic track at age 11; Belgium, the Netherlands, Switzerland, Indonesia, and Singapore at age 12; Bulgaria at age 13; Argentina, Italy, and Slovenia at age 14; France, Greece, Israel, Japan, Mexico, Portugal and many other countries at age 15. Other countries, like the US, UK, Australia, or the Scandinavian countries do not track during secondary school.

The model is built around a parsimonious theory of how school tracking affects the accumulation of skills in school. Our skill accumulation technology implies that each child has an optimal pace of instruction, which is increasing in children's skills. The pace of instruction can differ across school tracks but not within tracks. Policymakers choose the pace of instruction across tracks to maximize aggregate skills at the end of secondary school. We also allow for direct peer effects—children learn more if their school peers have higher skills. Because of the (endogenous) stratification in skills implied by school tracking, this is a further channel through which tracking affects skill accumulation in school. Under linear direct peer effects and absent any shocks to child skills during their time in school, the skill formation technology implies that the optimal tracking policy should perfectly stratify children when they start school. However, in the more realistic scenario where children's skills develop at different—and hard to predict—tempos as they grow older, early tracking may lead to lower aggregate end-of-school skills because of a mismatch between children's skills and the pace of instruction. Also, early tracking can increase inequality in educational outcomes. Another interesting implication of tracking is that children who lose in terms of skills are often concentrated in the track with the lower instruction pace. Thus, our child skill formation technology rationalizes some of the most robust empirical findings regarding school tracking in the literature and encompasses the main arguments about school tracking frequently made in the public discourse.[5]

We embed our theory of skill accumulation in school into a general-equilibrium life-cycle incomplete-markets framework of overlapping generations, in which parents care about their offspring in the tradition of Becker and Tomes (1986). Some aspects of the model are tailored to fit the German Education System. Children are tracked into two school tracks at the age of ten, typically at the end of four years of comprehensive primary school. Of the two school tracks, only one permits direct access to college (called academic track), but there is a second-chance opportunity for children in the other track (called vocational track). The school track children attend is decided by parents, who are altruistic and principally decide based on what's best for their children. However, parental preferences over school tracks

[5] Empirical estimates of the effects of (early) tracking on average learning outcomes of school children are often ambiguous (Hanushek and Wößmann, 2006). Evidence for the effects of tracking on inequality is more consistent, finding that tracking raises educational inequality and tends to predominantly disadvantage children from lower socioeconomic backgrounds (see, for instance, Meghir and Palme (2005), Aakvik et al. (2010), and Pekkala Kerr et al. (2013) for evidence from Scandinavian countries and Matthewes (2021) and Piopiunik (2014), for the case of Germany). While opponents of early tracking argue in favor of postponing the tracking age as a means to increase equality of opportunity in access to education for disadvantaged children (for example Wößmann (2020) in Germany), proponents argue that in a comprehensive school, children who learn quickly are thwarted, while slower-learning children overstrained, resulting in learning losses (see, for example Esser and Seuring (2020) in Germany).

can also play a role. End-of-school skills translate into adult human capital, which evolves stochastically over the working life and determines wages. Going to college allows access to the college-skill labor market, which affects the wage profile over the working years, but incurs an opportunity (time) cost and psychic costs that depend on end-of-school skills. The distribution of human capital across college and non-college workers affects equilibrium wages, which parents anticipate when they choose the school track for their children. Households can save into a non-state-contingent asset subject to life-cycle borrowing constraints, and parents can make a non-negative inter-vivos transfer when children become independent.

We solve for the steady-state equilibrium of the model numerically and calibrate the parameters in two steps. First, we estimate the skill formation technology parameters directly from German data on school children (Blossfeld et al., 2019) using a latent variable framework as in Cunha et al. (2010) and Agostinelli et al. (2023). In particular, we use information on achievement test scores to measure children's skills at different stages of their school careers. We then calibrate the remaining parameters to match a set of salient moments from representative German survey data. The model matches the data well, both in terms of aggregate moments and in terms of the distribution of skills across school tracks and parental backgrounds, as well as the transitions through the education system. To test the model's validity, we investigate non-targeted moments, such as the determinants of the school track choice. The model reproduces well the relationship between skills and school track choice by parental background. In addition, we also check our calibrated model against Dustmann et al. (2017)'s empirical observation that for children at the margin between two school tracks in Germany, school tracking is inconsequential for earnings later in life. We do so by computing the effects of the initial school track on later-in-life economic outcomes for a set of children who are around the margin between the two school tracks in equilibrium. Our model confirms that lifetime economic outcomes for these children are similar no matter which school track they attend.

To better understand the role of skill formation during the school tracking years for lifetime inequality, we implement a variance decomposition analysis in the spirit of Huggett et al. (2011) and Lee and Seshadri (2019). We find that around a third of the variation in lifetime economic outcomes is accounted for at age ten, just after the school track choice. This share rises to around two-thirds at age eighteen after the college choice, a number consistent with the literature. This suggests that the evolution of skills during the tracking years in secondary school is crucial for determining lifetime inequality, underscoring the importance of understanding skill formation during these years.

Our main policy interest is in the long-run and welfare effects of later school tracking or, equivalently, a longer period of comprehensive schooling. We find that a policy reform that postpones the school tracking age by four years to age fourteen—the average tracking age in OECD countries—entails an efficiency-mobility trade-off. On the one hand, postponing the tracking age improves social mobility as it decreases the intergenerational elasticity of income by around 2.2%. These mobility gains arise primarily because there is less heterogeneity in skill accumulation during comprehensive schooling, with children who would have gone to an academic track school losing and children who would have gone to a vocational track school gaining. This reduces the heterogeneity of end-of-school skills and, in particular, the skill differences across school tracks. On top of that, by the time the track decision is made, children from different parental backgrounds become more similar, and the later track choice depends less on parental background. There is also a source of mobility gains from labor market adjustments. The decrease in heterogeneity in end-of-school skills translates into smaller differences in human capital across college and non-college workers, and the wage premium falls. Moreover, as skills are linked between generations, this also reduces the initial skill differences of the next generation of children by parental background, reinforcing the effects on skill heterogeneity. Ultimately, inequality, as measured by the Gini coefficient of earnings, drops by 0.4%.

On the other hand, our results indicate that postponing tracking comes at the cost of a 0.1% drop in GDP and a 0.05% drop in consumption equivalent units. This is because 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 even though the later tracking decision occurs after more uncertainty about skills has been resolved. As a result, later tracking leads to lower aggregate end-of-school skills and lower aggregate output. The 0.1% drop in GDP depends on general equilibrium adjustments in the labor market that influence school track and college decisions. In partial equilibrium, holding wages fixed, aggregate output would increase as the share of academic-track children and college-educated workers rises. Abolishing tracking in favor of comprehensive schooling altogether would further exacerbate the efficiency-mobility trade-off.[6]

Finally, we evaluate the effects of reducing the direct influence of a child's socioeconomic background on the school track choice. Our data indicate that when parents go against the school track recommendation of their children's primary school teachers, it is generally in favor of their own educational path. We rationalize this as coming from parental preferences.[7]

[6] A similar trade-off has been highlighted in the literature about the effects of economic segregation on growth and inequality (see Bénabou, 1996) and more recently by Arenas and Hindriks (2021) where efficiency gains from unequal school opportunities arise because of positive assortative matching between parents who invest more in their children and better schools. In contrast, in our case, efficiency gains arise from matching similar-ability peers to tailored instruction levels for longer.

[7] These preferences may refer to actual track tastes, as parents might want their children to follow in their

Consistent with previous findings (e.g. Dustmann, 2004), our calibrated model yields that parental preferences play an important role for school track choices and may result in an inefficient allocation of children across tracks.[8] An important question is whether the consequences of such “misallocation” effects are visible not only in child skill outcomes but also in the aggregate and distributional outcomes in the economy. We show that eliminating the preference-driven influence of parental background on the school track leads to improvements in both social mobility and economic output in the range of 0.9% and 0.04%, respectively, and raises aggregate welfare by 0.04% in consumption equivalent units. The reason is that the initial school track choice depends more strongly on skills, which improves the teaching efficiency in each track and thereby raises the aggregate skill level. These results highlight that measures, such as well-executed mentoring programs, which provide information and support to children from lower socioeconomic families and have been effective in alleviating the negative influence of family background on school track decisions (Raposa et al., 2019), can not only improve the outcomes of these individual children but also lead to aggregate efficiency gains in the economy.

Related Literature

This paper links several strands of the literature: the quantitative macroeconomic literature on inequality and mobility, the literature on children's skill formation during school years, and the school tracking literature.

Much research in the quantitative macroeconomic literature on inequality and intergenerational mobility shares our focus on the role of skill formation, education, and education policies (e.g. Becker and Tomes, 1979, 1986; Restuccia and Urrutia, 2004; Lee and Seshadri, 2019; Abbott et al., 2019). While many papers in this literature focus on early childhood education and the role of parental investments (e.g. Daruich, 2022; Yum, 2023; Caucutt and Lochner, 2020; Lee and Seshadri, 2019), or study the role of college-education policies (e.g. Krueger and Ludwig, 2016; Abbott et al., 2019; Capelle, 2022), few papers explicitly include the secondary schooling stage into their analysis. An exception is Fujimoto et al. (2023), who study the importance of free secondary schooling for misallocation driven by borrowing

footsteps (Doepke and Zilibotti, 2017), or to some form of biased information about how each school track influences later outcomes or about the costs associated with completing it. In our data, we cannot tell the underlying reasons apart.

[8] For example, a college-educated parent may push 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 calibrate the extent of these asymmetric preferences in our model to replicate the share of deviations of the chosen school track from what had been recommended by the primary school teachers in our data.

constraints in Ghana. However, in their context, secondary schooling is the highest education level. In addition, recent research has highlighted the heterogeneous impact of school closures in the wake of the Covid pandemic on children at different stages of their schooling career (Jang and Yum, 2022; Fuchs-Schündeln et al., 2022) and across public and private secondary schools (Fuchs-Schündeln et al., 2023). Our contribution is to study a widespread feature of education policy during secondary school—tracking. Even though tracking occurs in many countries and (early) tracking policies are often made responsible for persistent inequality and social immobility in the public debate, an analysis of broad reforms to tracking policies in the macroeconomic literature is, to the best of our knowledge, missing.[9]

Our theory of skill formation during schooling years builds on the insights of the literature on child skill formation, which studies how children's skills evolve as a function of endowments, parental and environmental inputs, and schooling and teaching inputs (see, for instance, Cunha and Heckman, 2007; Cunha et al., 2010; Agostinelli et al., 2023, 2019; Duflo et al., 2011; Aucejo et al., 2022; Bonesrønning et al., 2022). To incorporate how tracking affects learning in secondary school, we consider two forms of peer effects.[10] First, similar to Agostinelli (2018), we incorporate direct peer effects, which capture the idea that children are affected by different-quality peer groups in a school track. Second, following Duflo et al. (2011)'s evidence in Kenyan primary schools and Aucejo et al. (2022)'s findings of complementarities between classroom composition and teaching practice in the US, we consider how the instruction levels across tracks adjust endogenously to the skill composition in that track. For that reason, school tracking is conceptually different from schools of different qualities (often related to neighborhood effects), which would mechanically disadvantage children in the lower-quality tracks. For example, Arenas and Hindriks (2021) provide a model of unequal school opportunity, defined as unequal school quality and access probability to the best schools, and quantify its effect on intergenerational persistence, highlighting the role of positive assortative matching between parents who invest more into their children and

[9] There is an extensive literature in education economics, which theoretically analyzes tracking policies, to which we relate (see Epple et al. (2002) and Betts (2011) for a general theoretical foundation of tracking). This literature tends to conclude that the effects of tracking on the level and distribution of educational outcomes are often theoretically ambiguous and depend on the shape of peer effects, resources between tracks, or the uncertainty surrounding child abilities. We make a similar point in Section 2.3. Brunello et al. (2007) offer an analysis of the optimal timing of tracking focusing on the role that an increasing demand towards more general skills plays, while Brunello et al. (2012) estimate the efficiency losses of deviating from the optimal tracking age across Europe, finding losses in the range of half a percent of GDP, on average. Our contribution to this literature is that we provide a richer framework that incorporates important macroeconomic effects of tracking on higher education and labor market outcomes and allows us to draw conclusions on the effects of tracking on mobility across generations.

[10] A summary of theoretical models of peer interactions and their implications for tracking policies can be found in Epple and Romano (2011). Sacerdote (2011) provides an overview of empirical approaches to measuring peer effects in education.

high-quality schools. In contrast, schools tracks differ endogenously in their instruction pace. Choosing a school track is thus less about choosing a "good" versus a "bad" school but more about choosing a school that fits a child's learning needs.[11]

We incorporate the skill formation technology with tracking into a standard life-cycle model with intergenerational linkages, such that the initial conditions of a new generation are endogenous, following the work in Daruich (2022); Lee and Seshadri (2019); Yum (2023). Moreover, we share with these papers the importance of considering GE effects when studying policy reforms that affect the skill composition in the economy. While the quantitative macroeconomic literature focuses almost exclusively on the US, where tracking across schools is uncommon (Fujimoto et al., 2023, is a notable exception as they focus on a developing country, Ghana), we focus on a country with a very early tracking system—Germany. At the same time, our school track model is general enough to be used in other countries that track between schools, such as many European countries, but also Asian countries like Korea and Singapore, or South American countries like Argentina or Uruguay (OECD, 2020b), and could even be adapted to be informative for countries where tracking occurs mostly within schools, across classrooms, such as in the US.

Lastly, this paper connects to an extensive empirical literature that estimates the effects of school tracking, and, in particular, its timing, on educational and later-in-life outcomes of students.[12] This literature typically either exploits temporal within-country variation in tracking practices (Meghir and Palme (2005), for Sweden; Aakvik et al. (2010), for Norway; Bauer and Riphahn (2006), for Switzerland; 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. Dustmann et al. (2017) use an individual-level instrumental variables strategy (the date of birth) and find no effect of the school track on educational attainment or earnings for students at the margin between two tracks. This result suggests that school tracking in Germany is largely inconsequential in the long run for children whose skills put them in between tracks when the decision was made, a result that our model accommodates. We add to this literature a quantitative model-based assessment of the long-term aggregate, distributional, and welfare effects of broad reforms to the school tracking age, which is

[11] Similarly, tracking is also different from having private schools or schools with different costs, where selection into schools is likely to depend on parental wealth directly. The degree to which private schools are similar to a tracking system will then depend on the correlation between child skills and parental wealth.

[12] See Betts (2011) for an excellent overview. We relate our model-based predictions of these effects to the findings in this literature in detail in Appendix 2.C.

difficult to establish empirically.

The remainder of the paper is organized as follows. Section 3.2 presents our model of overlapping generations and tracking during secondary school and introduces the skill formation technology. In Section 2.3, we build intuition about the model mechanisms underlying school tracking by deriving theoretical implications of that technology. Section 2.4 explains how we parameterize and calibrate the model. It also offers some validation exercises. In Section 3.3, we use the calibrated model to perform a series of counterfactual experiments to quantify the effects of different school tracking policy regimes. Finally, Section 3.4 concludes.

2.2 The Model

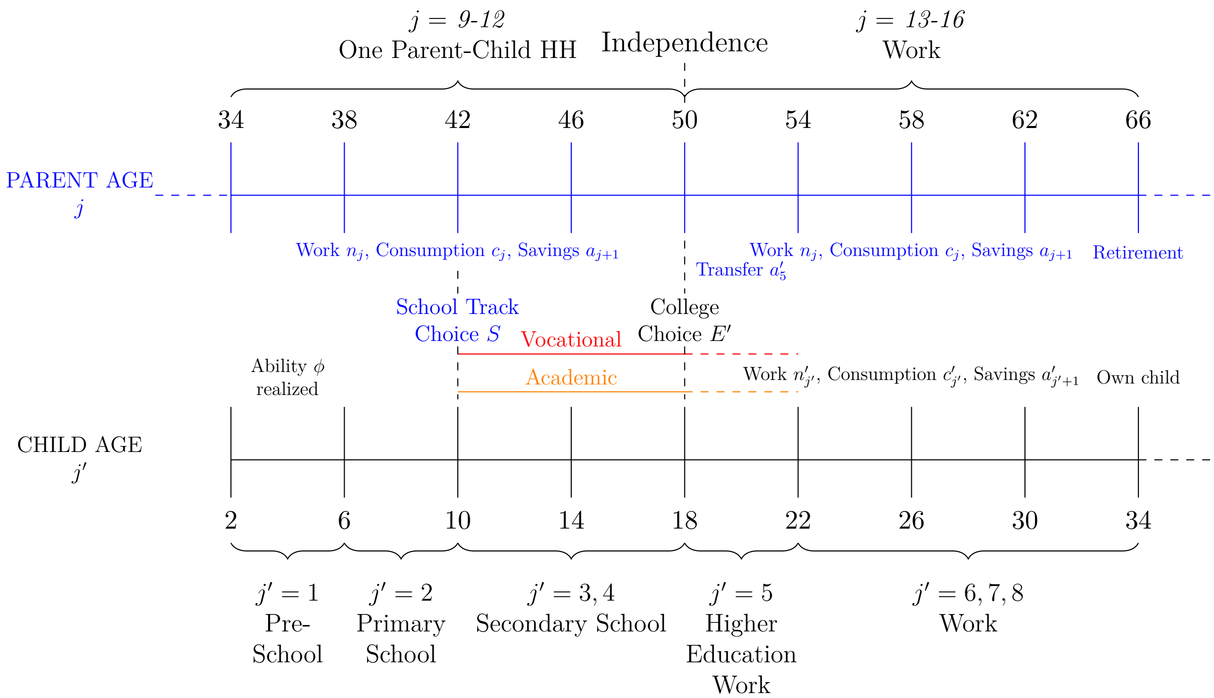
Time is discrete and infinite, and one model period, $j \in \{1, ..., 20\}$, corresponds to the four years between ages $[4j - 2, 4j + 2]$ in real life. Thus, agents enter the model as two-year-old children and exit at age 82.[13] This frequency allows us to investigate meaningful variations in school tracking ages. The structure implies that 20 generations are alive at every point in time. As in Lee and Seshadri (2019), we assume a unit mass of individuals in each period.

A life cycle can be structured into several stages, as illustrated in Figure 2.1: During the first four periods, a child lives with her parent, goes to school, and accumulates (school) skills. School tracking happens at the beginning of child period $j = 3$. At the beginning of child period $j = 5$, at age 18, the child becomes an independent adult, her skills are transformed into adult human capital, and she can decide to go to college. Both college and non-college-educated types of labor are used, next to capital, by a representative firm to produce the final consumption good. Adult agents decide how much labor to supply until they retire at the beginning of $j = 17$, at age 66. During the working periods, human capital grows stochastically. Finally, in $j = 9$, when they are 34 years old, adults become parents of a child. Adults make inter-vivos transfers to their children when they turn 18 and become independent.[14]

[13] We choose this perhaps unorthodox timing to capture that in Germany, children are ten years old when parents make the secondary school track decision, which resembles reality in Germany. Appendix 2.B.1 gives an overview of the German Education System.

[14] For the remainder of the text, we will denote all child variables with primes whenever both parental and child states are present. The child of a parent who is in period j is in period $j' = j - 8$.

Figure 2.1: Timeline of Life-cycle Events



2.2.1 Child Skill Formation

Every new child has an initial ability or skill endowment, ϕ, which is imperfectly transmitted from her parent.[15] When children enter primary school, at the beginning of $j = 2$, the initial ability translates into a first child school skill level, θ_2, expressed in logarithm, to which we refer to as skills henceforth.

$$\theta_2 = \log \phi. \tag{2.1}$$

We think of these skills as stage-specific competencies during the schooling periods, $j = 2$ to $j = 5$, that can be observed by everyone and are rewarded on the labor market for both college and non-college-educated workers.

Subsequently, the evolution of skills depends on the schooling system. During primary school ($j = 2$), the system is *comprehensive*, meaning that there is only one track to which all schools belong, denoted by $S = C$. During secondary school, there are two distinct school

[15] 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. The initial ability thus captures genetic components and investments made by parents into their child's development during early childhood, infancy, and even in-utero.

tracks, a *vocational* track $S = V$ and an *academic* track $S = A$.[16] School tracks can differ in their *pace of instruction*, denoted by P^S, which reflects the differences in the intensity and depth with which school subjects are taught.[17] Notably, the pace of instruction in each school track is endogenous in the sense that the education policymaker can choose it in every period to achieve her goals. For our analysis, we assume that the policymaker has an efficiency goal and maximizes aggregate end-of-school skills.[18] We further assume that all classrooms and schools in the same track are identical. Thus, if a child is allocated to a particular track, we can think of her as attending a "representative" classroom and school for that track. This implies that all children in a given track are exposed to the same set of classroom and school peers.

The technology of (log) skill formation during the school years $j = 2, 3, 4$, of a child in school track S, is then given recursively by:

$$\begin{aligned} \theta_{j+1} &= \kappa\theta_j + \alpha\bar{\theta}_j^S + g(\theta_j, P_j^S) + \zeta E + \eta_{j+1} \\ \eta_{j+1} &\sim \mathcal{N}(0, \sigma^2_{\eta_{j+1}}). \end{aligned} \tag{2.2}$$

Next period's skills are directly affected by past skills θ_j and parental education E, which we take as a proxy for the home environment in which a child grows up, including differences in

[16] While in principle a larger number of school tracks is conceivable, we restrict our analysis of tracking to two school tracks as this corresponds to a typical number across OECD countries. Typically, the two tracks serve the purpose of preparing children for academic higher education at a college or similar institution or to prepare children for a more vocational career.

[17] In Germany, the curricula and core subjects are no longer materially different across school tracks. The main difference between academic and vocational schools is that the former results in direct qualification to enter university, while the latter does not. In academic track schools, topics are generally taught more densely and comprehensively than in vocational track schools, preparing students for higher education. Moreover, students typically have more options for elective subjects at later stages of secondary school. Vocational track schools, by contrast, are less demanding in terms of the required learning effort, and graduation occurs after fewer years. A detailed comparison between the teaching intensity and learning goals across Germany is provided in Dustmann et al. (2017). Note that 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. In Appendix 2.B.1, we summarize information on expenditure per student as well as teacher quality across different school tracks in Germany.

[18] 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, methods, and specific topics that should be taught in each school track, subject, and 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. The concrete implementation of the curricula, however, is in the hands of the teachers and individual schools, who have some discretionary margin to adjust the instruction paces to the needs of their pupils. Therefore, we view the pace-setting process as a mix of overarching learning goals and individual adjustment across school tracks. To the best of our knowledge, there is no clear teaching goal about the distribution of end-of-school skills formulated by German education policymakers.

parental investments into child skills by parental background (Heckman and Mosso, 2014). By η_{j+1}, we denote unobserved i.i.d. shocks to the skills. This type of uncertainty in the formation of child skills is crucial for analyzing school tracking policies. We interpret these shocks as stemming, for example, from unexpected heterogeneity in child development speeds (such as late-bloomers), but also health shocks that can permanently influence the skill formation trajectory of a child.[19]

The school track can affect next period's skills in two ways: First, through direct interactions with peers in a track, which affect future skills linearly through the average skill level of other children in school track S, denoted by $\bar{\theta}_j^S$, as is common in the peer effects literature (Sacerdote, 2011).[20] Second, through the pace of instruction in her school track, P_j^S, as governed by the function g, which we assume takes the following form:

$$g(\theta_j, P_j^S) = \beta P_j^S + \gamma \theta_j P_j^S - \frac{\delta}{2}(P_j^S)^2. \tag{2.3}$$

This functional form implies firstly that for each skill level θ_j, there exists an individually-optimal instruction pace, $P_j^*(\theta_j)$, that maximizes future skills in each period. Secondly, if $\gamma > 0$, there is a positive complementarity between the individually optimal pace and the individual skill level, such that higher-skilled children also prefer a higher pace of instruction. This is motivated by evidence on the heterogeneous effects of teaching or instructional practices depending on prior student achievement and, in particular, by evidence on "match" effects between teaching practices and classroom skill composition (see Duflo et al. (2011), Aucejo et al. (2022) and references therein). As we will demonstrate theoretically in Section 2.3 and quantitatively in Section 3.3, this complementarity plays a central role in providing the rationale behind any efficiency argument in favor of school tracking policies.

Given (2.3), it is clear that aggregate learning is maximal if every child is taught at her preferred instruction pace in every period. However, there is only one instruction pace per school track. Given this constraint, a policymaker seeking to maximize expected future skills would then set the pace in each track to the one that is optimal for a child with exactly the

[19] Our assumption of shocks as the source of 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 a model with imperfectly observed child skills would imply in Appendix Section 2.D.

[20] 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. We note, however, that non-linear peer effects could have important implications for the assessment of tracking policies.

average skill level in that track, as summarized in Lemma 1.

