Education, Healthy Habits, and Inequality*

Jesús Bueren
EUI

Dante Amengual
CEMFI

Josep Pijoan-Mas
CEMFI and CEPR

January 2025

Abstract

Inequality in health outcomes has a strong socio-economic component. We argue that differences in health behaviors across education groups are key to understanding this fact. We start by estimating latent types of lifestyle and their impact on health dynamics by use of data on health behavior and health outcomes in the HRS and the PSID. We find that there is a large gradient in life expectancy across lifestyles (8 years at age 50) and that the higher frequency of health-protective lifestyles among the more educated individuals explains 40% of the education gradient in life expectancy. Next, to understand lifestyle formation and the correlation between education and health behavior choices, we build a life cycle model with idiosyncratic labor market and health risks. In the model, education and lifestyles are jointly chosen early in life by individuals who are heterogeneous in the utility cost of adopting protective lifestyles and acquiring education. Importantly, in our calibration these two early-life investments are complements, which endogeneously generates selection of heterogeneous individuals into each choice. Quantitatively, we find that the more educated individuals choose healthier lifestyles partly because of their income advantage, partly because of the higher yield of their health-protective behavior, and partly due to their better selection in terms of costs of adopting healthier behaviors. Finally, we find that the increase in the college wage premium over the last decades has widened the education gradient in lifestyles, resulting in one-year increase in the education gradient of life expectancy across cohorts born in the 1930s and 1970s. Of this increase, 40% is driven by the direct effect of wage changes, while 60% is due to changes in the composition of the shrinking set of high school dropouts.

*The authors thank the participants for their valuable feedback from presentations held at Bristol University, CEMFI, EUI, HEC Montreal, the 2024 BSE Summer Forum in Barcelona, and the 2024 EEA Annual Meetings in Rotterdam. Emails: amengual@cemfi.es, jesus.bueren@eui.eu, pijoan@cemfi.es. Josep Pijoan-Mas acknowledges financial support from Grant PID2022-142790NB-I00 funded by MICIU/AEI/ 10.13039/501100011033 and by ERDF/EU. A previous version of this work was circulated under the title "*Healthy Habits and Inequality*".

1 Introduction

The US economy has witnessed a sustained increase in income and health inequalities over recent decades. This trend is underpinned by two fundamental observations: a strong correlation between economic and health outcomes (Kitagawa and Hauser, 1973; Pijoan-Mas and Ríos-Rull, 2014; Chetty et al., 2016) and a widening educational gradient in health outcomes (Preston and Elo, 1995; Meara et al., 2008; Case and Deaton, 2015). However, the precise mechanisms linking economic status and health outcomes are not well understood. Our study seeks to address this gap by examining the role played by lifestyle factors and health behaviors.

Lifestyle factors and health behaviors —like exercise habits, dietary patterns, or smoking and drinking— are important determinants of health outcomes. McGinnis and Foege (1993) seminal paper estimates that around half of the deaths that occurred in the US in 1990 resulted from risk factors arising from bad health behaviors. Since then, many papers have shown that individuals who engage in healthier behaviors are more likely to experience positive health outcomes, while those with unhealthy habits are at greater risk of developing chronic diseases and experiencing premature mortality (Taylor et al., 2002; Li et al., 2018; Zaninotto et al., 2020). This literature has also consistently shown that individuals with higher levels of education tend to adopt healthier lifestyles (Lantz et al., 1998; Cutler and Lleras-Muney, 2010; Polvinen et al., 2013).

Our paper tackles three main questions. First, we want to measure the impact of different lifestyle habits —understood as the collection of different health-related behaviours— on health dynamics and health inequalities. Second, we seek to understand the joint determination of education and lifestyle habits early in life and why there is an education gradient in lifestyle habits. Finally, given these ingredients, we aim to quantify the effect of raising labor earnings inequality across education groups on the observed increase in the education gradients of lifestyles and health outcomes. In particular, we want to assess whether the observed increase in the education gradient of life expectancy is the result of the direct effect of diverging economic prospects across education groups or, rather, whether it is the result of changes in the composition of each population group.

To address the first question, we use data from the Health and Retirement Study (HRS) and the Panel Study of Income Dynamics (PSID). These data sets contain a rich array of health behavior indicators —including preventive tests, substance abuse, and exercise habits— that have not been systematically exploited before. However, one faces three challenges to use these data. First, the observed health behaviors are imperfectly correlated across individuals and over time, which suggests either a noisy measure of these variables or a non-deterministic relationship between different health behaviors and individual incentives. Second, estimating the effects of each health behavior on health dynamics is difficult, as different health behaviors are hard to isolate from each other and their health effects in the long run are difficult to identify. Finally, it is impractical to consider so many different variables in a dynamic model that tries to understand the determinants of health-related behavior. To address these issues, we make a contribution by developing a novel methodology to reduce the dimensionality of the data. In particular, we identify patterns in lifestyle behavior by assigning individuals to permanent types based on both their health behaviors and their health dynamics. Hence, these types summarize the propensity of individuals to engage in different health behaviors over time as well as the resulting health trajectories. As a result, our novel econometric procedure allows us to estimate the long-run effect of lifestyle choices on health outcomes, which is a key ingredient for our model of endogenous health dynamics.

We consider a parsimonious representation of health behavior in two distinct lifestyles, which we label *protective* and *detrimental*. This representation delivers four main results. First, lifestyles have a strong effect on health dynamics: at age 50 there is an 8.6 years life expectancy gap between individuals with *protective* and *detrimental* lifestyles. Second, there is a strong correlation between lifestyle habits and education, with harmful behaviors being more prevalent among the less educated. Indeed, this education gradient in health behaviors explains approximately 40% of the education gradient in life expectancy. Third, the life expectancy differences across lifestyles types are also large within education categories, and more so for college graduates than for high school dropouts. In particular, the life expectancy gap between individuals with *protective* and *detrimental* lifestyles is 6.6 years among high school dropouts and 9.6 years among college graduates. This uncovers an important complementarity between education and lifestyle choices, which we discuss below. Finally, and equally important, we find an increasing dispersion in lifestyles across education groups for individuals born in more recent cohorts, with the higher educated increasingly adopting *protective* behaviors, while the lower educated are progressively engaging in more *detrimental* ones.