Lemma 1. The pace of instruction a policymaker would set in each school track to maximize expected skills in the next period is given by

$$P_j^S = P_j^*(\bar{\theta}_j^S) = \frac{\beta + \gamma\bar{\theta}_j^S}{\delta} \tag{2.4}$$

where $\bar{\theta}_j^S$ is the average skill level of children in track S.

Proof. Follows from taking the first order condition of the conditional expected value $\mathbb{E}(\theta_{j+1}|S)$ in (2.2) with respect to P_j^S using (2.3) and under the i.i.d. assumption of η_{j+1}, and the fact that maximization of skills in each school track is a necessary condition for maximizing unconditional skills. □

According to Lemma 1, the instruction pace setting implies that future child skills depend non-linearly on her peers' skills. Specifically, skill gains decrease monotonically with the distance between a child's own skills and the average skill level in that track, or equivalently with the distance between her optimal instruction pace and the one she is currently taught at.[21] Consequently, for a child with a low skill level, going to a school track with a high instruction pace tailored to a higher average skill level can be harmful to the point that she actually learns less, despite being surrounded by better peers, than if she had attended a school with a lower pace.

After finishing school, at the beginning of $j = 5$, child skills are transformed one-to-one into the first adult human capital level, h_5,

$$h_5 = \exp(\theta_5). \tag{2.5}$$

2.2.2 Preferences

Preferences over consumption and labor supply of adults in each period are given by

$$u(c_j, n_j) = \frac{(c_j/q)^{1-\sigma}}{1-\sigma} - b\frac{n_j^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} \tag{2.6}$$

[21]See Appendix 2.A.1 for the derivation. Our formulation of learning hence implies that non-linear peer effects are driven by how the instruction levels are adjusted (as found, for example, in Duflo et al., 2011; Lavy et al., 2012). Moreover, it provides a micro-foundation for efficiency gains in average learning that stem from more homogeneous peer groups. We discuss the theoretical consequences of tracking under these assumptions in Section 2.3.

where c_j denotes household consumption and q is an adult consumption-equivalent scale that is larger than one whenever there is a child in the household and one otherwise. Risk aversion is captured by σ. Individuals incur disutility from working n_j hours, which is governed by b and the Frisch elasticity of labor supply, γ. All future values are discounted by β.

2.2.3 Educational Choices

There are two types of educational choices agents make during their life. The first and novel education choice is the secondary school track parents choose for their children. We assume that the utility of parents also depends on the track their children attend, through a stochastic academic-track utility cost $\chi(E) \sim H^E(\chi)$, whose distribution can depend on parental education E. This will allow us to capture that empirically, the school track decision is significantly affected by parental socio-economic status, even conditional on school performance, test scores prior to the track decision, and the track recommended by primary school teachers. Moreover, when parents deviate from the primary school teacher's recommendation, it is usually toward their own education path.[22]

There may be multiple reasons behind these parent-specific academic track costs. For example, there may be a cost associated with acquiring information about school tracks that is lower whenever a parent went to that track herself. Similarly, parents may feel better able to support their child in a track they are more familiar with. Parents may also systematically over- or underestimate their children's potential or have strong preferences for their child following in their footsteps. Whatever their exact reason, deviations in parent's track choice from the recommended path may lead to learning inefficiencies. For example, a child with low skills could be sent to the academic track by parents who have preferences for this track. This would lead to learning losses not only for the individual child but also create an externality for all other children as the instruction pace is endogenous to the peer composition.

Second, after finishing school, newly independent adults decide whether to go to college. In line with the literature (e.g. Daruich, 2022; Fuchs-Schündeln et al., 2022), we assume that going to college entails a "psychic" utility cost $\psi(S, \theta_5, \nu(E^p))$ that may depend on the secondary school track S, the end-of-school skills θ_5 and an idiosyncratic college taste shock,

[22]See Appendix 2.B.2 for some reduced-form evidence on the school track choice and deviations from the recommended tracks, by parental background. Importantly, children who deviate from the recommended school track perform differently than the others. Children who deviate from vocational to academic perform worse than the average kid in the academic track, and the reverse happens for children who deviate from academic to vocational. The fact that deviating from the recommended track does not seem to benefit children in terms of their achievements indicates that it is not the case that parents "know" the true potential of their child better or can support them better.

$\nu(E^p) \sim G^{E^p}(\nu)$, whose distribution may be influenced by the parent's education level E^p.[23]

This formulation can accommodate two important features of the transition between secondary and college education in the data. Firstly, the share of students with an academic track secondary school degree who get a college degree is significantly higher than those with a vocational secondary school degree.[24] Secondly, independently of the school track, the likelihood of college education in the data is increasing in the end-of-school skills.Finally, the random taste shocks reflect heterogeneity in the higher education decision coming from parental background or channels outside of the model, as is common in this literature.

2.2.4 Adult Human Capital, Labor Income and Borrowing

During the working career ($j = 5$ to $j = 16$), human capital grows according to

$$h_{j+1} = \gamma_{j,E}\, h_j\, \varepsilon_{j+1}, \quad \log \varepsilon_j \sim \mathcal{N}(0, \sigma_\epsilon^2) \tag{2.7}$$

following Yum (2023), where $\gamma_{j,E}$ are age- and education-specific deterministic growth rates and ε_{j+1} are market luck shocks, which follow an i.i.d. normal distribution in logs, with zero mean and constant variance σ_ε^2, as in Huggett et al. (2011). Human capital remains constant after retirement. Gross labor income is then given by

$$y_j = w_E\, h_j\, n_j \tag{2.8}$$

where w_E denotes the effective wage per unit of human capital paid to workers with higher education E.

Note that all prices, including w_E, implicitly depend on the distribution of agents in the economy, which we suppress for notational convenience. After retiring, each agent receives retirement benefits $\pi_j(h_{17}, E)$, which depend on the last education-specific human capital level before retirement.[25] Throughout their life, adult 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

[23] We add the superscript "p" here to indicate that E^p is the college education of the parent of a newly independent adult who chooses her own college education E.

[24] In Germany, every graduate from an academic track secondary school automatically obtains a college-entrance qualification, while graduates from vocational tracks do not. However, there exist a variety of "second-chance" opportunities to obtain college-entrance qualification for vocational-track graduates, such as evening schools (see Dustmann et al., 2017).

[25] As is common in the literature, we let benefits depend on human capital in this way to proxy for lifetime earnings, which form the basis of pension benefits in reality.

their children. The per-period borrowing constraint can thus be written as

$$a_{j+1} \geq \frac{-g}{1+r}. \tag{2.9}$$

In the following, we provide a recursive formulation of the agent's decisions in each life cycle stage.

2.2.5 Recursive Formulation of Decisions

At the beginning of each adulthood period prior to retirement, individuals learn about their market luck shock realization and, in case they have a child, about the child skill shock realization. Based on this information, they decide on consumption (c_j), savings (a_{j+1}), and hours worked (n_j). In addition, there are two education choices—the school track and the college decision—and parents decide on inter-vivos transfers in period $j = 13$. All decisions are subject to the human capital growth technology (2.7), the borrowing constraint (2.9), a working time constraint $n_j \in [0, 1]$ and a period budget constraint

$$c_j + a_{j+1} = y_j + (1+r)a_j - T(y_j, a_j) \tag{2.10}$$

where labor income is defined as in (2.8) and $T(y_j, a_j)$ gives taxes net of lump-sum transfers, which consist of labor income and capital taxes.

Parenthood (Age 34-50, periods $j = 9, ..., 13$)

Parent with a Young Child ($j = 9,\ 10$) The state space in these periods consists of the parent's education E, her human capital, h_j, and her assets a_j. Parents observe their child's initial ability ϕ at the start of the first period of the child's life, when she is two years old, which corresponds to the child's first school skill at age six, as given by (2.1).

Future skills $\theta_{j'+1}$ evolve according to (2.2) given the optimal pace of instruction as defined in Lemma 1. In particular, primary schools are comprehensive track schools, such that the evolution of a child's skills depends on the average skill level of all children in their cohort, $\bar{\theta}_{j'=2}$. The problem of the parent can then be written as:

$$\begin{aligned} V_j(E, h_j, a_j, \phi, \theta_{j'}) = &\max_{c_j, a_{j+1}, n_j} \left\{ u(\frac{c_j}{q}, n_j) + \beta\, \mathbb{E}\, V_{j+1}(E, h_{j+1}, a_{j+1}, \phi, \theta_{j'+1}) \right\} \\ &\text{s.t. } \theta_{j'+1} = \kappa\theta_{j'} + \ \alpha\bar{\theta}^S_{j'} + \ g(\theta_{j'}, P^*_{j'}(\bar{\theta}_{j'})) + \ \zeta E + \eta_{j'+1} \\ &\qquad (2.7) - (2.10) \end{aligned} \tag{2.11}$$

where expectations are taken over child skill shocks ($\eta_{j'+1}$), market luck shocks (ε_{j+1}), and in period $j = 10$ also over school track taste shocks $\chi(E)$.

The School Track Decision ($j = 11$) When the child turns ten, at the beginning of her third period of life, the parent decides on whether to send her to the vocational or academic track school, $S \in \{V, A\}$. The decision of parents is not constrained by any education policy (but parents do generally obtain a track recommendation from their children's primary school).[26] Once a child is tracked, she remains in that track for two periods, until the end of secondary school, when she turns 18.[27] Parents make the track decision by comparing the value of sending the child to a vocational track school with that of sending her to an academic track school. These (interim) values are given by

$$\begin{aligned} W_{11}(E, h_{11}, a_{11}, \phi, \theta_3, S) = \max_{c_{11}, a_{12}, n_{11}} & \left\{ u(\frac{c_{11}}{q}, n_{11}) + \beta \, \mathbb{E} \, V_{12}(E, h_{12}, a_{12}, \phi, \theta_4, S) \right\} \\ \text{s.t. } & \theta_4 = \kappa\theta_3 + \alpha\bar{\theta}_3^S + g(\theta_3, P_3^*(\bar{\theta}_3^S)) + \zeta E + \eta_4 \\ & (2.7) - (2.10) \end{aligned} \tag{2.12}$$

for each track S. They encode several incentives that influence the track decision. On the one hand, academic track attendance makes, ceteris paribus, college access more likely, which results in higher human capital growth and productivity over the life cycle. The returns to college education depend on the demand for college-type labor. On the other hand, parents know that her child's skill formation depends on the average skill level in a school track $\bar{\theta}_3^S$, both directly through peer interactions but also indirectly through the endogenous optimal instruction pace P_3^S. Thus, parents need to anticipate the distribution of children across tracks when making the track decision, which becomes an aggregate state, which we keep implicit.

On top of that, the track decision is also affected by the stochastic academic track utility

[26] 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.

[27] We abstract from track switches during secondary school, as these are relatively rare in the data. 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 it is likely an upper bound of the track switches between the vocational and academic tracks. However, this does not preclude track switches between the end of secondary school and the beginning of possible tertiary education, which we allow in our model.

shock, $\chi(E) \sim H^E(\chi)$. Parents then make the discrete track choice using (2.12) after observing a draw of $\chi(E)$. Thus, we can define the value of a parent after this shock realization at the beginning of period $j = 11$ as

$$\begin{aligned} V_{11}(E, h_{11}, a_{11}, \phi, \theta_3) = \max_{S \in \{V,A\}} \{ & W_{11}(E, h_{11}, a_{11}, \phi, \theta_3, S = V), \\ & W_{11}(E, h_{11}, a_{11}, \phi, \theta_3, S = A) - \chi(E)\}. \end{aligned} \tag{2.13}$$

Remaining Parenthood ($j = 12, 13$) In period $j = 12$, when the child is 14 years old and starts the second period of secondary school, the parent solves the following problem:

$$\begin{aligned} W_{12}(E, h_{12}, a_{12}, \phi, \theta_4, S) = \max_{c_{12}, a_{13}, n_{12}} & \left\{ u(\frac{c_{12}}{q}, n_{12}) + \beta \, \mathbb{E} \, V_{13}(E, h_{13}, a_{13}, \phi, \theta_5, S) \right\} \\ \text{s.t. } & \theta_5 = \kappa\theta_4 + \ \alpha\bar{\theta}_4^S + \ g(\theta_4, P_4^*(\bar{\theta}_4^S)) + \ \zeta E + \eta_5 \\ & (2.7) - (2.10) \end{aligned} \tag{2.14}$$

where the child's school track S, which has been decided in the previous period, is now included in the parent's state space.

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 $V_{j'=5}$. As in Daruich (2022), we model this as an interim decision problem and assume that the parent already knows the realization of her market luck shock and her child's final skill shock but does not know the realization of the college taste shock $\nu'(E)$. The transfer cannot be negative, so parents cannot borrow against the future income of their child. The strength of parental altruism is governed by the factor Λ. The value at the beginning of period 13 is then

$$\begin{aligned} V_{13}(E, h_{13}, a_{13}, \phi, \theta_5, S) = \max_{a_5' \geq 0} & \left\{ \widetilde{V}_{13}(E, h_{13}, a_{13} - a_5') + \Lambda \, \mathbb{E}_{\nu'} \, V_{j'=5}(\theta_5, a_5', \phi, S, E) \right\} \\ \text{s.t. } & \nu'(E) \sim G^E(\nu') \end{aligned} \tag{2.15}$$

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

$$\begin{aligned} \widetilde{V}_{13}(E, h_{13}, a_{13}) = & \max_{c_{13}, a_{14}, n_{13}} \{u(c_{13}, n_{13}) + \beta \, \mathbb{E} \, V_{14}(E, h_{14}, a_{14})\} \\ \text{s.t. } & c_{13} + a_{14} + a_5' = y_{13} + (1 + r)a_{13} - T(y_{13}, a_{13}) \\ & (2.7) - (2.9) \end{aligned} \tag{2.16}$$

so that the transfer a'_5 enters the budget constraint.

Work Life Without a Dependent Child (Age 18-34 and 50-66, periods $j = 5, 6, 7, 8$ and $j = 14, 15, 16$)

Independence ($j = 5$) After turning 18, the state space of a newly independent adult comprises the secondary school track she graduated from S, end-of-school skills θ_5, initial assets a_5, which she received from her parents, initial ability ϕ and her parent's education E^p, which affects the distribution of the stochastic college taste shock $\nu(E^p)$. Conditional on the realization of that shock, the young adult first decides whether to go to college ($E = 1$) or not ($E = 0$) by solving the problem

$$\begin{aligned} V_5(\theta_5, a_5, \phi, S, E^p) = \max_{E \in \{0,1\}} \{ & W_5(E = 0, h_5, a_5, \phi), \\ & W_5(E = 1, h_5, a_5, \phi) - \psi(S, \theta_5, \nu(E^p))\} \end{aligned} \tag{2.17}$$

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

$$\begin{aligned} W_5(E, h_5, a_5, \phi) = & \max_{c_5, a_6, n_5 \in [0, \bar{n}(E)]} \{u(c_5, n_5) + \beta \, \mathbb{E} \, V_6(E, h_6, a_6, \phi)\} \\ & \text{s.t. } (2.7) - (2.10) \end{aligned} \tag{2.18}$$

and end-of-school skills are transformed into adult human capital h_5 according to (2.5). While agents can work during college education, they only receive the vocational wage rate w_0. Moreover, obtaining a college education reduces the time available for work, as individuals spend part of their total time endowment studying, thus $\bar{n}(E = 1) < 1$.

Remaining Work Life ($6, 7, 8$ and $j = 14, 15, 16$) In periods 6 and 7, which correspond to ages 22 to 30, adults solve

$$\begin{aligned} V_j(E, h_j, a_j, \phi) = & \max_{c_j, a_{j+1}, n_j} \{u(c_j, n_j) + \beta \, \mathbb{E} \, V_{j+1}(E, h_{j+1}, a_{j+1}, \phi)\} \\ & \text{s.t. } (2.7) - (2.10). \end{aligned} \tag{2.19}$$

In period $j = 8$, when they are age 30 to 34, adults know that they will have a child at the start of the next period. For that reason, they take expectations over the initial ability of their future child, ϕ', on top of the expectations over the market luck shock. Thus, we obtain that in period 8

$$
\begin{aligned}
V_8(E, h_8, a_8, \phi) = & \max_{c_8, a_9, n_9} \{u(c_8, n_8) + \beta\, \mathbb{E}\, V_9(E, h_9, a_9, \phi')\} \\
\text{s.t. } & \log \phi' = \rho_\phi \log \phi + \epsilon_\phi, \qquad \epsilon_\phi \sim \mathcal{N}(0, \sigma_\phi^2) \\
& (2.7) - (2.10)
\end{aligned}
\tag{2.20}
$$

where ϵ_ϕ is an intergenerational shock. For periods $j = 14, 15, 16$, when they are age 54 to 66, adults are again without a child and solve the standard life-cycle savings problem

$$
\begin{aligned}
V_j(E, h_j, a_j) = & \max_{c_j, a_{j+1}, n_j} \{u(c_j, n_j) + \beta\, \mathbb{E}\, V_{j+1}(E, h_{j+1}, a_{j+1})\} \\
\text{s.t. } & (2.7) - (2.10)
\end{aligned}
\tag{2.21}
$$

where the initial ability ϕ has been transmitted to the child and does not enter the state space anymore. In the last period prior to retirement, $j = 16$, agents no longer need to take expectations over market luck shocks, as human capital remains constant during retirement.

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

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

$$
\begin{aligned}
V_j(E, h_{17}, a_j) = & \max_{c_j > 0, a_{j+1}} \{u(c_j, 0) + \beta V_{j+1}(E, h_{17}, a_{j+1})\} \\
\text{s.t. } & c_j + a_{j+1} = \pi_j(h_{17}, E) + (1 + r)a_j - T(0, a_j) \\
& \text{and } (2.9).
\end{aligned}
\tag{2.22}
$$

2.2.6 Aggregate Production, and Government

A representative firm produces output according to the Cobb-Douglas production function $Y = AK^\alpha H^{1-\alpha}$, where A denotes total factor productivity, K is the aggregate physical capital stock, and H is human capital defined by:

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

H_0 is the aggregate labor supply in efficiency units of non-college workers, and H_1 is that of workers with 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 $\tau_a r a_j$. All tax revenue is used to finance retirement benefits π_j and fixed lump-sum social welfare benefits g that are paid to every household.

2.2.7 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 age-period j is constant across cohorts. As is standard, 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 wasteful government spending (or spending that is linearly separable in the utility of households).

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 therefore requires that parents form expectations over the skill distribution across school tracks, which have to coincide with the actual distributions in equilibrium. Appendix 2.A.2 gives a detailed definition of the equilibrium.

2.3 Developing Intuition: Tracking and Skill Formation

Our formulation of the skill formation technology during the schooling years in (2.2) constitutes the novel cornerstone of our model. We now develop some intuition about what it implies for skill accumulation with and without school tracking. Our focus in this section is exclusively on the secondary schooling years (periods 3 and 4), and we ignore transitions to higher education and the labor market. Moreover, we simplify parents' preferences, such that they only care about their child's expected end-of-school skills and have no other preferences regarding the school track choice. Finally, we assume for simplicity that $\kappa = 1$ and that there are no direct parental influences, $\zeta = 0$, nor stochastic track costs, $\chi = 0$.

All other assumptions are maintained. In particular, policymakers set the instruction paces in each school track to maximize expected end-of-school skills, such that the pace-setting rule in Lemma 1 holds. Moreover, we assume that the distribution of child skills at the beginning of secondary school is normal and centered around 0.

2.3.1 Comprehensive School versus Tracking

We start by comparing a comprehensive schooling system (C), in which all children attend the same school track, to a tracking system (T) in which all children are tracked into a vocational or academic track. For simplicity, we consider only one period of schooling here. Thus, if θ_3 are the skills at the beginning of secondary school, θ_4 can be considered the skills at the end of school. A key implication of this simplifying assumption, when combined with the timing of skill shocks in (2.2), is that skill evolution during secondary school occurs as if there were no skill shocks during that time. As we will see, this implies that aggregate end-of-school skills are always greater with tracking than in a comprehensive system.

The Allocation of Children across Tracks

We consider two alternative allocation mechanisms. In the first one, a policymaker (or a teacher) allocates children across tracks directly. As before, the goal of the policymaker is to maximize the expected end-of-school skills across all children ($\max_S \mathbb{E}(\theta_4)$).

The second alternative consists of each parent making the track decision unilaterally for her child i with skill level $\theta_{i,3}$. A parent's only goal is to maximize her child's expected end-of-school skill level ($\max_S \mathbb{E}(\theta_{i,4})$). Parents know the distribution of θ_3. We can thus think of this mechanism as a simultaneous move game played among parents, where each parent's strategy set consists of the two tracks she can send her child to, and the next period's skills give the payoffs.

Proposition 1 shows that, in both alternatives, the track decision that results in the optimum or equilibrium is governed by a sharp cut-off skill level. A policymaker would optimally split the distribution exactly at its mean. Intuitively, this generates the highest aggregate end-of-school skills as it minimizes the variance of skills in each track, thereby creating peer groups that are as homogeneous as possible. In doing so, the policymaker internalizes that any effects coming from the direct peer externality offset each other across tracks. Thus, all gains achieved 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 decide regardless of the aggregate outcomes. The equilibrium of this implied game still features a sharp skill threshold, which is characterized by the skill level at which a child's expected end-of-school skills are exactly equal in both tracks. This threshold is smaller than the optimal threshold a policymaker would pick whenever the direct peer effects are positive ($\alpha > 0$). The reason is that, because of positive direct peer effects, children with skills just below the policymaker's threshold would benefit individually from going to the academic track (with higher average skills). As parents do not internalize the effect of their decision on average skills in each track, they will therefore send their children to the academic track.

Proposition 1. The allocation of children across tracks is characterized by a skill threshold $\tilde{\theta}_3$, such that all children with initial skills below $\tilde{\theta}_3$ go to one track and all children with initials skills above $\tilde{\theta}_3$ go to the other track.

- If the policymaker does the track allocation, the optimal skill threshold corresponds to the average initial skill level $\tilde{\theta}_3^* = \mathbb{E}[\theta_3] = 0$.
- If parents do the track allocation, the skill threshold depends on the direct peer externality α. With $\alpha > 0$, the threshold is smaller than $\tilde{\theta}_3^*$.[28]

Proof. In Appendix 2.A.1. □

Next, we compare the comprehensive and tracking systems in terms of their effects on end-of-school skills. We refer to an optimal tracking system, when the policymaker makes the track allocation with the goal to maximize end-of-school skills, as in Proposition 1.

The End-of-School Distribution

Proposition 2 shows that independently of the sorting mechanism, expected end-of-school skills in an optimal tracking system are always larger than in a comprehensive system, provided that $\gamma \neq 0$ and $\delta > 0$. Intuitively, this advantage comes from more homogeneous peer groups in each track in terms of their skills. Since learning decreases with the variance of skills among children in a track, more homogeneity on average increases end-of-school skills. Therefore, the gain from tracking increases the smaller the conditional variance of skills across tracks, as given in equation (2.24). The gain from tracking further increases in the complementarity between own skills and instruction pace, γ. The stronger the complementarity, the more it pays to stratify children by their skills. Moreover, the advantage increases in the variance of initial child skills $\sigma^2_{\theta_3}$ but decreases in δ, which ultimately governs the concavity of learning with respect to the instruction pace.

A full tracking system may lead to larger inequality in end-of-school skills. In particular, condition (2.25) states that the variance of end-of-school skills might be larger in a tracking

[28] 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.

system with positive peer externalities if tracking occurs at the optimal skill threshold. This is more likely to hold the larger the direct peer externality α.

Similarly, an optimal tracking system necessarily leaves some 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 $\tilde{\theta}_3 = 0$, these children thus occupy the center of the distribution and would, given a choice, prefer a comprehensive system. If there are no 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 the children at the bottom of the skill distribution suffer from a tracking system (e.g. Matthewes, 2021).

Proposition 2.

- Expected end-of-school skills in a full tracking system are larger than in a fully comprehensive system. This holds regardless of who makes the track decision, i.e., regardless of the tracking skill threshold $\tilde{\theta}_3$. The gain from tracking is given by

$$\mathbb{E}(\theta_4|T) - \mathbb{E}(\theta_4|C) = \frac{\gamma^2}{2\delta}\left(\sigma^2_{\theta_3} - \mathbb{E}(Var[\theta_3|S])\right). \tag{2.24}$$

- The end-of-school skill distribution in a full tracking system has a "fatter" right tail. In case of tracking at the optimal skill threshold $\tilde{\theta}_3 = \mathbb{E}(\theta_3)$, the variance of end-of-school skills in a full tracking system is larger than the variance in a fully comprehensive system iff

$$\alpha^2 + 2\alpha\left(1 + \frac{\beta\gamma}{\delta}\right) - (8-\pi)\frac{\gamma^4}{\pi\delta^2}\sigma^2_{\theta_3} > 0. \tag{2.25}$$

- 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.1. □

The main reason why T always beats C here in terms of aggregate skills is the simplifying assumption of no skill shocks during students' time in school. As a result, the tracking decision made at the start of secondary school is optimal throughout secondary school.

2.3.2 Early versus Late Tracking

Let us now consider a two-period secondary schooling system, like in our full model, where there can be skill shocks during the time students are in secondary school. In this case, the skills at the end of school are θ_5. 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 both cases, the allocation of children to tracks is done optimally by a policymaker, maximizing the expected aggregate end-of-school skills ($\max_S \mathbb{E}(\theta_5)$). The early tracking system is characterized by a track allocation in $j = 3$ that is maintained throughout secondary school. The late tracking system is characterized by all children going to a comprehensive school in the first period, followed by tracking at the beginning of the second secondary school period ($j = 4$). Hence, in the case of LT, children are allocated to school tracks after the skill shocks η_4 are realized, while in the case of ET, the school track decision was made before the realization of these shocks.

Proposition 3 says that expected 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, this represents the key disadvantage of early tracking. Since the first track allocation is maintained throughout secondary school, it does not correct for skill shocks during that time. As a result, some students are mismatched in the second period of secondary school ($j = 4$). The LT system avoids this mismatch by making the track allocation later. But this comes at the cost of less aggregate skill accumulation during the C stage. Hence, when students are subject to skill shocks during secondary school, there is a trade-off between the pace of learning in the first stage of secondary school and the quality of the student-track match in the second stage of secondary school.

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

$$\frac{\sigma^2_{\eta_4}}{\sigma^2_{\theta_3}} > 1 + \alpha + \alpha^2 + \beta + \frac{\beta^2}{2} + 2\alpha(1+\beta) + \frac{\gamma^2}{2\pi}\sigma^2_{\theta_3}. \tag{2.26}$$

Proof. In Appendix 2.A.1. □

These results illustrate that the skill technology alone entails non-trivial theoretical implications for the effects of school tracking on end-of-school skills. In particular, even when the track allocation is performed optimally, the timing of tracking balances a trade-off between efficiency gains from learning in more homogeneous peer groups and those from the ability to react to child skill shock realization.

In sum, our parsimonious skill technology can therefore accommodate the ambiguous empirical findings on the effects of tracking on the level of educational achievements, in addition to the estimated association of tracking with higher inequality and disproportional disadvantages among the lower-skilled groups. However, this alone does not allow us to quantify the macroeconomic effects of school tracking policies. Indeed, the quantitative importance of these forces for economic outcomes within and across generations not only depends on the estimates of the child skill technology parameters and the size of the skill shock variances but also on how they interact with other essential features of the model (and reality). For example, second-chance opportunities at the time of the college decision may make the effect of the (early) track choice less consequential for labor market outcomes. On the other hand, asymmetric parental preferences over school tracks may reinforce intergenerational persistence of education, while harming learning efficiency during the school years. Finally, the track decision is not just concerned with purely maximizing skills but takes into account future labor market prospects, which also depend on the share of children attending each track. To quantify these channels through the lens of our model, we now describe the calibration procedure.

2.4 Model Calibration

We calibrate the model to the German Education System (described in detail in Appendix 2.B.1) 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. Table 2.4 summarizes the externally calibrated parameters, and Table 2.5 presents the internally estimated ones.

2.4.1 Data and Sample Selection

The calibration is based on two data sources, and complemented by official statistics on education in Germany.[29] The first source is 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 participants in nationally representative samples in Germany, starting in 2010 (Blossfeld et al.,

[29]See Bildungsberichterstattung (2018) and Appendix 2.B.5 for details on these statistics and the sources of our target moments.

2019).[30] A key component of the information collected is regular standardized assessment tests of school children's competencies in areas such as mathematics, reading, sciences, or grammar.[31] In addition, there is information about school track recommendations by primary school teachers and the final school track choices. We restrict the sample to individual observations containing information on the school and class of a child in a given year.

The second data source is the German Socioeconomic Panel (SOEP), an annual representative household survey from which we use the 2010-2018 waves (Goebel et al., 2019). 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 we do is dropping those workers with hourly wages below the first and above the 99th percentile while keeping both workers and non-workers. 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 skill formation technology, as these constitute the most critical ingredient of our model. Then, we describe the functional forms and estimation strategies for all remaining parameters.

2.4.2 Estimation of the the Child Skill Formation Technology

We specify the empirical analog of the production technology of (the logarithm) of child i's skills that we take to the data as follows:[32]

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

[30] The NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi, Germany) in cooperation with a nationwide network. We use data from Starting Cohorts 2,3, and 4, survey waves 2011-2018 (NEPS Network, 2022).

[31] See also Appendix Section 2.B.3 for more details on the tests as well as the scaling procedure adopted by the NEPS.

[32] Following the work in Cunha et al. (2010), much of the empirical and quantitative literature using child skill formation technologies employs parametric specifications 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. In our formulation, all inputs into child skill formation, and in particular school inputs and parental inputs, are therefore substitutes, which is in line with the literature (Kotera and Seshadri, 2017). We also experimented with relaxing this assumption by including interaction terms between school inputs and parental education, which were, however, insignificant.

Note that (2.27) is a rearranged version of the skill technology (2.2) after substituting in (2.3) and the optimal pace of instruction in each school track as given by Lemma 1.[33] Moreover, in principle, we allow all parameters to be specific to the period j.

In the estimation, we also distinguish between $\bar{\theta}^S_{-i,j}$, which denotes the average skill level of the child i's *classroom* peers, and $\bar{\theta}^S_j$, which refers to the average skill level of all children in a school that belongs to track S. Note that in the model, $\bar{\theta}^S_{-i,j} = \bar{\theta}^S_j$, since we assume a representative school and class per track (or alternatively, identical classes conditional on school tracks). In the data, however, there is heterogeneity across classes, even within schools and tracks. 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 exploit this heterogeneity in the estimation.[34] Finally, the intercept $\omega_{0,j}$ can be a function of age and gender in the empirical estimation, and the parental educational attainment E is a time-constant dummy that equals one if child i comes from a household in which at least one parent is college educated.

As is common in the child skill formation literature (Cunha et al., 2010; Agostinelli and Wiswall, 2016), we think of skills θ_j as latent variables that are only imperfectly measured in the data. Therefore, we employ a log-linear measurement system for latent skills, using a series of achievement test scores as noisy measures of child skills in each period. The identification strategy of the scales and loadings of each measure using their covariances follows Cunha et al. (2010). We aggregate the individual measures into a composite unbiased index using Bartlett factor scores, as in Agostinelli et al. (2023), to account for measurement error. Appendix 2.B.4 details skills measurement and the estimation procedure.

We present our preferred estimates of the skill production technology parameters in Table 2.1 and provide robustness checks with different specifications in Appendix 2.B.4. These estimates are based on the NEPS Starting Cohort 3 data, between school grades 5 and 9, which corresponds to period 3 in our model. Since children are in a comprehensive primary school track before grade 5, we cannot estimate the age-, and track-specific coefficients for period 2. In addition, in grade 12, some parts of the tests are track-specific, which makes the estimates unreliable for period 4. For those reasons, we assume that the estimates of the

[33] The coefficients $\omega_{n,j},\ n = 0, ..., 5$ relate to those in (2.2) and (2.3) as follows: $\omega_0 = \frac{\beta^2}{2\delta}, \omega_1 = (\kappa + \frac{\beta}{\gamma}\delta), \omega_2 = \alpha, \omega_3 = -\omega_4 = \frac{\gamma^2}{2\delta}$, and $\omega_5 = \zeta$ for all j. We formally test the restriction $\omega_3 = -\omega_4$ after the estimation.

[34] Given that we control for school fixed effects in the estimation, our identification of the direct peer effects is therefore close to the literature on estimating peer effects using classroom-fixed-effects methods (see the discussion in Epple and Romano, 2011). This also has the added benefit that we can identify a model that includes $\bar{\theta}^S_{-i,j}, (\bar{\theta}^S_j)^2$, and the interaction $\theta\bar{\theta}^S_j$, even if we consolidate schools into a maximum of two school tracks in the data, which, as discussed in Appendix 2.B.1 resembles reality in Germany over the past decade.

Table 2.1: Child Skill Technology Parameters Estimates

Dependent Variable: $\theta_{i,j+1}$ Grade 9 on Grade 5		
Coefficient	Variable	
$\hat{\omega}_{1,3}$	$\theta_{i,j}$	0.664***
		(0.022)
$\hat{\omega}_2$	$\bar{\theta}^S_{-i,j}$	0.003
		(0.020)
$\hat{\omega}_3$	$\theta^2_{i,j}$	0.008*
		(0.004)
$\hat{\omega}_4$	$(\theta_{i,j} - \bar{\theta}^S_j)^2$	-0.011*
		(0.006)
$\hat{\omega}_{5,3}$	$E = 1$	0.034***
		(0.010)
Obs.		1,847

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. Standard errors are clustered at the school level. We control for year of birth, gender, and school-fixed effects. Source: NEPS.

skill technology parameters ω_2, ω_3, and ω_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 the logarithm of child skills. Hence, we can interpret the coefficients as elasticities. Thus, $\hat{\omega}_1 = 0.66$ means that a 1% increase in latent skills at the beginning of primary school is associated with a 0.66% increase in end-of-primary school skills. This own-skill productivity is close to the literature's common values (see estimates in Cunha et al., 2010; Agostinelli et al., 2019). During secondary school, the estimated coefficient $\hat{\omega}_2$ is positive but rather small and statistically insignificant. Existing estimates of linear-in-means peer effects models range from small negative effects to large positive effects of a one-unit increase in average peer test scores on student achievement.[35] Translating our estimates into such an effect, we find that a one-unit increase in average peers' test scores raises own future tests by around 0.01. As such, we are at the lower end of typical estimates during primary and secondary school, which is in line with other research that uses within-school classroom variation (see Epple and Romano, 2011) that typically arrive at lower estimates compared to studies that use some form of random assignment of peers. Finally, the estimated coefficient $\hat{\omega}_4$ is negative and statistically significant at 10%. It indicates that a 1% increase in the squared distance to the average skill level in a track is associated with an up to 0.011% 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. Importantly, we cannot reject the hypothesis that $\hat{\omega}_3 = -\hat{\omega}_4$, which is in line with our assumptions.[36]

The parameters we use in the skill formation technology in the model are then $\omega_2 = \hat{\omega}_2$, and $\omega_4 = \hat{\omega}_4$ as reported in Table 2.1. Moreover, we set $-\omega_3 = \omega_4 = \hat{\omega}_4$. The parameters $\omega_{1,3}$ and $\omega_{5,3}$ also come from Table 2.1, while $\omega_{1,2}$, $\omega_{1,4}$, $\omega_{5,2}$, and $\omega_{5,4}$ are estimated internally to match the own-skill elasticities from a regression of future skills on past skills and parental education, as reported in Table 2.2. Finally, the constant parameter ω_0 is set to zero.

[35] Table 4.2. in Sacerdote (2011) provides an overview about existing estimates using a variety of identification strategies.

[36] The estimated negative effect $\hat{\omega}_4$ is therefore conform with findings in the literature, which test the effects of skill-based tracking on later achievement directly (for example Duflo et al. (2011) argue that the large achievement gains of tracked students relative to non-tracked students are the result from indirect effects of peers that operate through the adjustment of teaching behavior) or test the effects of classroom heterogeneity on achievement. As summarized in Sacerdote (2011), many, but not all, findings in this literature point to the fact that classroom heterogeneity reduces test scores, which is consistent with the idea that tracking raises outcomes in both tracks. In terms of effect sizes, it is difficult to compare our estimates to existing ones as we are measuring the heterogeneity across tracks directly and not across classrooms.

Table 2.2: Evolution of Child Skills

Grade $\theta_{i,j+1}$ on Grade $\theta_{i,j}$ and E				
Dependent Variable:		Grade 4 (Cohort 2)	Grade 9 (Cohort 3)	Grade 12 (Cohort 4)
Panel A: All students				
	$\theta_{i,j}$	0.649***	0.811***	
		(0.011)	(0.016)	
	$E = 1$	0.072***	0.044***	
		(0.007)	(0.009)	
Obs.		4,023	2,070	
Panel B: Academic students				
	$\theta_{i,j}$	0.566***	0.745***	0.825***
		(0.019)	(0.025)	(0.019)
	$E = 1$	0.049***	0.035***	0.033***
		(0.011)	(0.012)	(0.009)
Obs.		1,371	1,195	2,327

Notes: This table presents the coefficients of regressions of current skills on past skills and parents' education dummy. Standard errors are clustered at the school level. Models control for year of birth, gender, and school-fixed effects. Source: NEPS.

2.4.3 Remaining Parameters

Preferences

We set the inverse elasticity of intertemporal substitution to $\sigma = 2$, a value that is common in the literature. The Frisch elasticity of labor supply is set to 0.5. The disutility shifter b is estimated internally to match the average time worked in the SOEP data, which is 0.36 when the total time available after sleep and self-care is assumed to be 13 hours on a weekday and normalized to 1.

We internally calibrate the time discount factor β, so the equilibrium interest rate amounts to 4% annually. The altruism parameter Λ is calibrated such that the ratio of average inter-vivos transfers to average labor income in the model corresponds to average higher education costs of children to average four-year labor income in the data. According to a 2016 survey by the German Student Association, the monthly costs of living during the higher education stages for a student without children are, on average, 830 Euros per month (Dohmen 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 four years (the length of time it takes on average to complete higher education studies). Then, the ratio of total costs to average 4-year labor income is approximately 0.49, which we take as our target moment.

Academic School Track Costs

The stochastic school track costs $\chi(E)$ are assumed to follow the distribution $\chi(E) \sim H^E(\chi) \equiv \mathcal{N}(\mu_{\chi,E}, \sigma_\chi^2)$. We parameterize the mean $\mu_{\chi,E}$ as follows:

$$\mu_{\chi,E} = \mu_{\chi,A} + \begin{cases} \chi_1 & \text{if } E = 1 \\ \chi_0 & \text{if } E = 0, \end{cases} \tag{2.28}$$

so that $\mu_{\chi,A} > 0$ represents a uniform utility cost of academic-track attendance (for example, stemming from the academic track being more demanding and psychologically taxing), and the parameters χ_0 and χ_1 represent asymmetric preferences or costs for the academic track by parental college education. We calibrate χ_0 and χ_1 to match the share of deviations from secondary school track recommendations by parental education in the data, while $\mu_{\chi,A}$ is calibrated to match the overall share of academic track recommendations (0.44).[37] The variance of the track tastes σ_χ^2 is calibrated to match the variance of the residuals coming from a regression of school track on end-of-primary-school skills, which is 0.166.

Initial Child Skills, and Child Skill Shocks

The transmission of initial ability ϕ, which equals the initial child skill level, across generations follows an AR(1) process with persistence coefficient ρ_ϕ and variance σ_ϕ^2. Since the initial ability is designed to capture any residual correlation in economic outcomes across generations, we calibrate it to match the intergenerational elasticity of incomes in Germany. Kyzyma and Groh-Samberg (2018) estimate an elasticity between the income rank of individual labor earnings between children and parents using the SOEP data of 0.24, which we take as our target statistic.[38] The variance σ_ϕ^2 is then estimated to match the variance of

[37]Primary school teachers typically give these recommendations before the transition to secondary school. They are based on both a reflection of the child's achievement during primary school and the teachers' assessment of the academic potential and success probability of the child in an academic track school. Thus, we argue that the recommendations are forward-looking and, since primary school teachers typically observe the children over multiple years every day during the week, based on a similar information set as the parents. Therefore, we consider the recommended school track in the model as the one a parent would have chosen without any specific school track taste ($\chi_1 = \chi_0 = 0$). Then, deviations from that unbiased track choice by parental education map into deviations from teacher recommendation. Details on these moments are given in Appendix 2.B.2.

[38]As is common in the literature, Kyzyma and Groh-Samberg (2018) compute the correlation of income ranks using average labor earnings over five years. We compute rank-rank correlations of four-year labor income, according to the period length of our model, and then compare the ranks of 30-34-year-old children to those of their parents when they were 46 to 50 years old, which is similar to the sample used by Kyzyma

pre-school skills in the data, which we normalize to 0.1.

As discussed in Section 2.3, the size of shocks to child skills has important implications for the effects of school tracking policies as they can give rise to efficiency losses from early tracking. 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 skill percentile rank across periods. 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 deterministic components of the skill formation technology or by track choices.

College Costs

We parameterize the "psychic" college cost function following Daruich (2022):

$$\begin{aligned} \psi(S, \theta_5, \nu(E^p)) &= \exp(\psi_0 + \psi_{S=V} + \psi_\theta \theta_5 + \nu(E^p)) \\ \nu(E^p) &\sim G^{E^p}(\nu) \equiv \mathcal{N}(\mu_{\nu,E^p}, \sigma^2_\nu). \end{aligned} \tag{2.29}$$

We estimate the two parameters ψ_0 and $\psi_{S=V}$ 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 who obtain a college education. We discipline the coefficient ψ_θ that multiplies end-of-school skills by matching the regression coefficient on test scores from a regression of a college graduation dummy on end-of-school test scores, controlling for the secondary school track.

We calibrate the two parental education-specific means of the college taste shock parameters to be symmetric deviations from 0, such that $\mu_{\nu,E^p=1} = \Delta(\mu_{\nu,E^p})$ and $\mu_{\nu,E^p=0} = -\Delta(\mu_{\nu,E^p})$ to match the ratio of the share of children from college-educated parents who themselves go to college (0.63) and the share of children from non-college-educated parents who go to college (0.20) in the data. Finally, we calibrate the variance of these shocks, σ^2_ν, to match the variance of the residuals from the above regression of college education on end-of-school skills and school track.

The final component of college costs is not a part of the "psychic" costs 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.[39] Thus, we set the maximum remaining time during the higher education stage to

and Groh-Samberg (2018). As in Lee and Seshadri (2019), we normalize average labor income across the entire working population to be one in the data and in the model. In the latter, we do this by setting the technology parameter A in the firm production function.

[39] A standard 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

$\bar{n}(E = 1) = 0.40$.

Human Capital Growth

We estimate the deterministic human capital growth profiles for both types of education, $\gamma_{j,E}$, for $j = 5, ..., 16$, using wage regressions in the SOEP data, following the approach in Lagakos et al. (2018).[40] The resulting experience-wage profiles for four-year experience bins are shown in Table 2.3, 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 (normalized) labor income across the working-age population in the data, which is around 0.86.

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 college and non-college human capital in the firm production equals 1.5 (Ciccone and Peri, 2005). The weight on non-college human capital in the CES aggregator, φ, 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. The TFP parameter A is calibrated such that the model produces average earnings of 1.

Regarding the tax and transfer system, 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 $\tau_a = 0.25$, corresponding to the final withholding tax rate on realized capital gains, interest, and dividends in Germany. The size of the lump

Germany is eight semesters or four years.

[40] Concretely we create, separately for each education group, four-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

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

We assume no experience effect on wage growth in the last eight years of work to disentangle time from cohort effects, following the HLT approach in Lagakos et al. (2018).

Table 2.3: 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

sum government transfers is set to $g = 0.06$, which in equilibrium amounts to 6% of average labor earnings. Finally, we set pension benefits to $\pi_j(h_{17}, E) = \Omega h_{17} w_E$ during retirement and calibrate the scale parameter Ω internally, such that the average replacement rate corresponds to 40% (Mahler and Yum, 2023).

2.4.4 Method of Simulated Moments Estimation Results

In total, we calibrate 26 parameters internally using the method of simulated moments to match 26 target data moments. The parameters, their estimated values, model-implied moments, and target data moments are presented in Table 2.5.

The model 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 college graduates in the simulated economy is 35%, which is in line with the German data in the 2010s. The share of children in an academic track school is 44%. The model also matches the transition rates from academic and vocational secondary school into college (at around 66% and 11%) and the effect of secondary school skills on college attendance, while it slightly overestimates the share of college graduates from non-college households.

Parental school track preferences significantly affect the school track decision, both in the model and the data. In particular, around 23% of college-educated parents overrule a recommendation for their child to go to a vocational track school, while 16% of non-college parents overrule an academic track recommendation in favor of a vocational track school.

Table 2.4: 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	HH Equiv. Scale	Jang and Yum (2022)
$\bar{n}(E=1)$	0.40	Time Cost of College	40 hours/week for 4 years
Firm			
σ_f	1/3	E.o.S (H_0, H_1)	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	0.25	Capital Tax Rate	Tax Rate on Capital Gains in Germany
g	0.06	Lump-sum Transfers	6% of Annual Labor Income

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

To match the correlation between child skill ranks across school periods, the model requires large child skill shocks, especially during primary school, with a standard deviation of 0.052. The estimated own skill elasticity increases between primary school ages (0.65) and end-of-secondary school ages (0.81). At the same time, the parental education intercept in the child skill technology decreases from 0.072 to 0.032.

2.4.5 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 2.6. The first set of moments pertains to child skills. Our model features slightly smaller differences in average child skills by parental education and comparable differences in average skills by school track. In both data and model, these differences increase between primary and secondary school, before staying relatively constant.[41] In general, differences

[41]See for instance Passaretta et al. (2022); Nennstiel (2022); Schneider and Linberg (2022) who investigate

Table 2.5: Internally Calibrated Parameters

Parameter	Value	Description	Target	Data	Model
Preferences					
β	0.935	Discount Factor	Annl. Interest Rate	0.04	0.04
b	20.7	Labor Disutility	Avrg. Labor Supply	0.36	0.36
Λ	0.31	Parental Altruism	Transfer/Income	0.49	0.49
School Track Tastes					
$\mu_{\chi,A}$	0.048	Uniform A-Track Costs	Share A-Track Recommend.	0.44	0.44
χ_0	0.0020	Mean A-Track Cost if $E=0$	Share of Dev. from A if E = 0	0.16	0.16
χ_1	-0.0036	Mean A-Track Cost if $E=1$	Share of Dev. from V if E = 1	0.23	0.23
σ_χ	$0.17 \cdot 10^{-3}$	Std. A-Track Cost Shock	Reg. S on θ: var(residuals)	0.166	0.168
Child Skill Technology					
$\omega_{1,2}$	0.65	Own Skill Elasticity ($j=2$)	Reg. θ_3 on θ_2 & E: coef. θ_2	0.649	0.649
$\omega_{5,2}$	0.072	Coefficient on E ($j=2$)	Reg. θ_3 on θ_2 & E: coef. E	0.072	0.072
$\omega_{1,4}$	0.81	Own Skill Elasticity ($j=4$)	$S=1$, Reg. θ_5 on θ_4 & E: coef. θ_4	0.825	0.812
$\omega_{5,4}$	0.032	Coefficient on E ($j=4$)	$S=1$, Reg. θ_5 on θ_4 & E: coef. E	0.033	0.032
Transmission of Initial Skills (Ability)					
σ_ϕ	0.032	Std. of Intergen. Shock	Variance of initial skills	0.10	0.12
ρ_ϕ	0.9	Persistence of Ability	IGE (income rank)	0.24	0.23
College Costs					
ψ	0.77	Intercept	Share in CL from A-Track	0.66	0.65
ψ_V	0.16	Add. Costs for V-Track	Share in CL from V-Track	0.11	0.11
ψ_θ	-0.35	Coefficient on θ_5	Reg. E on θ_4 & S: coef. θ_4	0.40	0.50
$\Delta(\mu_{\nu,E^p})$	0.034	Diff. in Means by E^p	Share in CL from Non-CL HH	0.20	0.28
σ_ν	0.008	Std. Taste Shock	Reg. E on θ_4 & S: var(residuals)	0.137	0.138
Idiosyncratic Shocks					
σ_ε	0.011	Std. Market Luck Shock	Std(Log Labor Income)	0.86	0.84
σ_{η_3}	0.052	Std. Learning Shock $j=3$	Rank$_{j=2}$-Rank$_{j=3}$	0.72	0.73
σ_{η_4}	0.030	Std. Learning Shock $j=4$	Rank$_{j=3}$-Rank$_{j=4}$	0.79	0.80
σ_{η_5}	0.032	Std. Learning Shock $j=5$	Rank$_{j=4}$-Rank$_{j=5}$ if $S=1$	0.74	0.75
Miscellaneous					
Ω	0.1	Pension Anchor	Replacement Rate	0.40	0.40
A	3.31	TFP	Avrg. Labor Earnings	1.0	1.0
φ	0.543	Weight Non-CL H_0	College Share	0.35	0.35

Notes: This table presents the internally calibrated parameters, targeted moments, and their model-generated counterfactuals. See Appendix 2.B.5 for details on the sources of the target moments.

across school tracks are larger than differences across parental education. Our model also produces realistic child skill rank-rank correlations *within* school tracks.