To understand the joint determination of education and health behavior choices, our second question, we propose a heterogeneous agents model comprising two distinct stages. In the first stage, when they are young, individuals simultaneously choose their education and health-related lifestyle. In this stage, individuals exhibit heterogeneity in both their utility costs associated with education and their utility costs associated with engagement in protective health behaviors. The utility costs of education may reflect individual taste, family background, or distance to education centers, while the utility costs of adopting protective health behaviors may be shaped by parental habits, peer effects, or access to health-promoting amenities during childhood. These costs will drive selection of hetereogeneous individuals into education and lifestyle choices. Then, in the second stage, during the working/retirement phase, individuals make consumption and savings choices in a standard life-cycle model with idiosyncratic labor income and health risks, with outcomes conditioned on their specific education and lifestyle choices made in the initial stage.

Our economic model incorporates complementarities between education and lifestyle investments for two important reasons. First, and as it is standard in models of health investments, an extra year of life is more valuable with higher consumption possibilities. This means that the value of a health-protective lifestyle is larger for the more educated, and the value of education is larger for individuals expecting to live more years. Second, as implied by the results of our econometric model, the benefit in terms of better health transitions of investing in *protective* health behavior is larger for higher-educated individuals.[1] In addition, these two complementarities between health and education investments endogenously shape the selection of individuals with lower (higher) costs of health investments into higher (lower) education.[2]

We calibrate our model to accurately replicate the joint distribution of education and lifestyle choices for cohorts born in 1930 and 1970, as well as to match the median wealth accumulation over the life cycle by individuals in different education and health behavior types. In the model, cohorts differ in their education-specific average labor earnings over the life cycle. Notably, our model effectively captures the overall wealth distribution and can explain 50% of the observed increase in life-expectancy inequalities across education groups between the 1930 and 1970 cohorts.

Our calibrated model allows us to quantify the relative importance of the mechanisms driving the observed education gradient in lifestyle behaviors, which we do for the 1970 cohort. First, the income advantage of the more educated explains 37% of the observed education gradient in health behavior and 1.3 years of the education gradient in life expectancy. Second, the higher yield of health-protective lifestyles of the more educated has the same quantitative effect as the income advantage. And third, endogenous selection into college education (high school dropouts) of individuals with lower (higher) costs of investing in their health explains 42% of the observed education gradient in health behavior and 1.5 years of the education gradient in life expectancy.

Finally, we use our model to study the recent increase in the education gradient of life expectancy, which is our third question. In particular, we quantify the effect of the changes (across cohorts) in the income prospects of individuals in different education groups. When comparing individuals born into the 1930 and 1970 cohorts, we find that the direct effect of worse (better) economic prospects for the high school dropouts (college graduates) in the 1970 cohort generates a widening of the education gradient of health behaviors, which leads to an increase in the education gradient of life expectancy of 5.3 months. The increase in the college premium of labor earnings also induces a large increase in the share of college graduates and a decline in the share of high school dropouts. This large change in education choices across cohorts sets in motion a selection mechanism that explains 7.5 months of the increase in the education gradient of life expectancy, as the smaller group of high school dropouts in 1970 contains a larger fraction of individuals for whom it is more costly to invest in their health. All in all, the increase in income inequality across education groups explains an increase in the education gradient of life expectancy of 1.1 years (55%

[1] Our econometric model is silent about the reasons behind this complementarity. One reason could be the higher propensity of more educated individuals to react to the results of preventive health care tests. For instance, it is well known that more educated individuals comply better with prescribed therapy (Gold and McClung, 2006) or that they are more likely to use newer drugs (Lleras-Muney and Lichtenberg, 2005).

[2] In our model, we abstract from considering potentially *higher prices* of adopting healthy behavior, which could also induce an education gradient in health behavior. While this factor could play a role, it is important to note that harmful behaviors like smoking and drinking are expensive and, hence, go in the opposite direction. Furthermore, Allcott et al. (2019) shows that the better diet of individuals in higher socio-economic groups is not due to differences in the quality of food supply across neighborhoods or to differences in prices of different types of food. Rather, these authors connect most of the variation in food choices to different food demands across people in different socio-economic groups, which aligns well with the focus of our paper.

of the observed increase) through the widening of the gradient in lifestyles. This result connects to the *deaths of despair* literature. Case and Deaton (2017) argue that the worsening of labor market opportunities for white males without a high school degree has led to an increase in risky health behavior (in particular, the use of opioids) for this population group. This has damaged the life expectancy of the less educated and widened the education gradient in life expectancy. Our results broaden the scope of changes in health-related behavior beyond substance abuse and provide a quantitative exercise for this type of argument. However, instead of looking at changes in behavior during the life cycle, we look at the early life determinants of lifestyle choices, which arguably are more important for comparisons across cohorts. Critically, our model finds a stronger impact from selection mechanisms (the composition of the shrinking set of high school dropouts) than from the direct effect of changes in labor market opportunities. This is consistent with the recent empirical findings of Novosad et al. (2022), who argue that the worsening of the health outcomes of non-college-educated males is partly due to selection. Hence, regarding the policy options to reduce health inequalities, our results suggest that early childhood interventions to foster the adoption of better lifestyles (like healthy food at school or exercise habits) may be more effective than labor market interventions to reduce wage inequality.

We note that our benchmark model assumes that individuals have perfect information about the average health consequences of their lifestyle choices. However, evidence suggests that earlier cohorts —particularly those born in the first half of the 20th century— may have had less accurate knowledge about the long-term consequences of health-related behaviors like smoking. To address this issue, we extend our model to account for imperfect information of earlier cohorts. We find that the improvement of information across cohorts is itself an engine of change in the educational gradient of life expectancy. In particular, with more informed choices, the complementarities between education and lifestyles increase, which widens the educational gradient in lifestyles and life expectancy. Furthermore, the role of selection becomes more prominent (due to the increase in the perceived complementarities), while the role of income inequalities decreases.

1.1 Related literature

Our estimation of latent types in health behavior relates to the very recent literature estimating unobserved fixed effects on health dynamics. De Nardi et al. (2024) were the first to show that health dynamics are not Markovian and that unobserved fixed effects are needed to explain long-run health dynamics. Contemporaneous work by Borella et al. (2024) and Hong et al. (2024) estimate unobserved fixed effects by k-means clustering of health trajectories. In contrast, our study classifies individuals into types using both health behavior data and observed health trajectories. By incorporating health behavior data, our approach offers a clearer interpretation of these fixed effects, helping to separate chance events from more persistent health traits. Our econometric framework allows us to deal with the selection of survivors and, hence, exploit longer trajectories of health data for clustering, which in turn helps disentangle persistent shocks from fixed effects.

Finally, our economic model incorporates type formation, which allows us to study the correlation between the unobserved fixed effects and socio-economic variables.