Table 2.6: Non-targeted Moments

Moment	**Data**	**Model**
Child Skill Moments		
Differences in average skills by parental education (in standard deviation)		
Primary School	0.53	0.44
Beginning Secondary School	0.66	0.53
Middle Secondary School	0.71	0.54
Differences in average skills by school track (in standard deviation)		
Beginning Primary School*	0.84	0.80
Beginning Secondary School	1.10	1.14
Middle Secondary School	1.11	0.95
Rank-rank coefficients		
$\text{Rank}_{j=2} - \text{Rank}_{j=3}$ if S = 1*	0.62	0.66
$\text{Rank}_{j=3} - \text{Rank}_{j=4}$ if S = 1	0.68	0.73
$\text{Rank}_{j=2} - \text{Rank}_{j=3}$ if S = 0*	0.64	0.67
$\text{Rank}_{j=3} - \text{Rank}_{j=4}$ if S = 0	0.74	0.74
Skill evolution during secondary school		
Reg. θ_4 on θ_3 and E: coef. θ_4	0.81	0.66
Reg. θ_4 on θ_3 and E: coef. E	0.04	0.04
Intergenerational Mobility and Inequality		
Parental Income Gradient (Dodin et al., 2024)	0.52	0.32
Q5/Q1 A-track on income (Dodin et al., 2024)	2.13	1.82
Q1 A-track on income (Dodin et al., 2024)	0.34	0.30
Gini Coefficient of Income	0.29	0.26
College Wage Premium	1.35	1.46

Notes: This table presents non-targeted moments and their model-generated counterfactuals.
* We exploit the panel structure of the datasets and group students by future school track assignation.

The second set of moments relates to further measures of intergenerational mobility and cross-sectional inequality. 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. (2024). Using a different data set than we, they regress academic-track school graduation of a child on the percentile income rank of her parents, finding that a ten-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 3.2 percentage point increase. Moreover, Dodin et al. (2024) 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. Our model-generated data squares well against

the NEPS data and find stable or growing socioeconomic status gaps in children's skills.

these external estimates (30% and 1.82, respectively).

We also investigate the model fit regarding the non-targeted determinants of the school track choice in relation to parental education and end-of-primary school skills. We delegate this discussion to Section 2.5.1, where we decompose the track determinants quantitatively.

Long-term effects of Track Choice for Marginal Students

Dustmann et al. (2017) analyze 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 born just before the cut-off age are less likely to go to an academic track secondary school simply because they are younger and, therefore, less developed than their class peers at the time of the track decision. 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.[42]

We use our model-simulated data to perform a similar exercise. In particular, we compare the later-in-life outcomes of children who are very similar in terms of their state variables at the time of school track choice but end up going to different school tracks. To that end, we calculate the average present values of lifetime income and lifetime wealth conditional on all states prior to entering secondary school—parental human capital, assets, education, and initial ability and skills—of children who go to an academic track school and children who go to a vocational track school.[43] Conditional on all other states, differences in the track allocation can only arise due to the stochastic track utility shock, which we can also interpret as arising from age-at-school-entry effects.

We find that going to the academic track instead of the vocational track is associated with a 6.6% higher present value of lifetime labor income, and a 4.4% higher present value of lifetime wealth for these, otherwise very similar children. While not zero, these differences seem relatively small in relation to overall inequality in these outcomes. For example, the 6.6% higher present value of lifetime labor income is around 1/10th of a standard deviation of lifetime labor income. Moreover, in our model, the track choice is only between one vocational

[42]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.

[43]Concretely, we partition all continuous states into 10 groups of equal size each. 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.

and one academic track, whereas Dustmann et al. (2017) consider three tracks, of which two can be classified as vocational. We would generally expect children at the margin of these two vocational tracks to show fewer differences in lifetime outcomes. In sum, we conclude that the implications of model with respect to the effect of tracking on *marginal* children are not at odds with the reduced-form evidence presented in Dustmann et al. (2017).

2.5 Quantitative Results

Our model allows us 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 quantify the main determinants of the secondary school track and the importance of skill accumulation during secondary school for lifetime inequality. We then use the model to study the effects of reforms of the timing of school tracking. Finally, we perform counterfactual analyses of economies in which we reduce the parental influence on the school track choice.

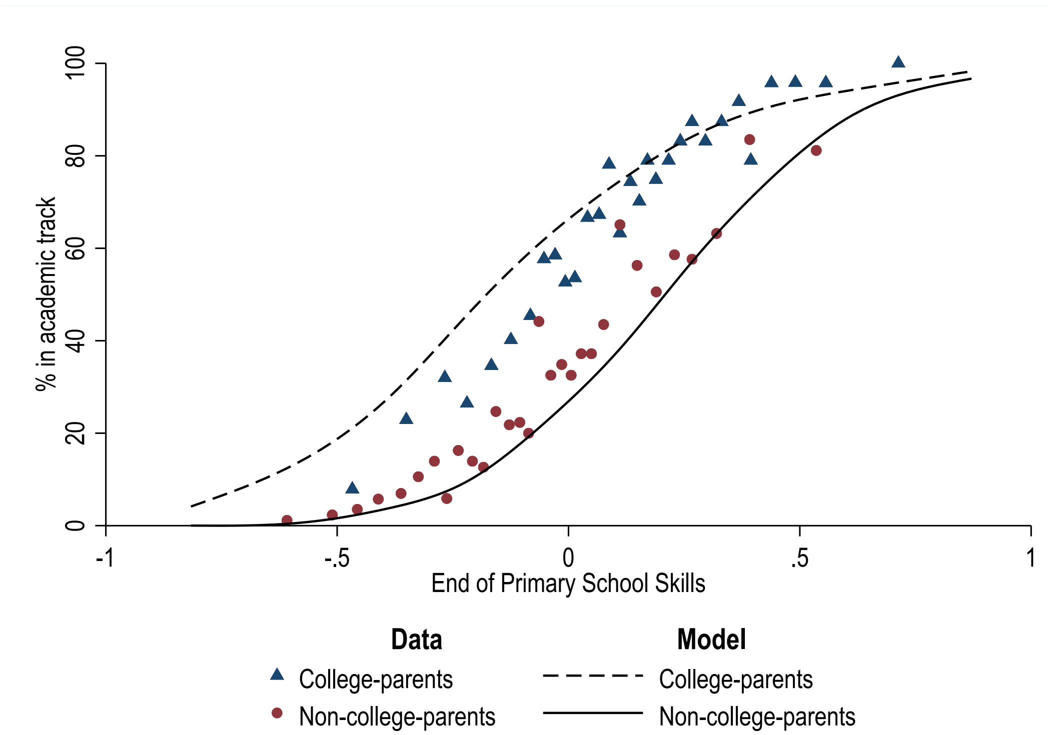
2.5.1 The School Track Choice and Sources of Lifetime Inequality

Determinants of the School Track Choice

Our calibrated model predicts that children's skills largely determine the school track choice. Figure 2.2 shows the relationship between skills and the academic track choice, separately by parental education. The model-generated data matches remarkably well the increasing, S-shaped probability of academic-track attendance in skills observed in the NEPS data. Parental education is another important independent driver of the school track choice as can be seen in Figure 2.2. Even for the same end-of-primary school skills, children from college-educated parents are significantly more likely to go to an academic track school than children from non-college parents, both in the data and the model.

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 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, ω_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 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

Figure 2.2: Probability of attending the Academic School Track



Notes: This figure shows the share of children attending the academic school track as a function of their skills. The triangles and dots are data moments and stand for children from college and non-college backgrounds, respectively; the baseline model simulated analogs are in dashed and solid lines. Data source: NEPS, cohort 3. All observations are weighted so that the shares of children in each track correspond to the targeted ones.

equal, increases the likelihood of college admission. Thirdly, even net of college tastes, parents face asymmetric academic track utility costs $\chi(E)$.

Table 2.7: School Track Choice Determinants

	Dependent Variable: $S = A$			
	(1)	(2)	(3)	(4)
	Baseline	$\omega_{5,j=3,4} = 0$	$\Delta(\mu_{\nu,E^p}) = 0$	$\chi_0 = \chi_1 = 0$
θ_3	0.78	0.79	0.86	0.82
$E = 1$	0.42	0.38	0.18	0.32

Notes: This table reports the coefficient estimates of regressions of an academic school track dummy on beginning of primary school skills and parental education, controlling for all other states at the time of the tracking decision and a constant. 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 (Columns (2)-(4) in Table 2.7), in which we isolate each effect by setting to zero the parental education effect parameter $\omega_{5,j=3,4}$, the means in college taste shocks across parental education $\Delta(\mu_{\nu,E^p})$, or the asymmetry in academic track costs χ_E.[44]

In all cases, the coefficient on parental education drops, and the coefficient on skills before the track decision increases relative to the baseline economy (in Column (1)). The magnitude of the effects, however, varies across the counterfactual scenarios. While shutting down the parental education parameter has little effect (Column (2)), shutting down the college taste shocks across parental education approximately halves the coefficient on parental education (Column (3)). Asymmetric academic school track costs also matter as shutting them down reduces the coefficient on parental education by around 24% (Column (4)).

Sources of Lifetime Inequality

In the spirit of Huggett et al. (2011) and Lee and Seshadri (2019), we can decompose how much of the variation in lifetime economic outcomes of our model agents can be explained by

[44] 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. In Column (2) in Table 2.7, we isolate the effects of the first channel by solving the model with $\omega_{5,j=3,4} = 0$ yet leaving $\omega_{5,j=3,4} > 0$ 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.

various factors at various ages. As before, we focus on (the present value of) lifetime labor income and wealth as our economic outcomes of interest.

Table 2.8: Contributions to Lifetime Inequality

Life Stage	States	Share of Explained Variance Lifetime Earnings	Lifetime Wealth
Independence (age 18)	$(S, \phi, h_5, a_5, E, E^p)$	69%	65%
	(S, ϕ, h_5, E^p)	60%	61%
	(S, ϕ, a_5, E, E^p)	49%	42%
School Track Choice (age 10)	$(S, \phi, \theta_3, h_{11}, a_{11}, E)$	30%	33%
	(S)	16%	15%
Pre-Birth (parent age 30)	(E, ϕ, h_8, a_8)	14%	21%

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

Row 1 of Table 2.8 summarizes that 69% of the variation in lifetime labor income can be accounted for by all states at the age of 18. These states are the school track in secondary school S, human capital h_5, transfers received from the parent a_5, the college choice E, parental college education E^p, and the initial ability ϕ. In terms of lifetime wealth, this number is around 65%.[45] Thus, our model suggests that lifetime outcomes are already largely predetermined when agents become independent and can enter the labor market. Note that all uncertainty regarding school skills has resolved and the college decision has been made at this stage. The remaining unresolved uncertainty over human capital (market luck) shocks during the working years has, therefore, more minor effects on lifetime inequality. The explained share of variation in lifetime outcomes remains high if we only condition on the states before the college decision has been made and the inter-vivos transfers have been realized (Row 2). This suggests that these states are not major sources of lifetime inequality. On the contrary, if we only exclude human capital h_5 (Row 3), the share of explained variance in lifetime earnings drops by 20 percentage points, and the share of explained variance in lifetime wealth by 23 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.

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 a third of lifetime earnings and wealth variation is explained (Row 4). Yet the explained share is significantly smaller than after school, suggesting that the learning

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

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 16% of lifetime earnings variation and 15% of lifetime wealth variation. However, this should not be interpreted as 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.

The last row of Table 2.8 shows the contribution of parental states prior to the birth of their children to their children's lifetime outcomes. At this stage, all uncertainty regarding child skills and human capital has not yet been realized (i.e., ϕ denotes the parent's ability). Around 14% 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 higher at 21%, pointing to the critical role of wealth transfers. For example, using the same decomposition of the unconditional variance of transfers into parental states pre-birth, we find that more than a third of the variation in transfers (35%) is predetermined before the child's birth. In contrast, only around 23% of the variation in human capital at age 18 is predetermined before birth, highlighting the role of the schooling years and shocks in shaping adult human capital.[46]

2.5.2 The Timing of School Tracking

In countries with an early tracking system, such as Germany, it is often argued that postponing the tracking age will improve social mobility without incurring efficiency losses. 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 intergenerational consequences or welfare effects of a large-scale reform that changes the timing of school tracking.

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 ten to fourteen or abolish tracking during secondary school altogether. In each experiment, we assume that in the periods preceding tracking, all children attend a school that belongs to a comprehensive school track, just like during primary school in $j = 2$. All parameters, including those governing school academic track costs and college costs, remain the same as

[46]In comparison to Lee and Seshadri (2019) in the U.S. case, our estimated contribution of parental states prior to the birth of a child to her eventual lifetime outcomes is somewhat smaller (in particular they find that almost half of the lifetime wealth variation is pre-determined at that stage. These differences may reflect that firstly, intergenerational mobility estimates in Germany tend to be smaller than in the U.S. Secondly, we incorporate explicitly the uncertainty in child skill realizations over the childhood years, while Lee and Seshadri (2019) focus on endogenous parental investments that could explain in particular the large explanatory power of pre-birth parental states for child human capital.

in the baseline economy.[47] We then compare steady-state equilibrium outcomes, which can be considered long-run outcomes of the policy change.

We present the effects of the counterfactual experiments on aggregate, distributional, and social mobility outcomes in Panel A of Table 2.9. In addition, we calculate the relative changes in average welfare, defined as the percent change in consumption that a newborn in the baseline economy would require in every period to be equally well off as in the policy counterfactual. Following the literature, we calculate this consumption equivalence welfare measure under the veil of ignorance, meaning that all policy functions remain unchanged.[48]

Postponing School Tracking by Four Years

Columns (1) and (2) present the results of postponing tracking from age ten to age fourteen, corresponding to the average tracking age in OECD countries (OECD, 2020b). In Column (1), wages (w_0, w_1) and the interest rate r remain at the same values as in the baseline case. That is, we compare the partial equilibrium outcomes of the policy counterfactual. In Column (2), prices adjust; that is, we compare the general equilibrium outcomes of the policy counterfactual. As before, the instruction paces during all school stages are set to the level that is optimally chosen by a policymaker given the allocation of children across tracks.

We find an efficiency-equity trade-off of postponing tracking in general equilibrium but not in partial equilibrium. In partial equilibrium, Column (1), both aggregate output Y and aggregate human capital H increase by 0.2%.[49] At the same time, cross-sectional inequality, as measured by the Gini coefficient of labor income, drops by 0.4%. Similarly, the college wage premium decreases, and the ratio of the 90th to 10th percentile of income decreases. Mobility is improved, as indicated, for example, by the intergenerational elasticity of income, which drops by 3.5%. In a similar vein, the dependence of going to an academic track school on parental income drops by 15%. These effects translate into an improvement in average welfare from postponing tracking, in the range of 0.18% consumption equivalent units.

In contrast, the model-predicted effects of postponing tracking change, once we allow for the adjustment of wages on the labor market and, therefore, general equilibrium effects of the human capital changes in the economy. Column (2) of Table 2.9 reports that, while the gains in terms of inequality and social mobility persist, albeit at a smaller level, the effects on aggregate human capital and output reverse as both decrease by 0.1%. Our quantitative

[47] In the case of no tracking, we assume that the fixed college utility costs $(\psi + \psi_v)$ are a weighted average of the baseline economy.

[48] Appendix Section 2.A.3 provides our welfare definition.

[49] Given that aggregate production is Cobb-Douglas in both physical capital and human capital, this implies that also aggregate physical capital increases.

Table 2.9: Timing of Tracking Counterfactual Experiments - Results

	Changes in %		
	(1)	(2)	(3)
Economy	PE	GE	GE
Tracking Age	14	14	Never
Panel A - Aggregate, Distributional and Intergenerational Outcomes			
Efficiency			
Output (Y)	+0.2	-0.1	-0.2
Human Capital (H)	+0.2	-0.1	-0.4
Cross-sectional Inequality			
Gini of earnings	-0.4	-0.4	-0.8
College wage premium	-4.0	-0.2	-2.8
90th/10th percentile of income	-0.1	-0.4	-0.8
Mobility			
Intergenerational income mobility ($-$income rank-rank coef.)	+3.5	+2.2	+23.9
Parental income on academic track (Dodin et al., 2024)	-15	-6.8	-
Welfare (CEV)	+0.18	-0.05	-0.08
Panel B - Educational Outcomes			
% Academic track	+5.4	+2.0	
... if college parents	+2.2	+1.5	
... if non-college parents	+6.8	+2.4	
% College	+3.9	-0.3	-0.2
... if college parents	+2.9	+0.5	-18.1
... if non-college parents	+3.4	-0.9	+16.7
... if academic track	-0.4	-1.7	
... if vocational track	+3.0	+0.5	
Average end-of-school skills ($\bar{\theta}_5$)	+4.3	-1.6	-2.7
Average middle-of-school skills ($\bar{\theta}_4$)	+7.5	-2.1	-1.8
Variance of end-of-school skills ($Var(\theta_5)$)	-0.3	-0.2	-2.1
Variance of middle-of-school skills ($Var(\theta_4)$)	-0.7	-0.9	-2.7
Correlation between academic track and initial skills	-20	-14	
Correlation between end-of-school skills and initial skills	-0.5	-0.1	-3.3
Correlation between college graduation and initial skills	-18	-12.6	-69.5
Correlation between college parents and end-of-school skills	-6.0	-6.0	-26.2
Correlation between college graduation and end-of-school skills	-4.1	-3.5	-12.9

Notes: This table presents changes in outcomes in % due to postponing the school tracking choice by four years (from the age of ten to the age of fourteen) or abolishing tracking altogether. Column (1) displays percentage changes due to postponing tracking in partial equilibrium, that is, if prices are unchanged. Column (2) shows the effects of postponing tracking in general equilibrium. Column (3) presents the effect of abolishing school tracking in general equilibrium.
Intergenerational mobility is measured as the negative of the income rank-rank coefficient.

results therefore indicate that postponing tracking incurs a trade-off between equality and mobility improvements on the one hand, and aggregate efficiency losses on the other hand, once general equilibrium effects are taken into account.[50] Moreover, average welfare when measured in terms of consumption equivalent units slightly decreases relative to the early tracking benchmark.[51]

Understanding the sources of the efficiency-mobility trade-off

In our model, aggregate efficiency in terms of output and human capital is driven by the level of skills learned during school as they translate into adult human capital. As we argued in Proposition 3 in Section 2.3, the effect of postponing tracking on end-of-school skills for *one cohort* of children is theoretically unclear even when the track decisions are made optimally, and depends in particular on the degree of uncertainty about the skill evolution. On top of that, over multiple generations, aggregate human capital also depends on the share of college-educated workers in the economy. This is because, on the one hand, college-educated workers mechanically experience steeper productivity growth over their working career (through $\gamma_{j,E}$), and on the other hand, college-educated parents provide higher inputs into the skill development of the next generation of children, which increases end-of-school skills and human capital.[52]

In partial equilibrium, when the college wage premium remains high, our model predicts that these effects lead to efficiency gains. As indicated in Panel B of Table 2.9, the share of college parents in the new steady state increases by 4.1%, and so does the share of children in academic track schools (+5.4%) and average end-of-school skills (+4.3%). In fact, because of the higher parental inputs, average child skills already before tracking at age ten are higher in the partial equilibrium late tracking case compared to the early tracking economy. This can be seen in Figure 2.3, where we plot the evolution of average child skills in the early tracking economy (red), the late tracking economy in partial equilibrium (green), and the late tracking economy in general equilibrium (blue).

[50] This result can be viewed in a similar spirit to the efficiency-mobility trade-off in Bénabou (1996), who has shown that policies aimed at improving mobility may entail penalties in terms of growth, or more recently in Arenas and Hindriks (2021), who argue that more equal school opportunities by parental income raises social mobility but come at the cost of modest efficiency losses in terms of human capital.

[51] It should be noted that the standard consumption equivalent welfare measure used by us and in the related literature does not take into account improvements of intergenerational mobility that occur across cohorts. Rather, the standard welfare measure (see definition in Appendix Section 2.A.3) only captures the trade-off between efficiency and redistribution within cohorts. Whether our welfare conclusions regarding a postponement of the school tracking age hold also if a planner takes into account mobility is an interesting question that requires future research.

[52] Labor supply can also affect aggregate human capital, but it stays approximately constant across the policy counterfactual experiments.

In the general equilibrium case, however, the college share remains approximately at its baseline level as college wages (in efficiency units) adjust downwards and non-college wages upwards. For that reason, parental inputs are approximately the same as in the baseline, early tracking case, which results in very similar average child skills at age 10 in steady state (see Figure 2.3). Our calibrated model then predicts that postponing tracking leads to losses in average skills. Concretely, average skills in period $j = 4$, that is right before late tracking, drop by 2.1%, and end-of-school skills, drop by 1.6%, on average. As explained in Section 2.3, these learning losses intuitively arise from the prolonged period of comprehensive school, during which instruction becomes less efficient. Moreover, our model predicts that these losses cannot be recuperated by learning efficiency gains that arise when more uncertainty about child skills is resolved in the late tracking case (see Figure 2.3).[53] The learning losses from longer comprehensive schooling also serve as an explanation for the raised incentive to send a child to an academic track school (+2% relative to early tracking), where average peer skill levels are higher, which partially compensates for less efficient learning.[54]

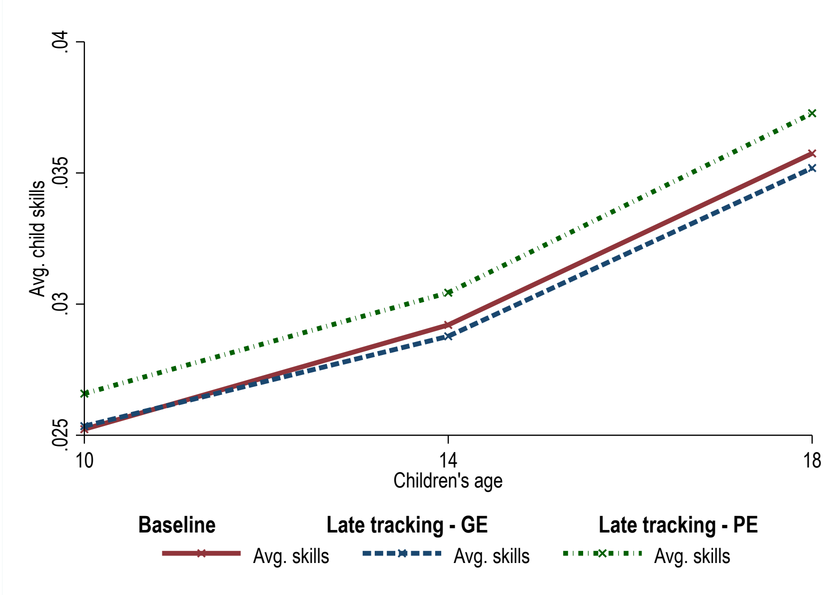
The effects of later tracking on inequality and mobility are fundamentally also rooted in the consequences of the policy change on the skill distribution. As reported in Panel B of Table 2.9, one more model period of comprehensive school decreases the overall heterogeneity in skills in the middle of secondary school (i.e. $Var(\theta_4)$ drops by 0.9%.). Intuitively, this is because children who would have gone to a vocational track school in the early tracking economy are now exposed to, on average better peers, while children who would have gone to an academic track school are now surrounded by, on average, lower skills.[55] On top of that, one more period of comprehensive school in the late tracking case harms relatively more children from college-educated households as they would have been more likely to go to an academic track school and benefits relatively more children from non-college households, who would have been more likely to go to the vocational track. Moreover, these children are more likely to occupy the center of the skill distribution, who, as we have argued before, are the children that gain most from comprehensive schooling. As a result, differences in

[53]The skill growth between the middle and end of secondary school in both the early and late tracking cases is quite similar. This is a consequence of the fact that the heterogeneity of skills in each track also remains at a similar level in both cases. As argued in Section 2.3, the conditional variances of skills in each track are necessarily smaller in the late tracking case when skill shocks are present and the track decision is made optimally. However, in our counterfactual experiment, parents make the late track decision, subject to the same asymmetric preference shocks as before. Quantitatively, this results in slightly larger deviations of the track choices from the recommended tracks compared to the baseline early tracking case (+2%).

[54]This also explains why, in the partial equilibrium case, the share of college workers increases in the first place. When learning becomes less efficient, more parents send their child to the academic track. This in turn, makes college education more likely, even when skills are lower, which raises the college share.

[55]As we have shown in Proposition 2 in Section 2.3, the effect of tracking on the overall variance of skills depends crucially on the presence of these direct peer effects.

Figure 2.3: Evolution of Average Child Skills in Counterfactual Experiments



Notes: This figure shows the evolution of average child skills from age 10 to 18. The baseline model simulated data is a red-solid line; the late tracking economy in partial equilibrium in a green-dotted-dashed line; and the late tracking economy in general equilibrium in a blue-dashed line.