The second stage of our model builds on a mature literature that employs life cycle models to quantify how heterogeneous health dynamics impact economic outcomes —see for instance De Nardi et al. (2010), French and Jones (2011), De Nardi et al. (2016), Ameriks et al. (2020), Nakajima and Telyukova (2023), or Bueren (2023)— and welfare —see Capatina (2015), Braun et al. (2019), Hosseini et al. (2021), and De Nardi et al. (2024). Importantly, the latter group of papers estimates significant welfare losses associated with adverse health episodes. The differential welfare losses of bad health across education groups are a key ingredient in our paper, as they shape the different incentives to invest in good health by individuals making different education choices.

Finally, the paper is also related to previous work featuring endogenous health dynamics. Fonseca et al. (2021) and Hong et al. (2024) consider monetary health investments, but they struggle to find a strong causal effect of medical spending on health outcomes. The same can be said of quasi-experimental evidence, like Aron-Dine et al. (2013), Finkelstein et al. (2012), and Baicker et al. (2013). Ozkan (2023) does find a stronger role for medical spending, but this is mostly due to the preventive health care component. Furthermore, the socioeconomic gradient in health outcomes is also big in the UK or Germany, countries characterized by free universal health care and low out-of-pocket medical spending, see Boháček et al. (2021) or Mahler and Yum (2022). This justifies our focus on health behavior choices as a potential driver of the educational gradient of health outcomes. Health behavior choices have been previously studied by Cole et al. (2019), Mahler and Yum (2022), Margaris and Wallenius (2023), and Bairoliya et al. (2024). These models allow agents to adjust their behavior over the life-cycle, taking education choices as given. They also assume equal or exogenously different utility costs of good health behavior across individuals in different education groups. Instead, our model endogenously allows for the sorting of individuals with ex-ante different utility costs of health behavior into different education groups and capitalizes on the persistence of health behavior, which is often established early in life, to model health behavior and education choices jointly. These features allow us to study type formation and to assess the extent to which selection drives differences in health outcomes across education groups. In this sense, our approach is more similar to Hai and Heckman (2022), who study the interaction between smoking, preferences, and education.

The remainder of the paper is organized as follows: Section 2 briefly summarizes the patterns in our data of health behavior. Section 3 and Section 4 present the econometric model used to identify lifestyles and its results, respectively. Section 5 outlines the economic model, while Section 6 details the calibration strategy. Section 7 presents the quantitative findings, followed by concluding remarks in Section 8.

Table 1: Mean health behavior and 4-year auto-correlation

	Mean					AC	
	HSD	HSG	CG	50-60	70-80	50-60	70-80
Drinking	0.07	0.08	0.07	0.13	0.05	0.48	0.42
Smoking	0.18	0.15	0.07	0.22	0.09	0.75	0.64
Cancer test	0.60	0.70	0.79	0.69	0.74	0.40	0.42
Cholesterol	0.74	0.81	0.86	0.75	0.85	0.33	0.29
Flu shot	0.57	0.60	0.66	0.46	0.74	0.55	0.58
Exercise	0.27	0.38	0.54	0.43	0.39	0.40	0.38

Notes: Data from the HRS. HSD: high school dropout; HSG: high school graduate; CG: college graduate; 50-60: sub-sample of individuals aged 50 to 60; 70-80: sub-sample of individuals aged 70 to 80. The last two columns show the autocorrelation (AC) of each health behavior with a 4-year lag.

2 Data on health behavior

We use data from the PSID (1999 to 2019) and the HRS (1996 and 2018). Our data on health behavior collects binary information on whether individuals engage in heavy drinking (having more than 2 alcoholic beverages on the day that the respondent drinks), whether they smoke, whether they have taken a preventive cancer test (males: prostate cancer screening; females: mammography), cholesterol test, and flu shot in the last year, and whether they have an exercise habit (some time participating in sports or other exercise activity during the week before being interviewed). All six variables are available in the HRS, but only the smoking and drinking variables are available in the PSID.

These data show three important patterns. First, there is a clear education gradient on all measures of behavior, whereby the more educated groups contain a higher fraction of individuals who engage in healthier behaviors (see Columns 1 to 3 in Table 1). This gradient suggests that health behaviors may be an important factor behind the education gradient of life expectancy. Second, there is persistence over time in health behaviors, but, except for smoking, this persistence is not high. For instance, among individuals between 50 and 60 years of age, the 4-year autocorrelations of heavy drinking or a flu shot are 0.48 and 0.55, respectively (see Columns 6 and 7 in Table 1). This weak persistence is partly the result of a life-cycle pattern, as the incidence of healthier behavior increases with age (except for exercise habits, which decline). For instance, the incidence of smoking falls from 0.22 of the population aged 50 to 60 to 0.09 of the population aged 70 to 80, while the incidence of flu shots rises from 0.46 to 0.74 (see Columns 4 and 5 in Table 1). And third, these different measures of health behavior are correlated across individuals as one would expect, but the correlations are small (see Table 2). For instance, the correlations across individuals aged 50 to 60 of cholesterol tests with flu shots or cancer tests are fairly large, 0.24 and 0.40, respectively, while the correlations of cholesterol tests with drinking, smoking, and exercise habits are much smaller, -0.06, -0.13, and 0.04, respectively.[3]

[3] The imperfect correlation over time and in the cross-section, facts two and three, have already been highlighted for healthy diet and exercise for a sample of nurses in the US, see Bairoliya et al. (2024).

Table 2: Cross correlation health behaviors

	Drinking	Smoking	Cancer test	Cholesterol	Flu shot	Exercise
Drinking	1.00	0.09	-0.00	-0.01	-0.03	0.02
Smoking	0.16	1.00	-0.09	-0.09	-0.08	-0.07
Cancer test	-0.08	-0.15	1.00	0.30	0.20	0.09
Cholesterol	-0.06	-0.13	0.40	1.00	0.23	0.06
Flu shot	-0.05	-0.07	0.19	0.24	1.00	0.02
Exercise	-0.00	-0.13	0.08	0.04	0.02	1.00

Notes: Data from the HRS. Upper diagonal: individuals aged between 70 and 80. Lower diagonal: individuals aged between 50 and 60.

The positive but low correlation of different good health behaviors over time and across individuals suggests the need for a latent factor model such that individuals of different types have a different propensity to engage in certain behaviors. Ideally, this propensity should change with age (as the incidence of behavior does) and with health (to allow for two-way relations between health and behavior). Ultimately, linking the types to observed health dynamics would serve to estimate the health consequences of behavior types, and to use this as a criterion to form the types.[4] We present such a factor model in the next Section.