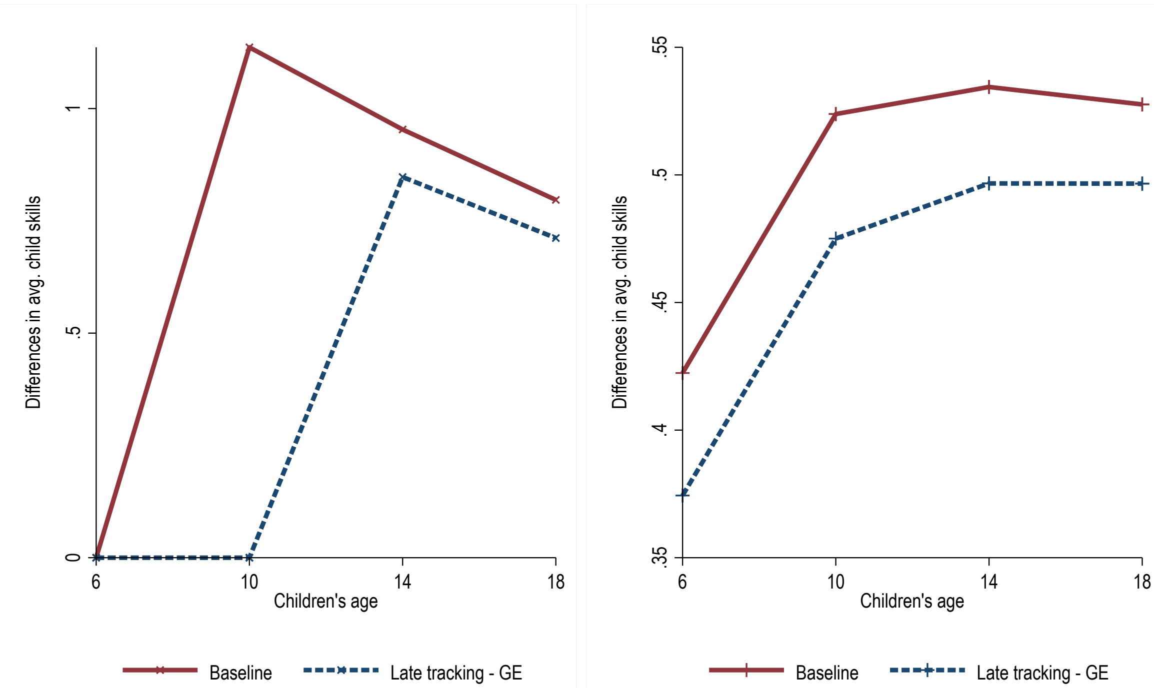
skills between parental backgrounds decrease, and relatively more children from a non-college parental background go to an academic track school once they are tracked in the late tracking case (+2.4%) relative to children from college parents (+1.5%). This can explain the increase in mobility as measured by the dependence of academic track graduation on the parental background.

The lower inequality in skills after one more period of comprehensive school translates into smaller differences in average skills between children in the academic and vocational track, once they are tracked. This is reinforced by the reduced differences between parental backgrounds in the track choice, and the fact that the track decision itself becomes less dependent on skills. The overall effect of postponing tracking on the child skill differences between tracks is plotted in the left panel of Figure 2.4, comparing the early tracking baseline economy (red) and the late tracking GE economy (blue). Smaller differences in skills across school tracks then entail smaller differences in adult human capital across college and non-college workers, which is again aided by the fact that relatively more children from the vocational school track go to college after the policy change. Given that college education and end-of-school skills and, thereby, human capital are the main determinants of income, overall earnings inequality therefore declines.

A consequence of these effects is that not only the school track but also the end-of-school

Figure 2.4: Differences in Average Child Skills in Early and Late Tracking Economy

(a) by school track (b) by parental education



Notes: These figures show the standardized differences in average skills between school tracks (left panel) and parental education (right panel) from age 6 to 18. The baseline model simulated data is a red-solid line, and the late tracking economy in general equilibrium is in a blue-dashed line.

skills, and the probability of going to college becomes less dependent on the initial skills, which are transferred from parents to children (see bottom rows of Panel B in Table 2.9). For that reason, intergenerational mobility, if defined as the dependence of economic outcomes of the child on parental economic outcomes, decreases. Interestingly, our model predicts that, in steady state, college attainment of children from college parents is still as likely or even slightly more likely than in the baseline, early tracking case. However, since college- and non-college parents become more similar in terms of their human capital, mobility in terms of income still improves.

Finally and importantly, the effects on inequality and mobility (as on efficiency) are reinforced through intergenerational linkages. For example, as college and non-college-educated parents become more similar in terms of their skills, and so do their children who inherit these skills. The differences in skills in terms of standard deviations between parental backgrounds are shown in the right panel of Figure 2.4. Notably, the relative differences between children of different parental backgrounds in the late-tracking GE case are reduced already at age 6.

We provide a comparison of our model's predictions regarding the effects of later tracking on learning outcomes to related findings from the empirical literature in Appendix 2.C. Overall, we maintain that our estimated effects are not at odds with existing empirical evidence. In particular, our model predicts learning gains for children from lower socioeconomic backgrounds and a decreased dependence on educational outcomes in secondary school on family background, which are among the most robust empirical findings in the literature. The fact that we estimate average learning losses from such a pervasive school tracking age reform can, in our eyes, not be refuted by existing evidence (nor can it be corroborated). It ultimately rests on the assumption of complementarity between child skills and the teaching practices in school, as highlighted in Section 2.3, which is itself based on empirical evidence (e.g. Duflo et al., 2011; Aucejo et al., 2022). The strength of our model-based approach is that it informs not just the short-term effect of school tracking on learning and educational outcomes of school children, but how these translate into higher education and labor market outcomes over multiple generations.

Abolishing School Tracking

Column (3) of Table 2.9 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, child skills become significantly more equal (i.e. $Var(\theta_5)$ drops by 2.1%). Moreover, as parents can no longer influence their children's skill evolution by choosing a specific school track, 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. In particular, children from non-college parents are 16.7% more likely to graduate from college than in the baseline economy, and children from college parents are 18.1% less likely to do so. Overall, mobility as measured by the (negative of the) intergenerational income elasticity improves significantly (+23.9%).

Similarly, a completely comprehensive school system reduces cross-sectional inequality markedly. For example, the Gini coefficient of earnings drops by 0.8%, as does the ratio of the 90th to 10th percentile of income. In addition, the differences in human capital across college and non-college workers become smaller, which decreases the college wage premium by 2.8%. On the other hand, abolishing tracking altogether makes learning even less efficient relative to the late tracking economy. On average, end-of-school skills are around -2.3% smaller in this economy. This leads to losses in aggregate human capital (-0.4%) and output

(-0.2%). Similarly and despite considerable equality gains, a completely comprehensive school system worsens average welfare in consumption equivalent units by 0.08%.

2.5.3 Limiting Parental Influence in the School Track Choice

In this section, we evaluate the effects of reducing parental influence on the school track choice without modifying the timing of school tracking. As discussed before, any force that impacts the school track allocation net of child skills is, in theory, detrimental to the efficiency of teaching and thus skill development in secondary school, if it dilutes the homogeneity of peer groups in each track. An interesting 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 of the economy. Our model provides a suitable environment to investigate such effects.

We evaluate two counterfactual scenarios: first, we shut down the asymmetry in academic track utility costs faced by parents of different education levels ($\chi_0 = \chi_1 = 0$). As we argued before, this asymmetry is a parsimonious way of capturing multiple reasons why parents systematically bias the school track choice toward their own educational path.[56] Second, 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.[57]

Table 2.10 shows that shutting down the asymmetry in academic track utility costs or enforcing a tracking threshold improves aggregate output. As before, this is mirrored by an

[56] We focus on this experiment as we view this as being the easiest to address by policies. For example, if the asymmetric school track costs are coming from information frictions, mentoring programs have proven very effective and almost cost-free in alleviating some of these frictions, as argued by Resnjanskij et al. (2024). While in the counterfactual scenario, we diminish parental influence from all socioeconomic groups, mentoring programs mostly target at-risk youths. Interventions that target all socioeconomic groups are rarer. An exception is Hakimov et al. (2022) who provide information about their chances of success at graduating college to all children, independently of their socioeconomic backgrounds. They find a reduction in the social elite college admission gap mostly driven by an increase in the admission of high-achieving low SES students to elite colleges.

[57] As derived in Section 2.3, 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 44% of children in terms of their skills being allocated to the academic track and the rest to the vocational track.

Table 2.10: Effects of School Track Choice Counterfactuals

	Changes in % (1)	Changes in % (2)
	$\chi_0 = 0$ $\chi_1 = 0$	Skill Threshold
Panel A - Aggregate, Distributional and Intergenerational Outcomes		
Output (Y)	+0.04	+0.12
Human Capital (H)	+0.05	+0.15
Gini of earnings	0.0	+0.8
Intergenerational income mobility ($-$ Income rank-rank coef.)	+0.9	-6.5
Parental Income on Academic Track (Dodin et al., 2024)	-25	+34
Welfare (CEV)	+0.04	-0.01
Panel B - Educational Outcomes		
% Acadmic track	-0.7	0.0
...if college parents	-8.6	-12
...if non-college parents	+9.6	+15.6
% College	0.3	0.0
...if college parents	-3.5	-8.8
...if non-college parents	+3.9	+7.2
...if academic track	+0.5	-3.5
...if vocational track	0.0	+14.4
Average end-of-school skills ($\bar{\theta}_5$)	+0.8	+3.0
Average middle-of-school skills ($\bar{\theta}_4$)	+0.2	+4.9
Average skills in V-Track upon tracking ($\bar{\theta}_3 \| S = V$)	-0.4	-50.0
Average skills in A-Track upon tracking ($\bar{\theta}_3 \| S = A$)	+1.2	+38.5
Variance of end-of-school skills ($Var(\theta_5)$)	+0.2	+1.9
Variance of middle-of-school skills ($Var(\theta_4)$)	-0.2	+0.9
Variance in V-Track upon tracking ($Var(\theta_3 \| S = V)$)	-0.4	-39.1
Variance in A-Track upon tracking ($Var(\theta_3 \| S = A)$)	-0.4	-18.1
Correlation between A-Track and Skills in period 3	+5.7	+59.4
Correlation between academic track and initial skills	+5.4	+79.5
Correlation between end-of-school skills and initial skills	-0.6	+0.9
Correlation between college graduation and initial skills	+0.9	+54.1

Notes: Column (1) displays percentage changes relative to the baseline economy entailed by the absence of parental preference for education, and Column (2) displays percentage changes entailed by skill treshold-rule for school tracking. All results are coming from the new general equilibrium distribution.
Intergenerational mobility is measured as the negative of the income rank-rank coefficient.

increase in aggregate human capital in the economy, in both cases. Note that both the share of college-educated agents and the share of children in academic track schools remain constant relative to the baseline case (see Panel B of Table 2.10). Both counterfactual scenarios lead to an increase in average skills at the end of secondary school. This increase arises from the fact that the variation in child skills within the school tracks becomes smaller, both when eliminating parental track preferences and especially when enforcing a tracking threshold. Since lower heterogeneity in skills within a school track improves learning efficiency, as derived in Section 2.3, this leads to higher end-of-school skills and thus higher adult human capital. This is consistent with the explanation of the learning efficiency-reducing misallocation effects that arise when parental background or any other factors drive the school track choice independently from skills. Unsurprisingly, without asymmetry in parental academic track costs and even more so with a sharp, purely skill-based allocation rule, the correlation of school track with parental education decreases, and skills themselves become more important in explaining the track choice.

However, while mobility, as measured by the negative of the intergenerational income elasticity, increases in the first counterfactual experiment (Column (1)), it decreases substantially when introducing a strict skill threshold (Column (2)). The reason for this is that purely skill-based tracking also increases the overall heterogeneity in skills markedly (i.e. $Var(\theta_5)$ increases by around 2%). In particular, while it increases learning on average, a cut-off-based school track allocation predominantly benefits the children in academic track schools. The argument is similar to Proposition 2 in Section 2.3: When factors other than skills determine the track choice, child skills in each track become more heterogeneous. In some sense, each track is thus more like a comprehensive school. As argued in Proposition 2, the learning losses from moving towards a stricter tracking system relative to a more comprehensive system are asymmetric and concentrated in the lower track, whenever the direct peer effects are positive, which is the case. Quantitatively, this effect can be seen in Panel B. in Table 2.10, where average skills in the vocational track at the point of the track decision decrease while they increase in the academic track.

As a result, cross-sectional inequality as measured by the earnings Gini coefficient rises by 0.8% in the case of a skill threshold compared to the baseline economy. Furthermore, larger inequality among college and non-college parents feeds into larger inequality in skills of their children. Consequently, the school track, end-of-school skills and college outcomes become significantly more dependent on the initial skill level in an economy with a strict skill-based separation, explaining lower mobility even though both education choices are less dependent on parental education (see Panel B of Table 2.10).

Perhaps surprisingly, overall welfare in terms of consumption equivalence variation in this counterfactual economy is slightly lower than in the baseline, despite increased output. On the one hand, this is due to the fact that cross-sectional inequality increases, which lowers welfare. On the other hand, when tracking is based on a skill threshold some parents no longer receive utility from being able to send their child to their preferred track. In contrast, when shutting down these parental track preferences directly, income inequality does not increase, as end-of-school skills are only slightly more dispersed. Given that aggregate human capital and output are higher, welfare is also increased by around 0.04% in this scenario.

In sum, these results point to an important role that measures such as mentoring programs, which have been shown to alleviate the influence of family background on school track decisions that is not justified by skill selection, can play in improving both aggregate efficiency and mobility at the same time. In contrast, reverting to a purely merit-based school track selection is, according to the predictions of our model, not welfare-enhancing and dampens equality and social mobility.

2.6 Conclusion

What is the role of education policies for aggregate productivity, inter-generational mobility, and inequality? We focus on the role of school tracking, a common—and controversial—education policy that has not been studied so far in the macroeconomic literature. As the long-run macroeconomic effects of school tracking involve the interaction of different markets and play out across generations, our analysis relies on a rich dynamic GE model with overlapping generations.

The key ingredient in our model is a parsimonious theory of skill formation in school. Skills are accumulated at a speed that depends on parental background, the pace of instruction in school, and the skills of classroom peers. The pace of instruction and the skills of classroom peers are, in turn, shaped by whether and when there is school tracking. We find that the theoretical implications of the model align with the empirical findings of the effect of tracking on educational achievements, as well as arguments in the public debate about tracking.

We tailor the model to fit the German Education System, where the track decision occurs when children are ten years old, and calibrate it using a variety of micro and macro data on child achievements, schools, and labor market outcomes. Our calibrated model predicts that the timing of school tracking involves a macroeconomic trade-off between efficiency and social mobility. Concretely, a policy reform that postpones school tracking by four years, which implies that children are in comprehensive school until age fourteen, decreases long-run GDP by 0.1% and lowers the inter-generational income elasticity by around 2.2%. 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 GDP loss mostly stems from lower learning efficiency due to more heterogeneous classrooms during the (prolonged) time in comprehensive school. The gain in social mobility is the result of comprehensive school reducing heterogeneity in skills, which implies that the school track depends less on parental background, and skill differences across tracks become smaller once the track decision is made.

Consistent with previous findings in the literature, our calibrated model also yields that parental background matters for the school track decision even when child skills are accounted for. We find that reducing this direct influence of parental background on the school track leads to improvements in both social mobility and economic output. Mentoring programs reducing the direct role of parental background (e.g. Raposa et al., 2019; Resnjanskij et al., 2024), can therefore simultaneously improve macroeconomic efficiency and social mobility.

Bibliography

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

Abbott, B., G. Gallipoli, C. Meghir, and G. L. Violante (2019): "Education policy and intergenerational transfers in equilibrium," *Journal of Political Economy*, 127, 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., M. Doepke, G. Sorrenti, and F. Zilibotti (2023): "It takes a village: the economics of parenting with neighborhood and peer effects," Working Paper w27050, National Bureau of Economic Research.

Agostinelli, F., M. Saharkhiz, and M. Wiswall (2019): "Home and School in the Development of Children," Working Paper w26037, National Bureau of Economic Research.

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

Arenas, A. and J. Hindriks (2021): "Intergenerational mobility and unequal school opportunity," *The Economic Journal*, 131, 1027–1050.

Aucejo, E., P. Coate, J. C. Fruehwirth, S. Kelly, and Z. Mozenter (2022): "Teacher effectiveness and classroom composition: Understanding match effects in the classroom," *The Economic Journal*, 132, 3047–3064.

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

Bauer, P. and R. T. Riphahn (2006): "Timing of school tracking as a determinant of intergenerational transmission of education," *Economics Letters*, 91, 90–97.

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

——— (1986): "Human capital and the rise and fall of families," *Journal of labor economics*, 4, S1–S39.

Bellenberg, G. and M. Forell (2012): "Schulformwechsel in Deutschland," *Durchlässigkeit und Selektion in den 16 Schulsystemen der Bundesländer innerhalb der Sekundarstufe I.*

Betts, J. R. (2011): "The economics of tracking in education," in *Handbook of the Economics of Education*, Elsevier, vol. 3, 341–381.

Bildungsberichterstattung, A. (2018): "Bildung in Deutschland 2018: ein indikatorengestützter Bericht mit einer Analyse zu Wirkungen und Erträgen von Bildung," .

Blossfeld, H., H. Rossbach, and J. von Maurice (2019): "Education as a lifelong process: The German National Educational Panel Study (NEPS)," *Edition ZfE.*

Bonesrønning, H., H. Finseraas, I. Hardoy, J. M. V. Iversen, O. H. Nyhus, V. Opheim, K. V. Salvanes, A. M. J. Sandsør, and P. Schøne (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.

Brunello, G. and D. Checchi (2007): "Does school tracking affect equality of opportunity? New international evidence," *Economic policy*, 22, 782–861.

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

Brunello, G., L. Rocco, K. Ariga, and R. Iwahashi (2012): "On the efficiency costs of de-tracking secondary schools in Europe," *Education Economics*, 20, 117–138.

Bénabou, R. (1996): "Equity and efficiency in human capital investment: the local connection," *The Review of Economic Studies*, 63, 237–264.

Capelle, D. (2022): "The Great Gatsby goes to College: Tuition, Inequality and Intergenerational Mobility in the U.S." Working paper.

Carlana, M., E. La Ferrara, and P. Pinotti (2022): "Goals and gaps: Educational careers of immigrant children," *Econometrica*, 90, 1–29.

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

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

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

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

Daruich, D. (2022): “The Macroeconomic Consequences of Early Childhood Development Policies,” Working Paper 2018-29, FRB St. Louis.

Dodin, M., S. Findeisen, L. Henkel, D. Sachs, and P. Schüle (2024): “Social mobility in Germany,” *Journal of Public Economics*, 232, 105074.

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

Dohmen, D., M. Thomsen, G. Yelubayeva, and R. Ramirez (2019): “Ermittlung der Lebenshaltungskosten von Studierenden: Aktualisierte Berechnung anhand der 21. Sozialerhebung des Deutschen Studentenwerks,” *FiBS - Forschungsinstitut für Bildungs- und Sozialökonomie*.

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

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

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

Epple, D., E. Newlon, and R. Romano (2002): “Ability tracking, school competition, and the distribution of educational benefits,” *Journal of Public Economics*, 83, 1–48.

Epple, D. and R. Romano (2011): “Peer effects in education: A survey of the theory and evidence,” in *Handbook of Social Economics*, Elsevier, vol. 1, 1053–1163.

Esser, H. and J. Seuring (2020): “Kognitive Homogenisierung, schulische Leistungen und soziale Bildungsungleichheit,” *Zeitschrift für Soziologie*, 49, 277–301.

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

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

Fujimoto, J., D. Lagakos, and M. Vanvuren (2023): “Aggregate and Distributional Effects of ‘Free’ Secondary Schooling in the Developing World,” Working Paper w31029, National Bureau of Economic Research.

Goebel, J., M. M. Grabka, S. Liebig, M. Kroh, D. Richter, C. Schröder, and J. Schupp (2019): "The German socio-economic panel (SOEP)," *Jahrbücher für Nationalökonomie und Statistik*, 239, 345–360.

Hakimov, R., R. Schmacker, and C. Terrier (2022): "Confidence and college applications: Evidence from a randomized intervention," Working paper, WZB Discussion Paper.

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

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

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

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

Holmlund, H. (2008): "Intergenerational Mobility and Assortative Mating: Effects of an Educational Reform. CEE DP 91." *Centre for the Economics of Education (NJ1).*

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

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

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

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

Kotera, T. and A. Seshadri (2017): "Educational policy and intergenerational mobility," *Review of economic dynamics*, 25, 187–207.

Krueger, D. and A. Ludwig (2016): "On the optimal provision of social insurance: Progressive taxation versus education subsidies in general equilibrium," *Journal of Monetary Economics*, 77, 72–98.

Kyzyma, I. and O. Groh-Samberg (2018): "Intergenerational Economic Mobility in Germany: Levels und Trends," Working paper, DIW.

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

Lavy, V., M. D. Paserman, and A. Schlosser (2012): "Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom," *The Economic Journal*, 122, 208–237.

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

Mahler, L. and M. Yum (2023): "Lifestyle Behaviors and Wealth-Health Gaps in Germany," Tech. rep., Available at SSRN 4034661.

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

Masters, G. N. (1982): "A Rasch model for partial credit scoring," *Psychometrika*, 47, 149–174.

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

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

Meyer-Guckel, V., J. Klier, J. Kirchherr, and F. Suessenbach (2021): "» Vom Arbeiterkind zum Doktor: Der Hürdenlauf auf dem Bildungsweg der Erststudierenden «," *Stifterverband/McKinsey Diskussionpapier*, 2.

Nennstiel, R. (2022): "No Matthew effects and stable SES gaps in math and language achievement growth throughout schooling: Evidence from Germany," *European sociological review*.

NEPS Network (2022): "National Educational Panel Study, Scientific Use File of Starting Cohort Kindergarten," .

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

OECD (2020a): "Education Policy Outlook in Germany," *OECD Education Policy Perspectives*.

——— (2020b): "PISA 2018 Results (Volume V): Effective Policies, Successful Schools," *PISA*.

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

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

Pekkarinen, T., R. Uusitalo, and S. Kerr (2009): "School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform," *Journal of Public Economics*, 93, 965–973.

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.

Pischke, J.-S. and A. Manning (2006): "Comprehensive versus selective schooling in England in Wales: What do we know?" Working Paper w12176, National Bureau of Economic Research.

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

——— (2013): "Scaling of competence tests in the National Educational Panel Study-Many questions, some answers, and further challenges," *Journal for Educational Research Online*, 5, 189–216.

Raposa, E. B., J. Rhodes, G. J. J. Stams, N. Card, S. Burton, S. Schwartz, L. A. Y. Sykes, S. Kanchewa, J. Kupersmidt, and S. Hussain (2019): "The effects of youth mentoring programs: A meta-analysis of outcome studies," *Journal of youth and adolescence*, 48, 423–443.

Rasch, G. (1960): *Probabilistic models for some intelligence and attainment tests.*, ERIC.

Resnjanskij, S., J. Ruhose, S. Wiederhold, L. Woessmann, and K. Wedel (2024): "Can Mentoring Alleviate Family Disadvantage in Adolescence? A Field Experiment to Improve Labor Market Prospects," *Journal of Political Economy*, 132, 000–000.

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

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

Sacerdote, B. (2011): "Peer effects in education: How might they work, how big are they and how much do we know thus far?" in *Handbook of the Economics of Education*, Elsevier, vol. 3, 249–277.

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

Waldinger, F. (2006): "Does tracking affect the importance of family background on students' test scores?" *Unpublished manuscript, LSE.*

Wössmann, L. (2020): "Gleiche Chancen? Je früher, desto besser! Bildungsgerechtigkeit im deutschen Schulsystem," *lautstark*, 07, 20–22.

Yum, M. (2023): "Parental time investment and intergenerational mobility," *International Economic Review*, 64, 187–223.