3 An econometric model of health dynamics with latent types

We combine our data from the HRS and the PSID to create an unbalanced panel of individuals $i = 1, \ldots, N$ followed for $t = 0, \ldots, T_i$ periods. For each individual and period, we observe standard demographic variables: cohort of birth $\mathrm{c}_i \in \{c_{10}, c_{30}, c_{50}, c_{70}, c_{90}\}$ (individuals born between 1900 and 1920, 1920 and 1940, *etc.*), gender $\mathrm{s}_i \in \{s_m, s_f\}$ (male, female), education $\mathrm{e}_i \in \{\text{HSD}, \text{HSG}, \text{CG}\}$ (high school dropout, high school degree, college degree), and age at first interview a_{i0}, plus a wide array of health-related variables, which we classify into two groups: health outcomes and health behavior. The health outcome $h_{it} \in H \equiv \{h_g, h_b, h_d\}$ takes three values: good health (h_g), bad health (h_b), or dead (h_d), which is an absorbing state. We build this variable using the 5-category self-rated health variable (where the best three categories form the good health state) and the information on survival. The health behavior vector $\mathbf{z}_{it} = \{z_{1,it}, z_{2,it}, \ldots, z_{N_z,it}\}$ contains information on N_z different categorical variables $z_{m,it} \in \{0, 1\}$ describing whether individual i in period t does some particular action. These actions are the health behavior variables described in Section 2.[5]

[4]The estimated relationship between types and health dynamics is different from estimating the causal effect of each behavior on health transitions, which is difficult for two reasons. Firstly, a positive but low correlation of different health-enhancing behaviors over time and across individuals complicates the isolation of singular behavior effects. Secondly, the impact of individual health behaviors on health outcomes may not manifest in observable health changes within a short period of time but may appear many years ahead. Our methodology shifts the focus from isolated behaviors to the holistic lifestyle patterns that shape long-run health trajectories.

[5]Some of the health outcome and health behavior variables for a given individual may be missing for some period t. Indeed, in the PSID we do not observe the variables for health *protective* behavior (cancer test, cholesterol test, flu shot). We take missing observations into account under the assumption that they occur completely at random, but we abstract from them in the model description to simplify the exposition.

We next assume that both the observed health behavior $\mathbf{z}_{it}$ and health outcome h_{it} depend on some unobserved time-invariant latent variable $\mathrm{y}_i \in Y \equiv \{y_1, y_2, \ldots, y_{N_y}\}$. We interpret the latent variable as lifestyle / health habit type —determined before the start of working life— that captures the idea that individuals differ in their propensity to undertake actions that are good for their health.

We model the joint sequence of health behaviors $\mathbf{z}_i^T$ and health outcomes $\mathbf{h}_i^T$ of each individual i conditional on the demographic variables and initial health by means of a mixture. That is, both observed health behavior and observed health transitions help inform the classification of individuals into types.[6] We can write the likelihood of the data as a mixture model:

$$\begin{aligned} p\left(\mathbf{z}_i^T, \mathbf{h}_i^T | \mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, a_{i0}, h_{i0}; \boldsymbol{\theta}\right) &= \sum_{y=1}^{N_y} p\left(\mathbf{z}_i^T | y, a_{i0}, \mathbf{h}_i^T; \boldsymbol{\theta}_y\right) \times p\left(\mathbf{h}_i^T | \mathrm{s}_i, \mathrm{e}_i, y, a_{i0}, h_{i0}; \boldsymbol{\theta}_y\right) \\ &\times \quad p\left(y | \mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, a_{i0}, h_{i0}; \boldsymbol{\theta}\right) \end{aligned} \tag{1}$$

where $\boldsymbol{\theta}$ is the vector of parameters and $\boldsymbol{\theta}_y$ is the subset of parameters relevant to type y. The three elements on the right-hand side of equation (1) are the probability of observing a sequence of health behaviors, the probability of observing a sequence of health outcomes, and the initial distribution of types. We explain them in more detail in Section 3.1, 3.2, and 3.3, respectively.

3.1 Health behavior

We assume that the probability of individual i in period t reporting the m^{th} behavior ($h_{m,it} = 1$) depends on the health behavior type y_i but also on age a_{it} and current health status h_{it}.[7] The idea is that the association of observed behavior, such as smoking or cancer tests, with types may differ over age and across health states. Instead, conditional on these variables, we impose that health behavior does not depend on cohort c_i, gender s_i, or education e_i. We do so because we want the definition of types to be stable across demographic groups, hence letting the variation in behavior across demographic groups arise from the different distribution of types.

Next, we assume that conditional on $\mathrm{y}_i = y$, h_{it}, and a_{it}, health behaviors are independent between them and over time so that we can model the probability of observing a history of health

[6] As an alternative, one might think of only using health behavior data to classify individuals into types, and estimate health dynamics conditional on type in a second stage. We do not do so for two reasons. First, we want the classification of individuals based on behavior to be meaningful in terms of health dynamics, that is, we want to avoid grouping over behaviors with very different incidence on health dynamics. Second, the joint use of health behavior and health dynamics data allows us to estimate the third term in equation (1), which controls the selection of better types into survival. That is, as we show in Section 3.3, this term absorbs the different frequencies of good and bad types over age due to differential mortality.

[7] Note that given a_{i0} and t, $a_{it} = a_{i0} + 2t$, as the average time between interviews in the HRS and PSID is 2 years.

behaviors as follows:

$$p\left(\mathbf{z}_i^T|y, a_{i0}, \mathbf{h}_i^T; \boldsymbol{\theta}_y\right) = \prod_{t=0}^{T}\prod_{m=1}^{N_z} p(z_{m,it}|y, a_{it}, h_{it}; \boldsymbol{\theta}_y) \tag{2}$$

where each element $p(z_{m,it}|y, a_{it}, h_{it}; \boldsymbol{\theta}_y)$ is modeled as a probit, see Appendix A.1 for details.

3.2 Health dynamics

We assume that health dynamics for individual i depends on gender s_i, education e_i, health behavior type y_i, and age a_{it}. The dependence of health dynamics on gender and education is meant to capture differences in health outcomes associated with these variables that are not captured by differences in health behavior types across these demographic groups. The absence of cohort c_i from the set of conditioning variables is an identification assumption. Since we do not observe full lifespans for individuals in different cohorts, we combine information on health dynamics at old ages from individuals born in earlier cohorts with information on health dynamics at young ages from individuals born in later cohorts. This is standard in the estimation of models of health dynamics with survey data, see, for instance, Pijoan-Mas and Ríos-Rull (2014). Hence, our assumption is that the health dynamics of individuals of a given gender and education are, conditional on type, identical across cohorts. Nevertheless, it is important to note that this allows gender and education health dynamics to differ across cohorts due to the different composition of types.

We next assume that conditional on gender s_i, education e_i, health behavior type $\mathrm{y}_i = y$, and age a_{it}, the evolution of health outcomes is markovian, that is, it only depends on one lag of health outcomes:

$$p\left(\mathbf{h}_i^T|\mathrm{s}_i, \mathrm{e}_i, y, a_{i0}, h_{i0}; \boldsymbol{\theta}_y\right) = \prod_{t=1}^{T} p(h_{it}|\mathrm{s}_i, \mathrm{e}_i, y, a_{i,t-1}, h_{i,t-1}; \boldsymbol{\theta}_y) \tag{3}$$

The specific functional form for $p(h_{it}|\mathrm{s}_i, \mathrm{e}_i, y, a_{i,t-1}, h_{i,t-1}; \boldsymbol{\theta}_y)$ is explained in Appendix A.2.

3.3 Distribution of health types

The final element that we need in order to evaluate the likelihood function is the fraction of individuals of each type y, that is, the mixing weights. In particular, we define $p\left(\mathrm{y}|\mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, a_{i0}, h_{i0}; \boldsymbol{\theta}\right)$ as the probability that individual i born into cohort c_i, of gender s_i, education e_i, first observed at age a_{i0} and health h_{i0} is of type y. In stationary mixture models, it is common to add this term to the likelihood function and estimate it non-parametrically from the data. However, in our model with health dynamics that are heterogenous by lifestyle types y, the mixing weights need to be different across initial age a_{i0} and health h_{i0}, as well as across $(\mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i)$ groups. To avoid estimating different mixing weights for each potential different entry age a_{i0}, we exploit the model of health dynamics in equation (3) to obtain a recursive expression describing the distribution of types conditional on observables at each age. This leaves us needing to estimate only the initial distribution of types at an arbitrary initial age, which we choose to be 25. That is, we only need to estimate the fraction $p\left(\mathrm{y}|\mathrm{c},\mathrm{s},\mathrm{e},25,h;\boldsymbol{\theta}_y\right)$ of individuals of age 25, cohort c, gender s, and education e, in health state h, who are of type y. The specific functional form for this probability (multinomial probit), as well as the formal description of the above argument, are explained in Appendix A.3.

3.4 Estimation

Given the need to estimate a large number of parameters and latent variables –one for each individual– we resort to Bayesian methods, which are particularly well-suited for models with latent structures. Furthermore, we implement a Gibbs sampler to break down the high-dimensional likelihood into smaller, simpler conditional distributions, which significantly reduces the computational complexity of the estimation process, see Appendix A.4 for details.

4 Results from the econometric model

We estimate the model using our panel data discussed in Section 2. In favor of parsimony, we choose $N_y = 2$ for our main estimation and discuss the cases with $N_y > 2$ in Appendix B.

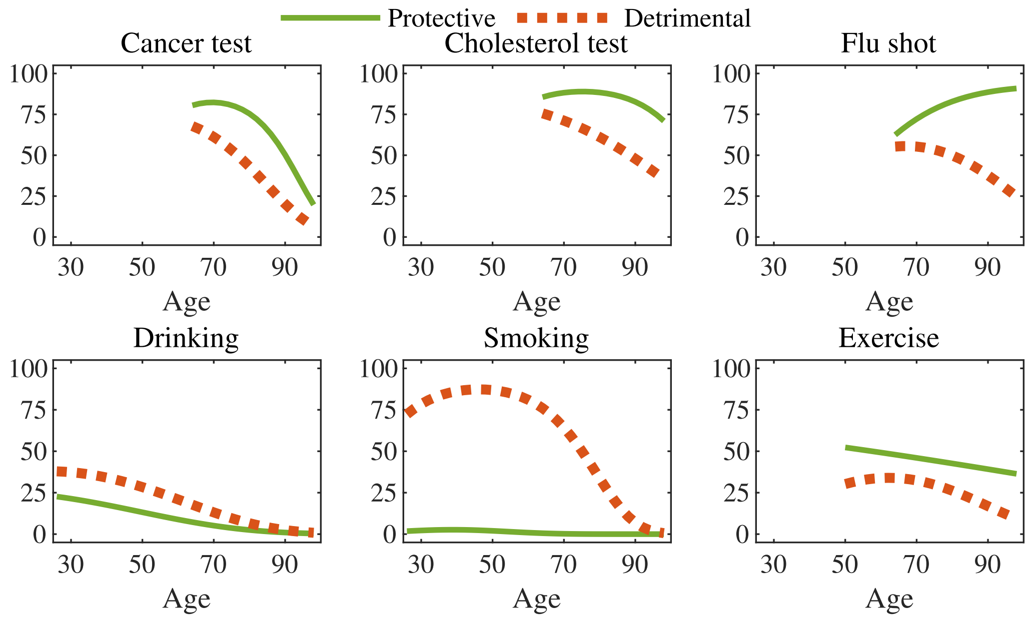
4.1 Health behavior

We start by showing how the estimated types are related to observed health behaviors. Figure 1 reports the probability of displaying each health behavior $z_{m,it}$ as a function of health type y_i, age a_{it}, and $h_{it} = h_g$ (the case $h_{it} = h_b$ is not too different). Individuals in one group, which we label *protective* (solid green line), have a higher likelihood of reporting health enhancing behaviors (cancer test, cholesterol test, flu shot, and exercise) and a lower probability of reporting harmful health behaviors (smoking and drinking). In turn, for individuals in the other group, which we label *detrimental* (red dashed line), the probability of reporting all the health enhancing behavior is lower and their smoking probability is very high. We note that, within each type, the probability of displaying each health behavior changes with age and it does so differently across types. For instance, the probability of taking a flu shot is similar across the two groups at age 50 but, as they age, individuals classified as *protective* increase their probability of getting a flu shot while individuals classified as *detrimental* do not.

4.2 Health dynamics

Next, we examine the relationship between health behavior types and health dynamics. We summarize this information by looking at the gradient in age-50 life expectancy across health behavior

Figure 1: Health habits and health behavior types



Notes: Estimation results. Probability of engaging in each health behavior by age and type, for male individuals in good health.

types.[8] In Table 3 (second column) we show that, on average, *protective* types live for 8.6 more years than *detrimental* types. The effect of lifestyle on health dynamics is also large within each education category: *protective* types live 6.6 more years than *detrimental* types among males without a high school degree, 6.8 more years among males with a high school degree, and 9.6 more years among males with a college degree (see columns 4, 6, and 8). Finally, it is important to note that the gradient is increasing in education, that is, the health yield of a *protective* lifestyle is larger for the more educated.