Appendices to Chapter 2

2.A Model Appendix

2.A.1 Proof of Propositions

Proposition 1

For the proof of this proposition, we denote by θ_3 the child skills at the beginning of secondary school and by θ_4 the skills at the end of secondary school. Moreover, we have assumed $\kappa = 1$, $\zeta = 0$, and $\chi = 0$ and that skills at the beginning of secondary school are normally distributed with mean zero and variance $\sigma^2_{\theta_3}$. First, we show that maximizing the aggregate end-of-school skills in a tracking system implies a threshold skill level $\tilde{\theta}_3$, such that all $\theta_3 < \tilde{\theta}_3$ go to one track, call it $S = V$ and all $\theta_1 > \tilde{\theta}_3$ go to the other track, $S = A$ (and those with $\theta_3 = \tilde{\theta}_3$ 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_4 in (2.2) of a child in a given school track S with instruction pace P^S using Lemma 1 as:

$$\theta_4 = \theta_3 + \alpha\bar{\theta}_3^S + \frac{\beta^2}{2\delta} + \frac{\beta\gamma\theta_3}{\delta} + \frac{\gamma^2\theta_3\bar{\theta}_3^S}{\delta} - \frac{\gamma^2(\bar{\theta}_3^S)^2}{2\delta} + \eta_4. \tag{2.30}$$

After adding and subtracting $\frac{\gamma^2}{2\delta}\theta_3^2$, this can be expressed as

$$\begin{aligned}\theta_4 &= \theta_3 + \alpha\bar{\theta}_3^S + \frac{\beta^2}{2\delta} + \frac{\beta\gamma\theta_3}{\delta} + \frac{\gamma^2\theta_3^2}{2\delta} + \eta_4 - \frac{\gamma^2}{2\delta}\left(\theta_3^2 - 2\theta_3\bar{\theta}_3^S + (\bar{\theta}_3^S)^2\right) \\ &= \theta_4(P^*_{\theta_3}) - \frac{\gamma^2}{2\delta}(\theta_3 - \bar{\theta}_3^S)^2,\end{aligned} \tag{2.31}$$

where $\theta_4(P^*_{\theta_3})$ denotes end-of-school skills if the child with skills θ_3 is taught at her individually optimal teaching pace $P^*_{\theta_3}$ (we suppress the j-index of P as we consider only one period in this case). Thus, in a given track, end-of-school skills are a strictly decreasing function of the *distance* to the average skill level $\bar{\theta}_3^S$ 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 $\mathbb{E}[\theta_4]$ 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, $\mathbb{E}(\theta_3|S = V) < \mathbb{E}(\theta_3|S = A)$. Then, because there is no strict threshold, this means that for any initial skill level θ_3, there must be at least two children with initial skill levels smaller or equal to θ_3 that go to different tracks or at least two children with initial skill levels larger or equal than θ_3 that go to different tracks. This implies that there exists a child with $\theta_3' \leq \mathbb{E}(\theta_3|S = V)$ that goes to track $S = A$, and/or a child with $\theta_3' \geq \mathbb{E}(\theta_3|S = A)$ that goes to track $S = V$, and/or two children with skills $\theta_3' < \theta_3''$, with $\theta_3', \theta_3'' \in [\mathbb{E}(\theta_3|S = V), \mathbb{E}(\theta_3|S = A)]$, 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 (2.31), this child with θ_3' 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 $\tilde{\theta}_3$ be the skill threshold and let S again indicate to which track a child is allocated, now with $S = V$ for all $\theta_3 \leq \tilde{\theta}_3$ and $S = A$ for all $\theta_3 > \tilde{\theta}_3$.

A policymaker solves

$$\begin{aligned} &\max_{\tilde{\theta}_3} \quad \mathbb{E}(\theta_4) \\ \iff &\max_{\tilde{\theta}_3} \quad \mathbb{E}(\mathbb{E}(\theta_4|S)) \\ &\text{subject to} \\ &\qquad P^S \text{ chosen optimally given Lemma 1.} \end{aligned} \tag{2.32}$$

Using (2.30) and the law of iterated expectations, this maximization problem boils down to

$$\begin{aligned} &\max_{\tilde{\theta}_3} \quad \frac{\beta^2}{2\delta} + \frac{\gamma^2}{2\delta}\,\mathbb{E}\left(\mathbb{E}(\theta_3|S)^2\right) \\ \iff &\max_{\tilde{\theta}_3} \quad \frac{\beta^2}{2\delta} + \frac{\gamma^2}{2\delta}\left(F(\tilde{\theta}_3)\,\mathbb{E}(\theta_3|\theta_3 \leq \tilde{\theta}_3)^2 + (1-F(\tilde{\theta}_3))\,\mathbb{E}(\theta_3|\theta_3 > \tilde{\theta}_3)^2\right), \end{aligned} \tag{2.33}$$

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 $\tilde{\theta}_3$. Using the law of total variance, the maximization problem can thus be rewritten as (dropping the constant term)

$$\begin{aligned} &\max_{\tilde{\theta}_3} \quad \mathbb{E}(\theta_4) \\ \iff &\max_{\tilde{\theta}_3} \quad \frac{\gamma^2}{2\delta}\left(\sigma^2_{\theta_3} - \mathbb{E}(Var[\theta_3|S])\right). \end{aligned} \tag{2.34}$$

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 $\tilde{\theta}_3^* = \mathbb{E}\,\theta_3 = 0$, 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 $\hat{\theta}_3$ is indifferent between tracks V and A iff

$$\begin{aligned} &\left(\alpha + \hat{\theta}_3\frac{\gamma^2}{\delta}\right)\mathbb{E}(\theta_3|\theta_3 \leq \hat{\theta}_3) - \frac{\gamma^2}{2\delta}\,\mathbb{E}(\theta_3|\theta_3 \leq \hat{\theta}_3)^2 \\ = &\left(\alpha + \hat{\theta}_3\frac{\gamma^2}{\delta}\right)\mathbb{E}(\theta_3|\theta_3 > \hat{\theta}_3) - \frac{\gamma^2}{2\delta}\,\mathbb{E}(\theta_3|\theta_3 > \hat{\theta}_3)^2 \\ \iff &\left(-\alpha - \hat{\theta}_3\frac{\gamma^2}{\delta}\right)\sigma_{\theta_3}\frac{f(\hat{\theta}_3/\sigma)}{F(\hat{\theta}_3/\sigma)} - \frac{\gamma^2}{2\delta}\sigma^2_{\theta_3}\frac{f(\hat{\theta}_3/\sigma)^2}{F(\hat{\theta}_3/\sigma)^2} \\ = &\left(\alpha + \hat{\theta}_3\frac{\gamma^2}{\delta}\right)\sigma_{\theta_3}\frac{f(\hat{\theta}_3/\sigma)}{1-F(\hat{\theta}_3/\sigma)} - \frac{\gamma^2}{2\delta}\sigma^2_{\theta_3}\frac{f(\hat{\theta}_3/\sigma)^2}{(1-F(\hat{\theta}_3/\sigma))^2} \end{aligned} \tag{2.35}$$

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 $\hat{\theta}_3$ that solves (2.35) numerically. In all cases with reasonable parameter values, (2.35) 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 $\hat{\theta}_3 = 0$, as can be directly seen from (2.35). When $\alpha > 0$, the root is smaller than 0, i.e. $\hat{\theta}_3 < 0$.

Proposition 2

The proof of this Proposition follows directly from (2.34). 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_4 is monotonically increasing in θ_3 in a given track (see (2.30)), this is the child with the highest initial skill, say $\theta_{3,max}$. Moreover, from the properties of a truncated normal distribution, we know that, for any skill threshold $\tilde{\theta}_3$, average skills in the A track, $\bar{\theta}_{3,A}$ are larger than the unconditional average, $\bar{\theta}_{3,C} = 0$. Thus, the squared distance between $\theta_{3,max}$ and $\bar{\theta}_{3,A}$ in a tracking system is smaller. Taken together, (2.31) implies that the child with initial skill $\theta_{3,max}$ 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_4 are monotonically increasing in θ_3 in every track, the range is characterized by the intersection of the linear function $\theta_{4,C}(\theta_3, \bar{\theta}_{3,C})$ with $\theta_{4,V}(\theta_3, \bar{\theta}_{3,V})$ and $\theta_{4,A}(\theta_3, \bar{\theta}_{3,A})$, which are just (2.30) if everyone was taught at the comprehensive, academic, or vocational pace. For any skill threshold, the lower intersection $\theta_{3,L}$ hence solves

$$\begin{aligned}
& \theta_{3,L} + \alpha\bar{\theta}_{3,C} + \frac{\beta^2}{2\delta} + \frac{\beta\gamma}{\delta}\theta_{3,L} + \frac{\gamma^2}{\delta}\bar{\theta}_{3,C}\theta_{3,L} - \frac{\gamma^2}{2\delta}\bar{\theta}^2_{3,C} + \eta_4 \\
= \; & \theta_{3,L} + \alpha\bar{\theta}_{3,V} + \frac{\beta^2}{2\delta} + \frac{\beta\gamma}{\delta}\theta_{3,V} + \frac{\gamma^2}{\delta}\bar{\theta}_{3,V}\theta_{3,L} - \frac{\gamma^2}{2\delta}\bar{\theta}^2_{3,V} + \eta_4 \qquad (2.36) \\
\iff \quad \theta_{3,L} = \; & \frac{1}{2}\bar{\theta}_{3,V} - \frac{\alpha\delta}{\gamma^2}.
\end{aligned}$$

Similarly, the upper intersection is given at

$$\theta_{3,U} = \frac{1}{2}\bar{\theta}_{3,A} - \frac{\alpha\delta}{\gamma^2}. \qquad (2.37)$$

For any skill threshold $\tilde{\theta}_3$, the interval $[\theta_{3,L}, \bar{\theta}_{3,U}]$ 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 $\tilde{\theta}_3 = 0$ even if parents endogenously sort their children. Hence, children with initial skills inside a symmetric interval around 0, $[\frac{1}{2}\bar{\theta}_{3,V}, \frac{1}{2}\bar{\theta}_{3,A}]$, lose relative to a comprehensive track, since $\bar{\theta}_{3,V} = -\bar{\theta}_{3,A}$ if $\tilde{\theta}_1 = 0$. The average loss of a child in this interval is equal to $\frac{\gamma^2}{2\delta}\bar{\theta}^2_{3,V} = \frac{\gamma^2}{2\delta}\bar{\theta}^2_{3,A}$.

If $\alpha > 0$, and the policymaker enforces the tracking skill threshold $\tilde{\theta}_3 = 0$, 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 $[\theta_{3,L}, 0]$ is allocated into the V track but loses relative to a comprehensive system. Similarly, every child with an initial skill inside $[0, \theta_{3,U}]$ is allocated to track A but loses relative to a comprehensive system. With $\alpha > 0$, $|\theta_{3,U}| < |\theta_{3,L}|$ and hence, the range of children in the A track that lose is smaller. The interval $[0, \theta_{3,U}]$ may even be empty in which case only children in the V track lose from tracking.

Proposition 3

For the proof of this proposition, we denote by θ_3 the child skills at the beginning of secondary school, by θ_4 the skills at the intermediary stage of secondary school and by θ_5 the skills at the end of secondary school. All other assumptions are maintained. First, we characterize the variance of θ_4. We start by collecting expressions for conditional and unconditional first and second moments.

The unconditional expected value of θ_4 in track V, if everyone went to V is

$$\begin{aligned}\mathbb{E}(\theta_{4,V}) &= \frac{\beta^2}{2\delta} + \alpha\bar{\theta}_{3,V} - \frac{\gamma^2}{2\delta}\bar{\theta}^2_{3,V} \\ &= \frac{\beta^2}{2\delta} - \alpha\sigma_{\theta_1}\frac{f(\tilde{\theta}_3/\sigma_{\theta_3})}{F(\tilde{\theta}_3/\sigma_{\theta_3})} - \frac{\gamma^2}{2\delta}\sigma^2_{\theta_3}\frac{f(\tilde{\theta}_3/\sigma_{\theta_3})^2}{F(\tilde{\theta}_3/\sigma_{\theta_3})^2}.\end{aligned} \tag{2.38}$$

The unconditional expected value of θ_4 in track A, if everyone went to A is

$$\begin{aligned}\mathbb{E}(\theta_{4,A}) &= \frac{\beta^2}{2\delta} + \alpha\bar{\theta}_{3,A} - \frac{\gamma^2}{2\delta}\bar{\theta}^2_{3,A} \\ &= \frac{\beta^2}{2\delta} + \alpha\sigma_{\theta_3}\frac{f(\tilde{\theta}_3/\sigma_{\theta_3})}{1 - F(\tilde{\theta}_3/\sigma_{\theta_3})} - \frac{\gamma^2}{2\delta}\sigma^2_{\theta_3}\frac{f(\tilde{\theta}_3/\sigma_{\theta_3})^2}{(1 - F(\tilde{\theta}_3/\sigma_{\theta_3}))^2}.\end{aligned} \tag{2.39}$$

The variance of θ_4 in a comprehensive system is

$$\begin{aligned} Var(\theta_{4,C}) &= \mathbb{E}(\Big(\theta_4 - \mathbb{E}(\theta_4))^2\Big) \\ &= (1+\beta)^2\sigma^2_{\theta_3} + \sigma^2_{\eta_4} \\ &\sigma^2_{\theta_4,C} + \sigma^2_{\eta_4}, \end{aligned} \tag{2.40}$$

where we define $\sigma^2_{\theta_4,C}$ to be the variance of θ_4 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

$$\begin{aligned} \mathbb{E}(\theta_{5,LT}) &= \mathbb{E}(\mathbb{E}(\theta_{5,LT}|S_{LT})) \\ &= \mathbb{E}(\theta_{4,LT}) + \frac{\beta^2}{2\gamma} + (\alpha+\beta)\,\mathbb{E}(\mathbb{E}(\theta_{4,LT}|S_{LT})) + \frac{\gamma}{2}\,\mathbb{E}(\mathbb{E}(\theta_{4,LT}|S_{LT})^2) \\ &= (2+\alpha+\beta)\frac{\beta^2}{2\gamma} + \frac{\gamma}{2}[\sigma^{\theta_4,LT} - \mathbb{E}(Var(\theta_{4,LT}|S_{LT}))], \end{aligned} \tag{2.41}$$

where $\mathbb{E}(\theta_{4,LT})$ and $\sigma^2_{\theta_4,LT}$ are just equal to the mean and variance of the comprehensive system in the one-period model (see equation (2.40)). The variable S_{LT} indicates the track selection in period 2, which follows the cut-off rule $S_{LT} = V$ if $\theta_{4,LT} \leq \tilde{\theta}_{4,LT}$ and $S_{LT} = A$ otherwise. The cut-off that maximizes (2.41) is $\tilde{\theta}^*_{4,LT} = \mathbb{E}(\theta_{4,LT}) = \frac{\beta^2}{2\gamma}$. This follows as (2.41) mirrors that of average end-of-school skills in the full tracking system of the one-period model in that average and aggregate $\theta_{5,LT}$ 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

$$\begin{aligned} \mathbb{E}(\theta_{5,ET}) &= \mathbb{E}(\mathbb{E}(\theta_{5,ET}|S_{ET})) \\ &= \frac{\beta^2}{2\gamma} + (1+\alpha+\beta)\,\mathbb{E}(\mathbb{E}(\theta_{4,ET}|S_{ET})) + \beta\frac{\gamma}{2}\,\mathbb{E}(\mathbb{E}(\theta_{4,ET}|S_{ET})^2) \\ &= \frac{\beta^2}{2\gamma} + (1+\alpha+\beta)\left(\frac{\beta^2}{2\gamma} + \beta\frac{\gamma}{2}[\sigma^2_{\theta_3} - \mathbb{E}(Var(\theta_{3,ET}|S_{ET}))]\right) + \beta\frac{\gamma}{2}\,\mathbb{E}(\mathbb{E}(\theta_{4,ET}|S_{ET})^2 \\ &= \frac{\beta^2}{2\gamma} + (1+\alpha+\beta)\left(\frac{\beta^2}{2\gamma} + \beta\frac{\gamma}{2}[\sigma^2_{\theta_3} - \mathbb{E}(Var(\theta_{3,ET}|S_{ET}))]\right) \\ &+ \beta\frac{\gamma}{2}[\sigma^2_{\theta_4,ET} - \mathbb{E}(Var(\theta_{4,ET}|S_{ET}))]. \end{aligned} \tag{2.42}$$

Comparing (2.41) and (2.42), the condition that governs if average end-of-school skills in a late tracking system are larger than in an early tracking system reads

$$\begin{aligned}
&\mathbb{E}(\theta_{5,LT}) - \mathbb{E}(\theta_{5,ET}) \\
&= \beta\frac{\gamma}{2}\left(\mathbb{E}(\mathbb{E}(\theta_{4,LT}|S_{LT})^2) - \mathbb{E}(\mathbb{E}(\theta_{4,ET}|S_{ET})^2)\right) \\
&- (1+\alpha+\beta)\beta\frac{\gamma}{2}\,\mathbb{E}(\mathbb{E}(\theta_3|S_{ET})^2) > 0.
\end{aligned} \tag{2.43}$$

The last term of (2.43) 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

$$\mathbb{E}(\theta_{4,LT}|S_{LT} = V) = \frac{\beta^2}{2\gamma} - \sigma_{\theta_4,LT}\frac{f(\tilde{\theta}_{4,LT}/\sigma_{\theta_4,LT})}{F(\tilde{\theta}_{2,LT}/\sigma_{\theta_4,LT})} \tag{2.44}$$

and

$$\mathbb{E}(\theta_{4,LT}|S_{LT} = A) = \frac{\beta^2}{2\gamma} + \sigma_{\theta_4,LT}\frac{f(\tilde{\theta}_{4,LT}/\sigma_{\theta_4,LT})}{1 - F(\tilde{\theta}_{4,LT}/\sigma_{\theta_4,LT})}, \tag{2.45}$$

where the unconditional variance of θ_4 in a late tracking system is given by $\sigma^2_{\theta_4,LT} = \sigma^2_{\theta_4,C} + \sigma^2_{\eta_4}$, i.e. by the one computed in equation (2.40). Since late tracking occurs *after* the realization of skill shocks in period 4, this variance additively *includes* the variance of these shocks.

Condition (2.43) 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 $\tilde{\theta}_3 = \mathbb{E}(\theta_3) = 0$ and $\tilde{\theta}_4 = \mathbb{E}(\theta_{4,LT}) = \frac{\beta^2}{2\gamma}$. In that case, we can write the expressions for the various expected square conditional expected values as follows:

$$\begin{aligned}
\mathbb{E}(\mathbb{E}(\theta_3|S_{ET})^2) &= 2\chi\sigma^2_{\theta_3} \\
\mathbb{E}(\mathbb{E}(\theta_{4,LT}|S_{LT})^2) &= \frac{\beta^4}{4\gamma^2} + 2\chi(\sigma^2_{\theta_{4,LT}} + \sigma^2_{\eta_4}) \\
\mathbb{E}(\mathbb{E}(\theta_{4,ET}|S_{ET})^2) &= \frac{\beta^4}{4\gamma^2} + 2\chi\sigma^2_{\theta_3}\left(\alpha^2 + \gamma^2 f(0)^2\sigma^2_{\theta_3} - \frac{\beta^2}{2}\right) \\
+2f(0)\sigma^2_{\theta_3}\left(\beta^2 + 2\alpha(1+\beta) - (2\gamma f(0)\sigma_{\theta_3})^2\right) &+ 2\chi(\sigma^2_{\theta_{4,LT}} + 2\chi\gamma^2\sigma^2_{\theta_3}).
\end{aligned}$$

Condition (2.43) then becomes

$$
\begin{aligned}
&\mathbb{E}(\theta_{5,LT}) - \mathbb{E}(\theta_{5,ET}) \\
&= \beta\frac{\gamma}{2}\Bigg(2\chi\sigma^2_{\eta 4} - 2\chi\sigma^2_{\theta_3}\Bigg(\alpha^2 + \gamma^2 f(0)^2\sigma^2_{\theta_3} - \frac{\beta^2}{2} \\
&+ \beta^2 + 2\alpha(1+\beta) - 4\gamma^2 f(0)^2\sigma^2_{\theta_3} + 2\chi\gamma^2\sigma^2_{\theta_3} + 1 + \alpha + \beta\Bigg)\Bigg) \\
&= \frac{\gamma}{\pi}\left(\sigma^2_{\eta 4} - \sigma^2_{\theta_3}\left(1 + \alpha + \alpha^2 + \beta + \frac{\beta^2}{2} + 2\alpha(1+\beta) + \frac{\gamma^2}{2\pi}\sigma^2_{\theta_3}\right)\right) > 0.
\end{aligned}
\tag{2.46}
$$

From this, Proposition 3 follows.

2.A.2 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 3.2. 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 $\{\bar{\theta}_{j,S}\}$ for $j = 2, 3, 4$ 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:

$$
H_0 = \sum_{j=5}^{J_r}\int_{X_j} n_j(x_j)\; h_j\,(x_j)\, d\Theta(X \mid E=0) + \sum_{j=5}^{5}\int_{X_j} n_j(x_j)\; h_j\,(x_j)\, d\Theta(X \mid E=1)
$$

where the first summation is the supply of high-school graduates while the second is that labor supply of college students while studying in $j = 5$.

For college level:

$$H_1 = \sum_{j=6}^{J_r} \int_{X_j} n_j(x_j)\ h_j\,(x_j)\, d\Theta(X \mid E = 1).$$

5. Asset market clears

$$K = \sum_{j=J_e}^{J_d} \int_{X_j} a_j(x_j) d\Theta(X),$$

which implies that the goods market clears;

6. The distribution of X is stationary: $\Theta(X) = \int \Gamma(X) d\Theta(X)$.

2.A.3 Welfare Measure

Our analysis centers on evaluating aggregate welfare under different policy scenarios. 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 by $\mathcal{V}_5^C(\theta_5, a_5, \phi, S, E^p, \Delta)$ the welfare of agents in the initial state of the economy ($j = 5$) if their consumption (and that of their descendants) were multiplied by $(1 + \Delta)$:

$$\begin{aligned}\mathcal{V}^C(\theta_5, a_5, \phi, S, E^p, \Delta) = \mathbb{E}^C \sum_{j=5}^{j=20} \beta^{j-5} v_j \left(c_j^{*C}(1+\Delta), n_j^{*C}, E^{*C}, \theta_5, S, E^p\right) \\ + \beta^{13-5} \delta \mathcal{V}_{j^5}^C \left(\theta_5', a_5', \phi', S', E^{*C}, \Delta\right),\end{aligned}$$

where E^p is the education of the parent, and for $j = 6, ..., 10, 12, ..., 20$

$$v_j(c_j, n_j, E, \theta_5, S, E^p) = \frac{(c_j/q)^{1-\sigma}}{1-\sigma} - b\frac{n_j^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}, \tag{2.47}$$

for $j = 5$

$$v_j(c_j, n_j, E, \theta_5, S, E^p) = \frac{(c_j/q)^{1-\sigma}}{1-\sigma} - b\frac{n_j^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} - 1\{E=1\}\ \psi(S, \theta_5, \nu(E^p)), \tag{2.48}$$

and for $j = 11$

$$v_j(c_j, n_j, E, \theta_5, S, E^p) = \frac{(c_j/q)^{1-\sigma}}{1-\sigma} - b\frac{n_j^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} - 1\{S=A\}\ \chi(E). \tag{2.49}$$

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:

$$\bar{\mathcal{V}}^0(\Delta^C) = \bar{\mathcal{V}}^C(0).$$

2.B Empirical and Calibration Appendix

2.B.1 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) or OECD (2020a). Figure 2.B.1 illustrates a simplified structure of the system, starting in Grade 4 and ending with tertiary education.

Generally, schooling is compulsory in Germany for every child starting at age six and lasting until age 18. However, the obligation to go to school typically lasts until grade 9 or 10, after which it shifts to a vocational training obligation if no upper secondary school is attended. At age six, all children visit a comprehensive primary school that lasts the first four grades.[58] After that, children are allocated into traditionally three different secondary school tracks: A lower vocational track, a medium vocational track, and an academic track.

[58] In two federal states, Berlin and Brandenburg, comprehensive primary school lasts the first 6 grades.

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.B.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 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 into 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.

Figure 2.B.1: Simplified Structure of the German Education System

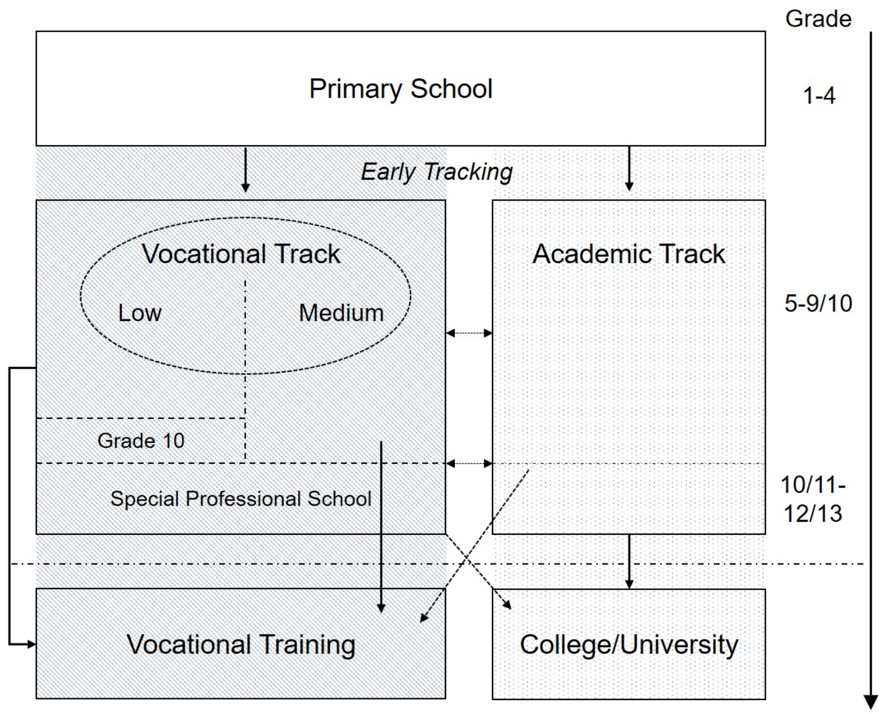


Table 2.B.1: Per-Student Public Expenditures across School Types and Years

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). All amounts in euros.