4.3 Distribution of health types

Our estimation assigns a different fraction of individuals to each type y depending on cohort c, gender s, and education e. For the 1970 cohort, 74.4% of individuals are classified as *protective* and 25.6% as *detrimental* (see first column in Table 3). Our main finding is that there is a large educational gradient of health behavior types: the share of *protective* individuals grows from 44.3% to 93.6% as we move from high school dropout to college graduate (see columns 3, 5, and 7), which reflects a strong correlation between education and lifestyle types. In a sense, this correlation is not too surprising: it is well known that the incidence of smoking, drinking, and obesity declines with education, and Table 1 before shows how in our data the share of individuals taking cancer tests, cholesterol tests, and flu shots increases with education. It is therefore natural that our classification of individuals into types according to the observed health-related behavior retains the education

[8] For more details, Figure E.1 in the Appendix shows the actual health transitions and survival functions by health behavior type.

Table 3: Life expectancy at age 50 across education and lifestyles: males born in 1970s

	All		HSD		HSG		CG		Δ_eLE (CG-HSD)		
	%	LE	%	LE	%	LE	%	LE	Data	(a)	(b)
All	100.0	29.3	100.0	24.7	100	28.1	100.0	32.8	8.1	4.5	3.6
PRO	74.9	31.6	46.1	28.8	69.7	30.3	93.4	33.4	4.6		
DET	25.1	22.6	53.9	21.2	30.3	23.0	6.6	24.5	3.4		
Δ_y	49.9	9.0	-7.7	7.7	39.3	7.3	86.9	8.9	1.2		

Notes: This table reports the share of male individuals of each lifestyle and the life expectancy (LE) at age 50 for different population groups. Column (a) corresponds to the counterfactual LE gradient when the distribution of behavior types among high school dropouts is the same as for college-educated individuals. Column (b) is the difference between the actual LE gradient and the counterfactual in column (a), that is, it corresponds to the gradient explained by difference in lifestyles across education groups for given health dynamics.

gradient of these latter variables. Nevertheless, our classification of individuals into types also uses longitudinal information on health dynamics *conditional on education*, that is, it also uses the fact that within education, *protective* types have better health dynamics than *detrimental* types.

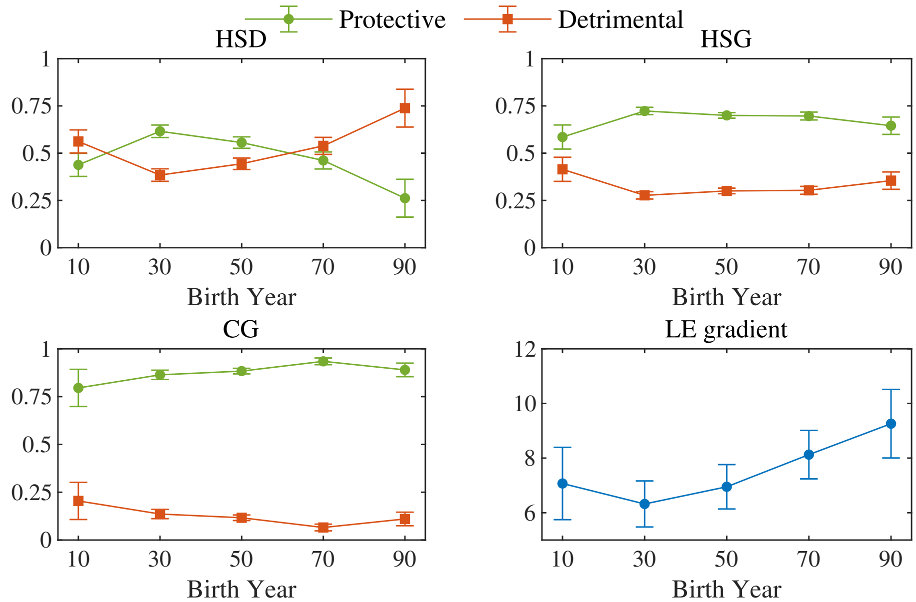
Next, we look at the evolution over time of the joint distribution of health types and education. In Panels (a) to (c) of Figure 2 we report the distribution of types by education group from the 1910 to 1990 cohorts. Within the high school dropouts, the *detrimental* types increased monotonically from 40% in the 1930 cohort to 75% in the 1990 cohort. This implies a severe deterioration in the lifestyle of individuals in the least educated group, which reverses a slight improvement in the type distribution between the 1910 and 1930 cohorts. In contrast, among college-educated individuals, there is a smaller change, with a slight increase in the share of *protective* and a slight decline of the *detrimental*. All in all, this implies that the educational gradient in lifestyles has widened remarkably between the 1930 and the 1990 cohort. As we will see in the next Section, this will generate an increasing life expectancy gap across education groups.

4.4 Decompositions

Combining all previous results, we can show that the different type composition across education groups explains an important fraction of the educational gradient of life expectancy, which is 7.8 years for the 1970 cohort. In particular, we compute counterfactual health dynamics for males without a high school degree for the case in which they have a distribution of health behavior types as the males with a college degree. In this counterfactual case, we find that their life expectancy would rise by 3.2 years, reducing the education gradient of life expectancy from 7.9 to 4.7 years. Hence, the different distribution of health behavior types across education groups explains around 40% of the gradient. The rest, 4.7 years, is the result of different health dynamics across education groups for fixed distribution of lifestyle types (see the last two columns in Table 3).

Finally, in Panel (d) of Figure 2 we report the age-50 predicted education gradient in life expectancy for different cohorts. In our estimation, different cohorts have different life expectancies

Figure 2: Distribution of types by education and cohort (males)



Notes: The first three panels display the fraction of individuals of each behavior type in each different cohort for a given education group. The bottom-right panel reports the estimated life expectancy difference between college graduates and high school dropouts in each cohort.

because of different compositions of latent types as described in Panels (a) to (c) of Figure 2, but health dynamics conditional on type are identical across cohorts. This allows us to infer health dynamics at old ages of younger cohorts, for which most individuals are still alive today. Our findings show a growing education gradient in life expectancy: from 6.8 years in the 1930 cohort to 9.1 years in the 1990 cohort, which follows a 0.7 year decline in the gradient between the 1910 and the 1930 cohorts.

5 An economic model

Our economic model considers two distinct life stages. The *early life* stage is a static problem where young individuals belonging to a cohort c choose their education $e \in \{\text{HSD}, \text{HSG}, \text{CG}\}$ and their lifestyle or health-related behavior $y \in \{\text{DET}, \text{PRO}\}$. This problem is a stand-in for choices and investments made by either parent in early childhood or young adults before entering the labor market. The objective is to maximize the expected value of starting working life with a given type (education and lifestyle) minus the (idiosyncratic) costs of choosing each type. This stage serves to account for the observed correlation between education and health types. The *adult life* stage is a dynamic life-cycle consumption-saving problem under uncertainty in both labor market and health outcomes where individuals differ in fixed characteristics (education and lifestyle). This stage serves to link inequality in health and economic outcomes, and provides the value of starting life in each type, which is used in the *early life* stage.

5.1 Stage 1: early life

Let $V_0^{\mathrm{c,e,y}}$ be the value of starting working life with a type (e,y) for an individual in cohort c. Before entering the labor market, young individuals choose their type by solving

$$\max_{\mathrm{e,y}} \left\{ V_0^{\mathrm{c,e,y}} - \tau_{\mathrm{e}} - \tau_{\mathrm{y}} \right\} \tag{4}$$

where τ_{e} and τ_{y} represent the cost of choosing an education e and a lifestyle y, respectively. These education and health behavior costs are heterogeneous in the population.[9] This formulation imposes that education and healthy habit choices are taken together at a young age and never change. We want to comment on this assumption. First, there is empirical evidence on the early-age adoption of health-related habits.[10] Second, these habits tend to be very persistent, as exemplified by the small effects of interventions to change them during adulthood.[11] Third, we note that, according to our econometric specification in Section 3, a permanent choice of a latent lifestyle type y still allows for changes in observed behavior (like smoking, exercise, or preventive tests) over the life-cycle and across health states, see for instance Figure 1. And fourth, if lifestyles are very persistent over the life cycle, understanding their initial formation seems more relevant than understanding their changes when the research question is the study of differences in lifestyles across education groups or across cohorts.

Going into details, we normalize to 0 the cost of a *detrimental* lifestyle and the cost of not finishing high school ($\tau_{\mathrm{DET}} = 0$ and $\tau_{\mathrm{HSD}} = 0$). Next, we assume that the cost of *protective* behavior τ_{PRO} is heterogeneous in the population and described by some CDF, $F(\tau_{\mathrm{PRO}})$, while the costs of graduating from college and high school $(\tau_{\mathrm{CG}}, \tau_{\mathrm{HSG}})$ are also heterogeneous in the population and described by another CDF, $F(\tau_{\mathrm{CG}}, \tau_{\mathrm{HSG}})$. Throughout the paper, we impose that these two distributions are independent from each other. Hence, the selection of individuals with different τ_{PRO} into each education category (and the selection of individuals with different τ_{CG} into each lifestyle category) will arise endogeneously through our model mechanisms.

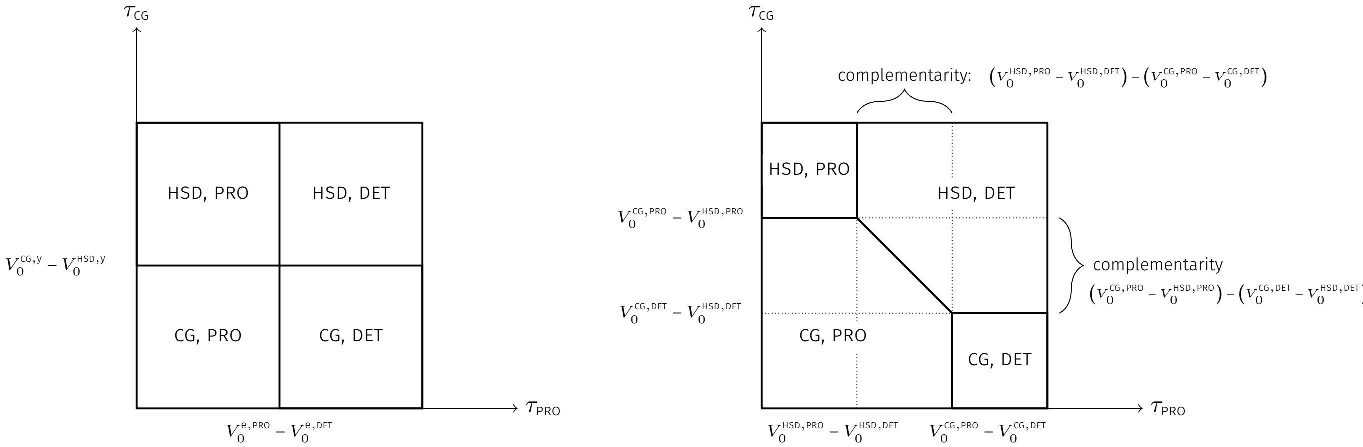
In this framework, the choices of education and lifestyle are independent of each other if and only if $V_0^{\mathrm{CG,PRO}} - V_0^{\mathrm{CG,DET}} = V_0^{\mathrm{HSG,PRO}} - V_0^{\mathrm{HSG,DET}} = V_0^{\mathrm{HSD,PRO}} - V_0^{\mathrm{HSD,DET}}$; that is, if the value of pursuing a *protective* lifestyle is the same no matter what is the chosen education level (or, likewise, if the value of choosing higher education is the same regardless the lifestyle choice). Such a situation is illustrated in Panel (A) of Figure 3 for the particular case of only two education

[9] The heterogeneous costs of education τ_e may arise due to many different factors, such as differences in family background (Hauser and Featherman, 1976), distance to quality education centers (Card, 1995), or taste (Willis and Rosen, 1979). However, because τ_e does not directly affect earnings, we should not think of it as labor market ability in the manner of Keane and Wolpin (1997). Less is known about the heterogeneous utility costs of healthy habits choices. A recent literature relates differences in brain characteristics of young adolescents with later-in-life health risk behavior (Xiang et al., 2023), which highlights the importance of individual-level variation in τ_y.

[10] For instance, see Farrell and Fuchs (1982) or Hai and Heckman (2022) for the early adoption of smoking and how it correlates with later decisions of college education.

[11] See Conner and Norman (2017) and references therein.