2.B.2 Empirical Evidence on School Track Selection

In this section, we present reduced-form evidence on the effect of parental background on the school track choice for their children.

Table 2.B.2 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. Consistently, we find that 54% of children from college-graduated parents receive a teacher recommendation for the academic track versus 39% of children from non-college-graduated parents.[59] In addition, Table 2.B.2 shows that while around 23% of parents who themselves have a college education overrule a vocational recommendation, only 4% 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. As argued before, one reason for these deviations may be that 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.B.2 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 6% of children of college-educated parents who deviated from the vocational track recommendation belong to the top quartile of skills in the academic track four years later in Grade 9. In contrast, the same number reaches 34% among those who received an academic track recommendation and followed that recommendation. This suggests that the support provided by college-educated parents who deviate from a vocational track recommendation and send their child to an academic track school 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 almost half of them belonging to the top quartile of skills in the vocational track. As a comparison, 22% 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

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

Table 2.B.2: School Track Choice

			% in the top 25% by track in G9	
Recommendation	Shares	% deviate	if followed	if deviated
College Parents				
Academic	56%	4%	34%	68%
Vocational	44%	23%	34%	6%
Non-college Parents				
Academic	38%	16%	21%	44%
Vocational	62%	12%	22%	14%

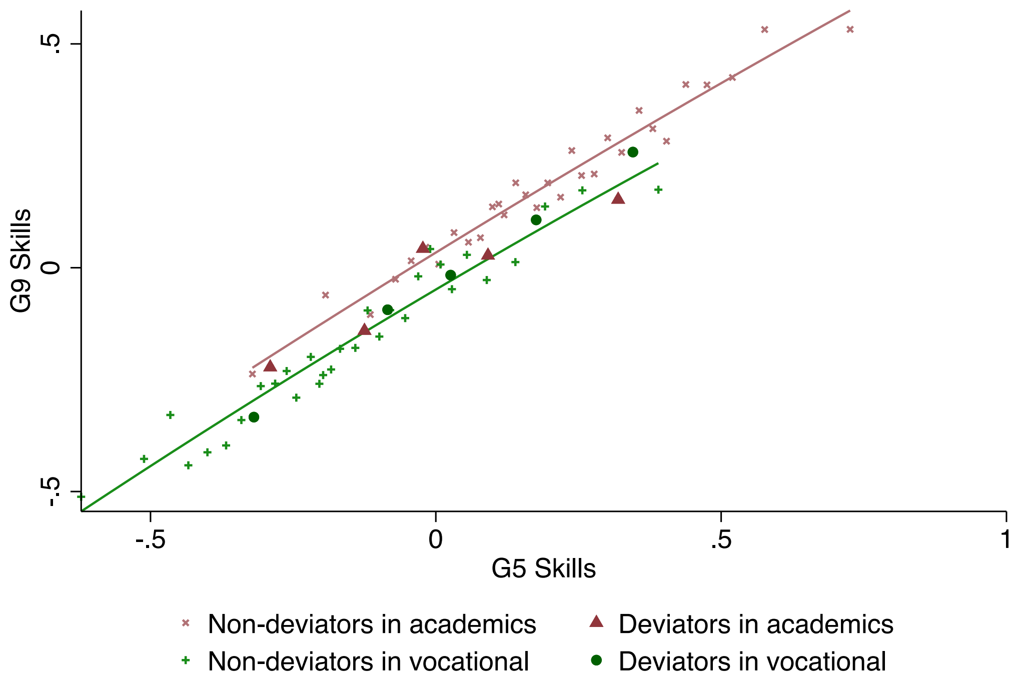
Notes: This table provides information on school track choice by parental education and teacher recommendation. Source: NEPS, Cohort 3.

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.

Moreover, Figure 2.B.2 plots the relationship between skills at the beginning of grade 5, that is right after the track decision and skills in grade 9, when children are typically 15 years old. The red and green lines are fitted values of grade 9 skills among the children who went to an academic or vocational track school, following the track recommendation they received from their primary school teacher. The triangles are the (grouped) learning outcomes of children who received a vocational track recommendation but go to an academic track school, the dots are the outcomes of children who received an academic track school recommendation but deviated to a vocational track school, instead. Visually, the learning outcomes of the deviators in academic track schools are on average below the predication of outcomes of children with similar skill levels in grade 5. The learning outcomes of children who deviated to vocational track schools are visually in line with or slightly above the prediction of what one would have expected a child with that skill level to learn in a vocational track school.

This is confirmed more formally by the regression results of grade 9 skills on grade 5 skills of children in academic or vocational track schools, when including a dummy variable that equals one whenever a child goes to that track against the recommendation of the primary school teacher, as shown in Table 2.B.3. When deviating to an academic track school, children incur statistically significant learning penalty. When deviating to a vocational track school despite an academic track school recommendation, children learn on average more, even though the effect is smaller and statistically insignificant. These results suggest that going against the track recommendation of the primary school teachers does not seem to benefit children's skill formation, underpinning our argument that it is likely not the fact that parents on average know better about the skills of their children than teachers and

Figure 2.B.2: Past and Future Skills by School Track and Deviator Status



Notes: This figure shows the skills of children in grade 9 by school track and deviation status from the recommended track as a function of their skills in grade 5. Non-deviating children are grouped in 30 bins by school track and represented in '×' and '+' symbols for those in the academic and vocational tracks, respectively. The lines show the quadratic prediction of grade 9 skills using grade 5 skills and grade 5 skills squared for non-deviating students in each school track. Deviating children are grouped in 5 bins by school track and represented in triangles and circle symbols for academics and vocational, respectively —data source: NEPS, Cohort 3.

that there are other reasons, perhaps preferences, that drive the decision to deviate from the recommended track.

2.B.3 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 lifespan (Blossfeld et al., 2019). To that end, the NEPS carefully designs and implements regular tests of the respondents' competencies along several domains, including reading comprehension, mathematical competence, and scientific literacy, which we use for the estimation of the child skill technology, but also domains such as information and communication technologies (ICT) literacy. In line with the guidelines set by the Program for International Student Assessment (PISA), the tests are generally designed to assess the extent to which

Table 2.B.3: Current Skills on Past Skills and Deviator Status

	Dependent Variable: Grade 9 Skills
Panel A: Cohort 3 - Academic Track Recommendations	
Grade 5 Skills	0.757***
	(0.026)
Downward Deviators (n = 84)	-0.062***
	(0.023)
Obs.	1,101
Panel B: Cohort 3 - Vocational Track Recommendations	
Grade 5 Skills	0.760***
	(0.033)
Upward Deviators (n = 84)	0.031
	(0.022)
Obs.	591

Notes: This table presents the coefficients of regressions of skills in grade 9 on past skills in grade 5 and deviation status for children with academic track teacher recommendations (Panel A) and for children with vocational track teacher recommendations (Panel B). Models control for parental education. Source: NEPS, Cohort 3.

children have learned the content of school curricula but also to judge a child's ability to use domain-specific knowledge to constructively engage with real-life problems (Neumann et al., 2013). The math test, for example, 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 for the majority simple multiple-choice questions with four response options. In addition, the sometimes include more complex multiple-choice questions, as well as short-constructed responses.[60] Each domain is tested using between 20 and 25 items, which usually takes around 30 minutes (Pohl and Carstensen, 2013).

In order to use these questions for the analysis of latent competencies, they need to be scaled. For reading comprehension, mathematical competence, and scientific literacy, the NEPS (similar to the PISA) uses a scaling procedure that is based on 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

[60] 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).

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).

The scaling model used by the NEPS for dichotomous items is the Rasch model (Rasch, 1960).[61] 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:

$$p(X_{vi} = 1) = 1 - p(X_{vi} = 0) = \frac{\exp(\theta_v - \sigma_i)}{1 + \exp(\theta_v - \sigma_i)}, \tag{2.50}$$

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.[62] This model is estimated via (weighted) conditional maximum likelihood under a normality assumption on latent ability.

There are several challenges that arise when scaling the test items: These include dealing with different response formats, the treatment of missing responses, adaptive testing, and linking tests across cohorts. An overview about the approaches undertaken by the NEPS to overcome these challenges is given in Pohl and Carstensen (2013). Table 2.B.4 exemplary describes our available NEPS samples of mathematics assessments by starting cohort and grade level.

2.B.4 Details on Child Skill Technology Estimation

Following the literature on child skill formation, we employ a linear measurement system for the logarithm of latent skills in each period that is given by

$$M_{i,k,j} = \mu_{k,j} + \lambda_{k,j}\theta_{i,j} + \epsilon_{i,k,j}, \tag{2.51}$$

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 in the domains of reading, maths and scientific literacy. The parameters $\mu_{k,j}$, and $\lambda_{k,j}$ denote the location and factor loading of latent log

[61] For polytomous items, the Partial Credit Model is used, which is a generalization of the Rasch model (Masters, 1982).

[62] It is interval-scaled as Ballou (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.

Table 2.B.4: 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.

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.[63] 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):

$$\lambda_{k,j} = \frac{Cov(M_{k,j}, M_{k',j})}{Cov(M_{1,j}, M_{k',j})} \tag{2.52}$$

for all $k, k' > 1$ and $k \neq k'$. Rescaling the measures by their identified location and scale

[63]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.

parameters then gives us error-contaminated measures of latent skills for each period as

$$\theta_{i,j} = \frac{M_{i,k,j} - \mu_{k,j}}{\lambda_{k,j}} - \frac{\epsilon_{i,k,j}}{\lambda_{k,j}} = \widetilde{M}_{i,k,j} - \frac{\epsilon_{i,k,j}}{\lambda_{k,j}}. \tag{2.53}$$

Equipped with identified latent variables up to measurement error for all periods, we can plug these into the empirical analogue of the child skill technology (2.27), which yields

$$\begin{aligned}\widetilde{M}_{i,k,j+1} = \kappa_{0,j} + \kappa_{1,j}\widetilde{M}_{i,k,j} + \kappa_{2,j}\widetilde{M}^2_{i,k,j} + \kappa_{3,j}\overline{\widetilde{M}}_{-i,j,S} \\ + \kappa_{4,j}(\widetilde{M}_{i,k,j} - \overline{\widetilde{M}}_{j,S})^2 + \kappa_{5,j}E_i + \zeta_{i,k,j+1},\end{aligned} \tag{2.54}$$

where $\overline{\widetilde{M}}_{-i,j,S}$ refers to the average 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 average 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 (2.54) will be biased. To account for that, we follow the literature and use Bartlett factors scores to aggregate the different measures into an unbiased score (Agostinelli et al., 2023). As indicated before, we use maths, reading, and science test scores, which we have available across different cohorts and grades (years) in school: We use Cohort 2 for grades 1 and 4, corresponding to the primary school stage in our model (i.e. period $j = 2$); Cohort 3 for grades 5 to 9, which correspond to the first stage of secondary school in the model (i.e. period $j = 3$); and Cohort 4 for grades 9 to 12, corresponding to the second stage of secondary school in the model (i.e. period $j = 4$). Note that in grade 7, children only take two tests, which is why we cannot construct the latent skills. In addition, in grade 12, the maths test differs by track, and only children in the academic track take the science test. Consequently, in grade 12, we can only create latent skills for children in the academic school track. [64]

Table 2.B.5 summarizes the estimated coefficients of the child skill technology (2.27) using the identified latent variables as describes above in columns (1) and (2), or using math test scores directly in columns (3) and (4). The estimates differ slightly depending on whether we use longitudinal weights or not, but overall are quite consistent. Table 2.B.6 performs the estimation where the squared distance to track average term in (2.27) is distributed, such that we include directly the interaction between own skill and track average. The estimated coefficient is positive, statistically significant in most specifications and not statistically different from $-2\hat{\omega}_4$, lending support to our modeling assumptions.

Table 2.B.5: Robustness Checks: Child Skill Technology Parameters Estimates

Grade 9 on Grade 5		Skills		Math scores	
Dependent Variable: $\theta_{i,j+1}$		(1)	(2)	(3)	(4)
$\hat{\omega}_{1,3}$	$\theta_{i,j}$	0.664***	0.647***	0.519***	0.517***
		(0.022)	(0.025)	(0.025)	(0.030)
$\hat{\omega}_2$	$\bar{\theta}_{-i,j,S}$	0.003	0.028	0.022	0.025
		(0.020)	(0.021)	(0.024)	(0.031)
$\hat{\omega}_3$	$\theta^2_{i,j}$	0.008*	0.006	0.010**	0.015**
		(0.004)	(0.005)	(0.005)	(0.006)
$\hat{\omega}_4$	$(\theta_{i,j} - \bar{\theta}_{j,S})^2$	-0.011*	-0.013**	-0.012*	-0.020**
		(0.006)	(0.006)	(0.007)	(0.008)
$\hat{\omega}_{5,3}$	$E = 1$	0.034***	0.033***	0.033***	0.045***
		(0.010)	(0.012)	(0.012)	(0.014)
Obs.		1,847	1,676	2,084	1,708
Weights		No	Yes	No	Yes

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. In Columns (2) and (4), all observations are weighted using longitudinal weights, while in Columns (1) and (3), they are not. Standard errors are clustered at the school level. Columns (1) and (2) present the results for latent skills corrected for measurement errors, while columns (3) and (4) present the results for uncorrected latent skills of maths grades. Models control for year of birth, gender, and school-fixed effects. Source: NEPS.

[64] 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.

Table 2.B.6: Robustness Checks: Alternative Child Skill Technology Parameters Estimates

Grade 9 on Grade 5		Skills		Math scores	
Dependent Variable: $\theta_{i,j+1}$		(1)	(2)	(3)	(4)
$\hat{\omega}_{1,3}$	$\theta_{i,j}$	0.657***	0.626***	0.515***	0.505***
		(0.021)	(0.024)	(0.023)	(0.028)
$\hat{\omega}_2$	$\bar{\theta}_{-i,j,S}$	0.001	0.024	0.020	0.018
		(0.020)	(0.021)	(0.024)	(0.030)
$-2 * \hat{\omega}_4$	$\theta_{i,j} * \bar{\theta}_{j,S}$	0.018**	0.014	0.022**	0.029**
		(0.009)	(0.010)	(0.010)	(0.012)
$\hat{\omega}_{5,3}$	$E = 1$	0.034***	0.034***	0.033***	0.048***
		(0.010)	(0.012)	(0.012)	(0.014)
Obs.		1,847	1,676	2,084	1,708
Control for $\bar{\theta}^2_{j,S}$		Yes	Yes	Yes	Yes
Weights		No	Yes	No	Yes

Notes: This table presents the coefficients of regressions of skills in grade 9 on skills in grade 5, the average skill level of peers, the interaction between child skills and the average skill in the track, the average skill in the track squared, and parent's education dummy. In Columns (2) and (4), all observations are weighted using longitudinal weights, while in Columns (1) and (3), they are not. Standard errors are clustered at the school level. Columns (1) and (2) present the results for latent skills corrected for measurement errors, while columns (3) and (4) present the results for uncorrected latent skills of maths grades. Models control for year of birth, gender, and school-fixed effects. Source: NEPS.

2.B.5 Details on the Data Moments used in the MSM Estimation

In this section, we present details of the data moments that we use as calibration targets in the method of simulated moments estimation.

Table 2.B.7 presents the distribution of students across school tracks and education levels. We use two main sources to compute those shares. First, whenever available, we use official statistics that are reported in the education report: 44% of students are in the academic track, and children from college parents are 2.27 more likely to graduate from college than students from non-college parents (p. 156 in Bildungsberichterstattung, 2018). Second, we complement this data using Cohort 4 of the NEPS dataset: 35% of the parents graduated from college, children from college parents are 1.95 more likely to attend the academic school track than children from non-college parents and students in the academic track are 5.23 more likely to attend college than students in the vocational track.[65] In addition, 23% of college

[65]In the NEPS dataset, we only have college attendance and not graduation. We use the ratio of college attendance by groups as a proxy for the ratio of college graduation.

parents deviate from the vocational recommendation, and 16% of non-college parents deviate from the academic recommendation as argued above. Finally, the model is in stationary equilibrium, which implies that 35% of students graduate from college, the same share as the share of college parents. All the remaining shares are computed so that the model distribution is internally consistent.

Table 2.B.8 describes the evolution of child skills over time and across groups using the identified latent variable (Columns (1) and (2)) or maths scores directly (Columns (3) and (4)). As before, we use different cohorts of NEPS for the estimation: Cohort 2 for grades 1 to 4, Cohort 3 for grades 5 to 9, and Cohort 4 for grades 9 to 12. We also report the results for grade 12 using Cohort 3, but we prefer the results from Cohort 4 as the number of observations is greater (see Table 2.B.4). For a given individual, the correlation across skills increases over time, from 0.61 between grades 1 and 4 to 0.74 between grades 9 and 12 (Table 2.B.8 Column (1), using the latent skill and longitudinal weights). The differences in average skills across groups are also increasing over time: from 0.541 SD in grade 1 to 0.677 SD in grade 9 by parents' education, and from 0.847 SD in grade 1 to 1.036 SD in grade 9 by school track (Table 2.B.8 Column (1), using the latent skill and longitudinal weights).[66]

Tables 2.B.9 and 2.B.10 report details on the estimation of academic school track attendance on child skills at the beginning of secondary, or end of primary school, as well as the estimation of college attendance on past skills and school track. We use the latter estimates to calibrate the college costs in our model, while the former serve as untargeted tests.

2.C Comparison of Model-predicted Effects with Empirical Estimates

Generally, empirical estimates of the effects of between-school tracking policies on the average learning outcomes of children offer no clear consensus, as identification of causal effects is made difficult by severe endogeneity issues (Hanushek and Wößmann, 2006). Using the same dataset as we do, Matthewes (2021) compares the mathematics and reading achievement outcomes of school children in non-academic school tracks in Germany who benefit from two more years of comprehensive school in some federal states versus those that are already tracked in these years in a difference-in-differences framework. He finds that later tracking even improves average achievement outcomes. However, in contrast to our setup, his analysis does not consider children in academic track school who are already separated but compares

[66] To compute the difference by school track in grade 1, we use the panel structure of NEPS, and allocate students in grade 1 to school track according to their actual school track in grade 7.

Table 2.B.7: Distribution of Students, School Tracks and Parental Education

Statistics	Value	Source	Comment
% of college parents	35%	NEPS Cohort 4	
Track choice			
% in ac. track	44%	Education report p.110	42% in NEPS Cohort 4
Ratio % ac. track if college parents to % if non-college parents	2.06	NEPS Cohort 4	
% in ac. track if $E = 1$	66%	Implied	
% in ac. track if $E = 0$	32%	Implied	
Track recommendation			
Deviation if recom. S = 0 and $E = 1$	23%	NEPS Cohort 4	
Deviation if recom. S = 1 and $E = 0$	16%	NEPS Cohort 4	
% ac. recom.	44%	Implied	
% ac. recom. if $E = 1$	56%	Implied	
% ac. recom. if $E = 0$	38%	Implied	
College graduation			
% who graduate from college	35%	Model assumption	
Ratio % college if academics to % if vocational	6.27	NEPS Cohort 4	
% college if academics	66%	Implied	
% college if vocational	11%	Implied	
Ratio % college if college parents to % if non-college parents	3.20	Meyer-Guckel et al. (2021)	Ratio computed from Figure 1, where 64% of children from college parents are bachelor graduates versus 20% of children from non-college parents.
% college if college parents	63%	Implied	
% college if non-college parents	20%	Implied	

Notes: This table provides information on the distribution of students by school track, college education, and parental education with corresponding sources.

Table 2.B.8: Evolution of Skills

Statistics	Skills (1)	Skills (2)	Math grades (3)	Math grades (4)	Source
Group Differences					
Differences in average skills by parental education (in standard deviations)					
Grade 1	0.530	0.541	0.459	0.462	NEPS Cohort 2
Grade 5	0.658	0.647	0.605	0.579	NEPS Cohort 3
Grade 9	0.672	0.774	0.598	0.697	NEPS Cohort 3
Grade 9	0.710	0.677	0.659	0.623	NEPS Cohort 4
Differences in average skills by school track (in standard deviation)					
Grade 1	0.840	0.847	0.767	0.769	NEPS Cohort 2
Grade 5	1.104	1.022	1.067	0.986	NEPS Cohort 3
Grade 9	1.058	1.089	1.040	1.113	NEPS Cohort 3
Grade 9	1.110	1.036	1.062	0.998	NEPS Cohort 4
Rank-Rank correlations					
Panel A: All students					
Grades 1 to 4	0.72	0.72	0.59	0.58	NEPS Cohort 2
Grades 5 to 9	0.79	0.79	0.71	0.71	NEPS Cohort 3
Panel B: Academic students					
Grades 1 to 4	0.62	0.61	0.46	0.45	NEPS Cohort 2
Grades 5 to 9	0.68	0.69	0.57	0.59	NEPS Cohort 3
Grades 9 to 12	0.74	0.72	0.65	0.66	NEPS Cohort 3
Grades 9 to 12	0.74	0.74	0.66	0.59	NEPS Cohort 4
Panel C: Vocational Students					
Grades 1 to 4	0.64	0.64	0.53	0.50	NEPS Cohort 2
Grades 5 to 9	0.74	0.75	0.63	0.64	NEPS Cohort 3
Weights	No	Yes	No	Yes	

Notes: This table provides information on average differences in skills in one standard deviation unit by parental background and school track over time as well as skill rank-rank correlations. In columns (2) and (4), all observations are weighted with longitudinal weights, while in columns (1) and (3), they are not. Columns (1) and (2) present the results for latent skills corrected for measurement errors, while columns (3) and (4) present the results for uncorrected latent skills of maths grades. Sources are mentioned in the last column.

Table 2.B.9: School Track on Past Skills

Dependent Variable: Academic School Track	
Panel A: Cohort 3 - Grade 5	
$\theta_{i,j-1}$	0.877***
	(0.019)
Obs	3,888
Panel B: Cohort 2 - Grade 4	
$\theta_{i,j-1}$	0.745***
	(0.027)
Obs	2,299

Notes: This table presents the coefficients of regressions of the academic school track on past skills in grade 4 (Panel A) or in grade 5 (Panel B). Models control for year of birth and gender fixed effects. Source: NEPS.

Table 2.B.10: College on Past Skills and School Track

Dependent Variable: College Attendance	
Panel: Cohort 4 - Grade 9	
$\theta_{i,j}$	0.395***
	(0.015)
S	0.407***
	(0.011)
Obs	10,074
Variance of residuals	0.137

Notes: This table presents the coefficients of regressions of college attendance on past skills (grade 9) and school track. We control for year of birth and gender fixed effects. Source: NEPS.

children in two different non-academic tracks.[67] The results in Matthewes (2021), therefore, are not directly informative about the effects of a broad comprehensive school reform, which places *all* school children in schools of the same track for a longer period of time. However, they suggest that the effects of tracking may be heterogeneous in the sense that it could be particularly children from lower socio-economic backgrounds that benefit from de-tracking reforms.

This is corroborated by empirical evidence from Scandinavian countries, who have all undergone comprehensive school reforms in the last 60 years and often find that longer comprehensive schooling decreases the effect of family background on educational attainment (see, for instance Meghir and Palme, 2005; Aakvik et al., 2010; Pekkala Kerr et al., 2013).[68] Similarly to Meghir and Palme (2005), who study an increase of compulsory schooling to nine years from seven or eight years in Sweden, we find a negative effect of our policy on attainments for children from college parents (-3% in end-of-school skills) but a positive effect for children from non-college parents (+5% in end-of-school skills). Our results are also in line with evidence that attendance at academic track schools becomes less dependent on the parental background when tracking occurs later. In particular, we find an increase in academic shares for college and non-college parents' children, but relatively more so for non-college parents' children.

In terms of the effects of school tracking policies on inequality in learning outcomes, most existing evidence comes from comparisons of early and late tracking systems across countries (Hanushek and Wößmann, 2006; Brunello and Checchi, 2007).[69] They consistently find that tracking raises educational inequality as measured by child achievement test scores. Our result of lower heterogeneity in child skills during secondary school is thus in line with these findings.

Finally, while the reduced impact of the family background on educational attainment in *secondary school* is often already interpreted as evidence for improvements in social mobility following de-tracking reforms, it does not necessarily follow that such improvements lead to

[67] Traditionally, the school system in many federal states in Germany consisted of three tracks. One academic track (*Gymnasium*) and two non-academic tracks that differ much less in terms of their curriculum than between academic and non-academic tracks.

[68] Since the reforms in these countries came together with other education policy changes, in particular with more mandatory schooling years, the estimated effects can often not be unequivocally attributed to the tracking regime change. Studies about de-tracking reforms in Britain (e.g. Pischke and Manning, 2006) also often arrive at mixed results. Piopiunik (2014), who study an *increase* in tracking in one of the federal states in Germany, Bavaria, also led to learning losses for the lower-skilled children.

[69] As pointed out by Waldinger (2006) or Betts (2011), these studies often come with significant identification challenges given the unclear classification of countries in early and late tracking systems or the possibility of unobserved differences driving the results.

a lower association between child and parental outcomes later in life.[70] Similarly, smaller inequality in test scores does not necessarily need to translate into lower cross-sectional inequality in terms of labor market outcomes. As argued before, an assessment of the effects of school tracking policies on these outcomes is challenging as it requires the consideration of general equilibrium effects that a change in the skill composition of students may entail on the labor market. Most existing empirical evidence, however, comes from relatively short-term evaluations of tracking reforms that cannot consider such effects.

To the best of our knowledge, the only empirical estimates on the effect of a broad comprehensive school reform on the intergenerational elasticity of income come again from the Nordic countries (Holmlund, 2008; Pekkarinen et al., 2009). In particular, the reform in Finland, undertaken subsequently across regions in the 1970s, is similar in scope to our experiment as it postponed the track allocation from age 11 until age 16. Pekkarinen et al. (2009) find that the elasticity between fathers' and sons' relative earnings declined by seven percentage points due to the reform (from 0.3 to 0.23). Our model also predicts a decrease in the intergenerational income elasticity, yet the effect is quantitatively smaller. Some of this difference is likely due to the fact the reform in Finland simultaneously also changed the average length of schooling, and the content of the curriculum in schools towards a more academic orientation and went from a largely private to a public school system.

2.D 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

[70] For example, Malamud and Pop-Eleches (2011) analyses the effects of a school-tracking age postponement in Romania. While they find that children from disadvantaged backgrounds were significantly more likely to attend and graduate from academic track schools following the reform, this did not lead to an increased share in the college graduation probabilities of these disadvantaged children, which they attribute to the same overall share of college slots available pre- and post-reform. The quantitative results of our model similarly predict, that while postponing tracking increases mobility in school track choice, this does not lead to higher mobility in college attainment. This effect is driven by parental-education-specific college tastes.

function f, defined by:

$$\begin{aligned}\theta_{j+1} &= f(\theta_j, P_j^S, \bar{\theta}_j^S, E) && (2.55)\\ &= \kappa\theta_j + \ \alpha\bar{\theta}_j^S + \ + \beta\ P_j^S + \gamma\ \theta_j\ P_j^S - \frac{\delta}{2}\ P_j^{S2} + \ \zeta E, && (2.56)\end{aligned}$$

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:

$$\begin{aligned}\hat{\theta}_j &= \theta_j + \epsilon_{\theta,j}\\ \epsilon_{\theta,j} &\sim \mathcal{N}(0, \sigma^2_{\epsilon_\theta}).\end{aligned} \tag{2.57}$$

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:[71]

$$\begin{aligned}\tilde{\theta}_j &= k\ \hat{\theta}_j + (1-k)\tilde{\theta}_j^P\\ \sigma_j^2 &= \sigma_{j-1}^2 - k\sigma_{j-1}^2\\ k &= \frac{\sigma_{j-1}^2}{\sigma_{j-1}^2 + \sigma^2_{\epsilon_\theta}},\end{aligned} \tag{2.58}$$

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

$$\begin{aligned}\theta_{j+1} &= f(\theta_j, \bar{\theta}_j^S, E)\\ &= \frac{\beta^2}{2\delta} + \ (\kappa + \frac{\beta\gamma}{\delta})\theta_j + \ (\alpha)\bar{\theta}^S - \ \frac{\gamma^2}{2\delta}\bar{\theta}^{S2} + \ \frac{\gamma^2}{\delta}\theta_j\bar{\theta}^S + \zeta\ E\end{aligned}$$

[71] We could assume the first initial prior to be equal to the signal they receive in $j = 1$.

$$= \omega_0 + \ \omega_1 \ \theta_j + \ \omega_2 \ \bar{\theta}^S + \ \omega_4 \ \bar{\theta}^{S^2} - \ 2\omega_4 \ \theta_j \bar{\theta}^S + \ \omega_5 \ E.$$

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. The noise in the preference shifter can be regarded as a reduced form of capturing the imprecision in the parents' perception of their child's skills.

Chapter 3

Aggregate and Distributional Effects of School Closure Mitigation Policies: Public versus Private Education[1]

with Minchul Yum[2]

Abstract: Recent studies highlight the adverse effects of school closures in terms of average lifetime income loss, cross-sectional inequality, and intergenerational mobility. We use a simple model of human capital formation to compare two policy instruments that can address these negative consequences: direct public provision, such as through an extension of school time, and the provision of private education subsidies. We demonstrate that the effects of these policies on inequality and mobility depend crucially on the degree of substitutability between private and public inputs in the production of human capital.

[1] We thank Youngsoo Jang for useful discussions. Financial support from the German Research Foundation (DFG) through CRC TR 224 (Project A04) is gratefully acknowledged.
[2] Department of Economics, University of Southampton

3.1 Introduction

During the COVID-19 pandemic, many governments unprecedentedly closed schools for extensive periods. This not only incurs considerable learning losses among affected children in the short run (Blanden et al., 2023; Werner and Woessmann, 2023), but may also entail significant adverse long-run consequences in terms of future income and welfare losses (Agostinelli et al., 2022; Fuchs-Schündeln et al., 2022; Jang and Yum, forthcoming). Moreover, as the learning losses and parental behavioral responses are heterogeneous across the income distribution, the school closures may lead to higher inequality and impair intergenerational mobility (Jang and Yum, forthcoming).

Against this backdrop, various policy interventions have been discussed to counteract the detrimental consequences of school closures (Zviedrite et al., 2021). To highlight the potential long-run implications of such mitigation policies, we present a model, which is simple yet considers the sophisticated nature of how private and public education investments interact as inputs into the production of human capital. We explore two mitigation policies: extending public schooling time, such as during the summer, and implementing means-tested subsidies for private education. Our results suggest that the elasticity of substitution between private and public inputs plays an important role in shaping the effects of these policies. In particular, if private and public inputs are highly substitutable, both mitigation policies can bring down inequality and improve intergenerational mobility. If the two inputs are complementary, however, extra public schooling can aggravate inequality and harm mobility, which is in contrast to government subsidies to private education.

This is informative, as even though much of the existing macro literature assumes that the degree of substitutability between public and private is very large, empirical evidence on their elasticity is far less clear. For example, Gelber and Isen (2013) find evidence that larger public investments *crowd in* parental investments into their children, which is incompatible with a model in which public and private investments are substitutes (e.g., Becker and Tomes, 1979). Generally, the degree to which private investments can replace public schooling in the production of child human capital likely depends, among other things, on the period length, the age and education stage of the child, and the presence and quality of private education markets.

The contribution of our paper is thus to offer a novel policy insight as to the importance of substitutability between public and private education for the long-term effects of mitigation policies on inequality and mobility. We thereby complement the analysis in Fuchs-Schündeln et al. (2023), who consider a school-time extension mitigation policy using a rich quantitative model, but do not consider mobility consequences and the role of substitutability between public and private education.

3.2 Model

The model consists of three periods, indexed by $t = 0, 1, 2$. A household consists of an adult parent and a child and draws a time-constant endowment $m \in \{m_l, m_h\}$ with an equal probability, such that we have low- and high-income households. We abstract from savings.

At the beginning of $t = 0$, the child draws a learning ability $\phi \in \{\phi_l, \phi_h\}$, which is correlated with m. Specifically, for $k = l, h$, households holding m_k draw ϕ_k with a probability of p_ϕ. The learning ability affects the production of a child's human capital over time, h_t.[3] The initial human capital level h_0 is set to one. Human capital then evolves as a function of past human capital, learning ability, and private parental-, as well as public schooling inputs. We think of the initial period $t = 0$ as capturing the early education stage of a child (e.g., pre-and primary school), and of period $t = 1$ as the second education stage of a child (e.g., secondary school). The final period $t = 2$ then captures the adult period of the child generation, where the final human capital level h_2 realizes.

In $t = 0$, a household with endowment m and child learning ability ϕ then solves:

$$
\begin{aligned}
V(\phi, m, t = 0) &= \max_{c_t, e_t \geq 0} \{\log c_t + V(h_{t+1}, \phi, m, t = 1)\} \\
\text{subject to} & \\
c_t + e_t &= m \\
h_{t+1} &= \phi \left\{ (e_t/\bar{e})^\psi + (\varsigma g)^\psi \right\}^{\frac{\alpha}{\psi}} h_t^{1-\alpha},
\end{aligned}
\tag{3.1}
$$

where c_t denotes consumption and e_t denotes all private investments into the production of human capital, divided by its mean.[4] Private investments and time-invariant public investments, which we denote by g, are aggregated using a CES aggregator. The elasticity of substitution between the two investments is shaped by $\psi \leq 1$. The parameter $\varsigma < 1$ captures the productivity loss of public schooling due to school closures and is used to simulate (unexpected) school closure shocks later (Fuchs-Schündeln et al., 2022; Jang and Yum, forthcoming). As is common, the production of human capital is then of the Cobb-Douglas form, where total investments and past human capital are the input factors with unit elasticity of

[3] The correlation of the learning ability of children and their parents' endowments allows us to parsimoniously capture various sources of intergenerational persistence not due to endogenous investments.

[4] Private investments may also include the time parents spend with their children, which we can think of as incurring an opportunity cost measured by foregone wages.

substitution and the factor shares are given by α.[5] Finally, the learning ability ϕ plays a role of total factor productivity. We abstract from future discounting.

The decision problem in $t = 1$ is similar,

$$\begin{aligned} V(h_t, \phi, m, t=1) &= \max_{c_t, e_t \geq 0} \{\log c_t + \eta \log h_{t+1}\} \\ \text{subject to} & \\ c_t + (1 - s(m))e_t &= m \\ h_{t+1} &= \phi \left\{ (e_t/\bar{e})^{\psi} + (\gamma g)^{\psi} \right\}^{\frac{\alpha}{\psi}} h_t^{1-\alpha}, \end{aligned} \tag{3.2}$$

where we assume a warm-glow altruism motive for parents governed by $\eta > 0$. Moreover, we introduce two policy tools that have been discussed as measures to counteract the learning losses induced by school closures: (i) prolonged school periods that make up for (some) of the lost time in public schools as governed by $\gamma > 1$ (Fuchs-Schündeln et al., 2023); and (ii) means-tested subsidies to private education (Yum, 2023), given by $s(m_l) = s \geq 0$ and $s(m_h) = 0$.

3.3 Results

3.3.1 Calibration of Two Baseline Model Economies

In light of the unclear evidence on the elasticity of substitution between private and public investments in the human capital formation of children in the literature, we calibrate two versions of the baseline model: In Model 1, we set $\psi = 0.6$, such that the substitution elasticity is 2.5, so that private and public investments are gross substitutes, albeit imperfect ones.[6] In Model 2, we set $\psi = -1$ (so that the elasticity is 0.5), implying that both inputs are gross complements, reflecting pertinent findings in the micro literature (Gelber and Isen, 2013).

In both versions of the baseline model, there are no school closures, such that $\varsigma = 1$, and no government interventions (i.e., $\gamma = 1$ and $s(m) = 0$). Moreover, we set $\alpha = 0.25$ throughout, but our qualitative results are robust to this parameter. We parameterize $m_l = 1 - m_\delta$, $m_h = 1 + m_\delta$, and $\phi_l = \phi_\mu(1 - \phi_\delta)$ and $\phi_h = \phi_\mu(1 + \phi_\delta)$. We then have five parameters, $\{p_\phi, \eta, m_\delta, \phi_\delta, \phi_\mu\}$, which we internally calibrate to ensure that the equilibrium distribution

[5] We have explored a specification that allows for strong dynamic complementarity, and our qualitative findings remain robust to this consideration.

[6] For example, Kotera and Seshadri (2017) estimate an elasticity of substitution of 2.43, and Arcalean and Schiopu (2010) one of 1.31 for primary and secondary education stages.

Table 3.1: Internally Calibrated Parameters

Parameter	**Model 1** $\psi = 0.6$ (ES = 2.5)	**Model 2** $\psi = -1.0$ (ES = 0.5)	Target Statistics Description	Value
p_ϕ	0.567	0.604	IGE	0.34
η	0.920	1.251	Avg e/income	0.10
m_δ	0.800	0.800	Gini Adult	0.40
ϕ_δ	0.419	0.449	Gini Child	0.40
ϕ_μ	0.723	1.140	Avg h_2	1.00

in both baseline model versions exactly matches the intergenerational elasticity, the ratio of average monetary investments to average income, and the Gini coefficient of incomes in the US, as summarized in Table 3.1. The last parameter, ϕ_μ, determines the scale of ϕ, which is used to normalize the average h_2 to one.

3.3.2 Aggregate and Distributional Effects of School Closures

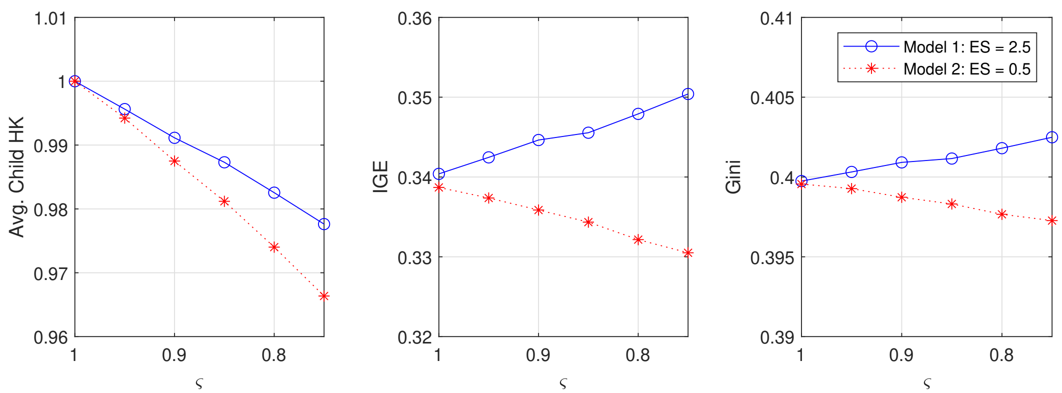
We first assess our simple model's predictions regarding the aggregate and distributional effects of school closures in $t = 0$ by varying the parameter $\varsigma \leq 1$, which lowers the public (schooling) input into the human capital formation in $t = 0$. We focus on three outcomes: (i) average final human capital of the child generation in $t = 2$, h_2, which serves as our measure of long-run efficiency; (ii) the intergenerational elasticity, the slope coefficient from regressing $log(h_2)$ on $log(m)$, which we take as our measure of intergenerational mobility; and (iii) the Gini coefficient of h_2, serving as our measure of long-term cross-sectional inequality for the child generation.

As shown in Figure 3.1, Model 1 (blue solid line) predicts that school closures lead to aggregate losses in terms of human capital, lower intergenerational mobility, and larger cross-sectional inequality. Thus, this simple model qualitatively replicates the findings of Jang and Yum (forthcoming). In contrast, Model 2 (red dotted line) predicts that, when the elasticity of substitution between private and public inputs in the production of human capital is very low, school closures can even lead to higher mobility and lower inequality, while average human capital still drops at a faster rate than in Model 1.[7]

Fundamentally, the different results arise because when private investments are good substitutes for lower public investments, richer parents can more easily offset the effects of lost schooling time by increasing their private investment than less wealthy parents. Thus,

[7] Jang and Yum (forthcoming) do not explore the case where public and private education investments are gross complements, as investigated in Model 2 in this study.

Figure 3.1: Effects of School Closures without Mitigation Policies



Notes: A longer duration of school closures is captured by a lower value of ς.

the human capital of children from different parental backgrounds diverges, which increases inequality and reinforces the correlation between child and parental economic outcomes in Model 1. If private and public inputs are more complementary in producing child human capital, however, this mechanism reverses. That is, the higher private inputs of richer parents become less effective when schools close as they are only productive when complemented with public inputs. Thus, differences between children from different parental backgrounds can even decrease, resulting in lower inequality and higher intergenerational mobility.

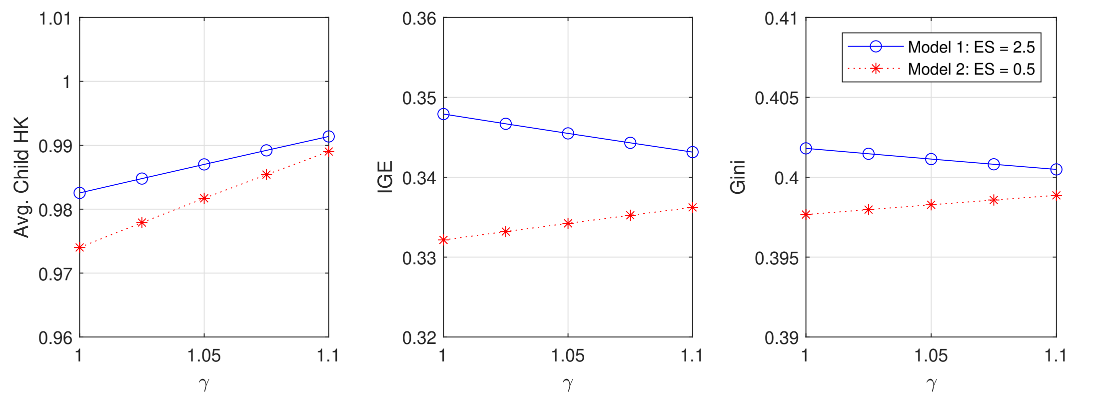
3.3.3 Mitigating the School Closure Effects

As our main exercises, we now explore the effects of two school closure mitigation policies: extra public schooling and means-tested private education subsidies. To that end, we consider different degrees of each mitigation policy in $t = 1$ in both model versions, after all decisions in $t = 0$ with school closures are made.[8] We focus on the case with $\varsigma = 0.8$.

In the first exercise, we increase $\gamma \geq 1$ in $t = 1$. Analogously to the school closures, this can be interpreted as prolonged schooling during the second education stage, for example by leaving schools open during the holidays, or by extending regular schooling days. As shown in the left panel of Figure 3.2, such policies indeed succeed in alleviating the human capital losses associated with school closures. This is true for both Model 1 and 2. However, while make-up schooling raises intergenerational mobility and lowers cross-sectional inequality in Model 1, it does the opposite in Model 2. This again reflects the argument that when private and

[8] Note that, under our logarithmic assumptions on the utility from h_2 in (3.2), optimal private investments e_1 are independent of h_1. For that reason, even if the policy interventions were anticipated, our results remain unchanged.

Figure 3.2: Mitigation Policy 1: Prolonged Schooling

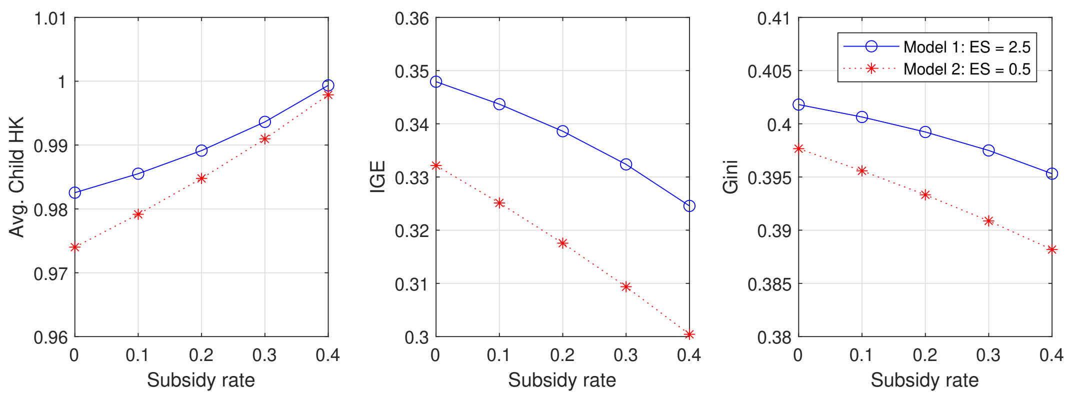


Notes: Prolonged schooling (γ) is considered as a mitigation policy following school closures with $\varsigma = 0.8$.

public investments are substitutable, as in Model 1, children from poorer households benefit relatively more from make-up public schooling, as their parents could not compensate the learning losses resulting from closures through private inputs as effectively as richer parents. Thus, the differences in human capital between rich and poor children decrease, and the correlation between parental and child outcomes drops. In contrast, when the two inputs are complementary, children from richer households disproportionately gain from prolonged schooling as they also benefit from higher private inputs that make schooling productive. In Model 2, universal make-up schooling can therefore aggravate inequalities and hamper social mobility.

In the second exercise, we raise $s \geq 0$ in $t = 1$, the subsidy rate for private education spending such as coupons for purchasing a tablet or online courses, given to parents of the low-endowment type. The effects are shown in Figure 3.3. In both Model 1 and Model 2, the policy successfully mitigates the average human capital losses resulting from school closures. At the same time, intergenerational mobility increases and cross-sectional inequality falls, regardless of the substitutability between public and private investments in human capital production. Thus, a means-tested private education subsidy can potentially prevent the exacerbation of inequality and adverse effects on mobility, in cases where the elasticity of substitution is especially low. Of course, in such a world, inequality and immobility after school closures would already be lower to begin with.

Figure 3.3: Mitigation Policy 2: Means-tested Private Education Subsidy



Notes: Education subsidies provided to low-income parents ($s(m)$) are considered as a mitigation policy following school closures with $\varsigma = 0.8$.

3.4 Conclusion and Discussion

This paper demonstrates that the long-term consequences of school closure mitigation policies in terms of inequality and intergenerational mobility depend crucially on the elasticity of substitution between public and private investments in the human capital formation of children.[9] Our results illustrate that in a stark case when both inputs are complementary, untargeted mitigation policies such as universal schooling extensions may lead to the perhaps unintentional consequences of increasing inequality and lowering mobility. An important task for researchers and policymakers in the design of such policies is thus to consider how private, parental, and public schooling investments interact across different contexts, such as in the short- or long run, at different ages and educational stages of children, across different domains of skills like cognitive and non-cognitive, or in the presence of professional private education and tutoring markets.

Finally, despite our calibration, our analysis here serves the purpose of delivering these arguments mostly qualitatively. A serious quantitative evaluation of school closure mitigation policies would require a richer overlapping generations model that incorporates several further potentially important aspects. For example, a more accurate comparison of policies should take into account the financing costs of policies. In addition, a combination of the two policies we consider or a targeted prolongation of public schooling just for disadvantaged children is conceivable. These interesting and important investigations are left for future work.

[9] See Glomm and Kaganovich (2003) and Aliprantis and Carroll (2018) who make related points about the sensitivity of distributional or sorting outcomes to this elasticity in different contexts.

Bibliography

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

Aliprantis, D. and D. R. Carroll (2018): “Neighborhood dynamics and the distribution of opportunity,” *Quantitative Economics*, 9, 247–303.

Arcalean, C. and I. Schiopu (2010): “Public versus private investment and growth in a hierarchical education system,” *Journal of Economic Dynamics and Control*, 34, 604–622.

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

Blanden, J., M. Doepke, and J. Stuhler (2023): “Educational inequality,” in *Handbook of the Economics of Education*, Elsevier BV, 405–497.

Fuchs-Schündeln, N., D. Krueger, A. Kurmann, E. Lalé, A. Ludwig, and I. Popova (2023): “The fiscal and welfare effects of policy responses to the Covid-19 school closures,” *IMF Economic Review*, 71, 35–98.

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

Gelber, A. and A. Isen (2013): “Children’s schooling and parents’ behavior: Evidence from the Head Start Impact Study,” *Journal of Public Economics*, 101, 25–38.

Glomm, G. and M. Kaganovich (2003): “Distributional effects of public education in an economy with public pensions,” *International Economic Review*, 44, 917–937.

Jang, Y. and M. Yum (forthcoming): “Aggregate and intergenerational implications of school closures: a quantitative assessment,” *American Economic Journal*.

Kotera, T. and A. Seshadri (2017): “Educational policy and intergenerational mobility,” *Review of Economic Dynamics*, 25, 187–207.

Werner, K. and L. Woessmann (2023): “The legacy of COVID-19 in education,” *Economic Policy*, eiad016.

Yum, M. (2023): “Parental time investment and intergenerational mobility,” *International Economic Review*, 64, 187–223.

Zviedrite, N., J. D. Hodis, F. Jahan, H. Gao, and A. Uzicanin (2021): “COVID-19-associated school closures and related efforts to sustain education and subsidized meal programs, United States, February 18–June 30, 2020,” *PloS one*, 16, e0248925.

This dissertation is the result of my own work, and no other sources or means, except the ones listed, have been employed.

June 3, 2024, Mannheim

Lukas Mahler

Curriculum Vitae

2018 - 2024	University of Mannheim
	Ph.D. in Economics
2016 - 2018	University of Mannheim
	M.Sc. in Economics
2013 - 2016	WHU - Otto Beisheim School of Management
	B.Sc. in Business Administration