Figure 3: Early life choices of education and lifestyle



(A) Independent choices

(B) Complementarity

Notes. Eary life choices of education and lifestyle for the case with only two education categories ($\tau_{\text{HSG}} \to \infty$). Panel (A) displays the choices for the case in which $V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} = V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$, while Panel (B) displays the choices for the case $V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} > V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$. See Appendix C for details.

categories (we drop out e = HSG).[12] In this situation, an individual chooses a *protective* lifestyle if and only if $\tau_{\text{PRO}} < V_0^{\text{e,PRO}} - V_0^{\text{e,DET}}$ and a *detrimental* lifestyle otherwise. Here e stands for both education categories as the differences in values is the same. Likewise, an individual chooses a college education if and only if $\tau_{\text{CG}} < V_0^{\text{CG,y}} - V_0^{\text{HSD,y}}$ and drops out of high school otherwise.

In contrast, whenever $V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} > V_0^{\text{HSG,PRO}} - V_0^{\text{HSG,DET}} > V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$ there are complementarities between education and lifestyle choices; that is, the value of a *protective* lifestyle is larger for the more educated. We plot this situation in Panel (B) of Figure 3 (again, for the particular case of only two education categories). In this case, the value of pursuing each lifestyle depends on the chosen education (and the value of pursuing a particular education depends on the chosen lifestyle). In particular, individuals with an intermediate cost of *protective* behavior (τ_{PRO}) choose a *protective* lifestyle and a college education when the cost of studying (τ_{CG}) is low. However, when the cost of studying is high, they choose a lower level of education (e = HSD) and they may also switch to *detrimental* behavior, as the value of a *protective* lifestyle is lower when not going to college.

Finally, an important consequence of the complementarity of education and lifestyle choices shown in Panel (B) of Figure 3 is that it generates a distinct pattern of selection of individuals into education groups and lifestyle types. In particular, for the case with two education categories, it can be shown that the distribution of τ_{PRO} conditional on e = HSD first-order stochastically dominates the distribution of τ_{PRO} conditional on e = CG. That is, $\Pr[\tau_{\text{PRO}} \leq \tau | \text{HSD}] \leq \Pr[\tau_{\text{PRO}} \leq \tau | \text{CG}] \;\; \forall \tau$,

[12] See Appendix C for details on the construction of both panels in Figure 3.

with strict inequalities for some τ, see Proposition 1 in Appendix C. In words, when there is complementarity of education and lifestyle choices, individuals with a college education will on average be more likely to choose a *protective* lifestyle than individuals without a high school degree because of the self-selection of low (high) τ_{PRO} individuals into college (high school dropout) education.

5.2 Stage 2: adult life

In this Section, we model adult life, which generates the value functions $V_0^{\mathrm{c,e,y}}$ used in the early life stage.

5.2.1 Demographics, preferences, and shocks

The model period corresponds to two years. Individuals live for at most T periods but survival is stochastic every period. During the first $R-1$ periods of life people are exposed to health shocks, medical expenditure shocks, and labor income shocks. Individuals retire at age R, when they start receiving a retirement pension instead of stochastic labor income.

Preferences over consumption flows c_t are described by a standard CRRA period utility:

$$u(c_t) = \frac{(c_t/\bar{n}_t)^{1-\sigma} - 1}{1-\sigma} + \bar{b},$$

where σ is the risk-aversion parameter, $\bar{n}_t$ is an age-specific household size, and $\bar{b}$ is a positive term to ensure that individuals in our model value their life. In the period when they die, individuals also derive utility from leaving a bequest of size k_t:

$$v(k_t) = b_0 \frac{(k_t + b_1)^{1-\sigma} - 1}{1-\sigma}$$

where b_0 drives the strength of the bequest motive and b_1 the extent to which preferences for bequest increase with wealth.

Following the empirical model in Section 3, health h_t can be either good (h_g) or bad (h_b) and it evolves according to the age-dependent Markov chain $\Gamma_t^{\mathrm{e,y}}(h_t)$, which depends on age t, education e, and health type behavior y. The survival probability $s_t^{\mathrm{e,y}}(h_t)$ depends on age t and health h_t, but also (possibly) on education e and health-behavior type y.

Every period of their working life, individuals receive an exogenous income, which we model in two components. First, there is an employment shock $\ell_t \in \{0, 1\}$ that determines if the individual has the chance of working in the labor market ($\ell_t = 1$) or not ($\ell_t = 0$). We model $\mathrm{Prob}(\ell_t = 1|\mathrm{c,e},t,h_t)$ as an *i.i.d* shock conditional con cohort, education, age, and health. This component aims to capture the higher probability of individuals in bad health being non-employed, which is at the core of the health gradient of labor income, and also differences in employment across education and cohorts. Second, conditional on working, individuals receive labor income given by,

$$\log w_t^{\mathrm{c,e}}(h_t, \xi_t, \epsilon_t) = \log \omega_t^{\mathrm{c,e}}(h_t) + \xi_t + \epsilon_t, \tag{5}$$

where $\omega_t^{\mathrm{c,e}}(h)$ is a deterministic component depending on cohort, education, age, and health, while ξ_t and ϵ_t are persistent and transitory shocks, respectively. The initial value of the persistent component ξ_0 is drawn from a normal distribution with mean zero and variance $\sigma^2_{\xi_0}$. The stochastic persistent component is assumed to follow a Gaussian AR(1) process with persistence ρ_ξ and variance of the innovations σ^2_ξ. The transitory component ϵ_t is Gaussian white noise with mean 0 and variance σ^2_ϵ.

Finally, medical expenses are given by,

$$\log m_t^{\mathrm{e}}(h_t, \zeta_t) = \mu_t^{\mathrm{e}}(h_t) + \zeta_t$$

where $\mu_t^{\mathrm{e}}(h_t)$ is a deterministic component, as a function of education e, age t, and health h_t, while ζ_t is an i.i.d gaussian white noise process with mean zero and variance $\sigma^e_{\zeta,t}(h)$.

5.2.2 Taxation and social transfers

We model the tax system as follows. Working households pay payroll taxes, which include the Medicare tax (τ_{MCR}) and the Social Security tax (τ_{ss}), with the latter only affecting earnings below w_{ss}. Following Heathcote et al. (2020), we define a progressive labor income tax function $T(w) = w - a_{\tau 0} w^{1-a_{\tau 1}}$. We represent several existing means-tested programs in a stylized way through a public safety-net program guaranteeing every household a minimum income floor $\underline{x}$. Retirees receive Social Security benefits. In practice, these payments depend on an individual's history of earnings. To capture the existing variation in pension benefits without increasing computational costs, we approximate the benefits using the following approach. First, we divide individuals into groups based on their labor force participation just before retirement, their last draw of the persistent productivity shock, and their education and health behavior type. Then, for each group, we compute average earnings over the 17 model periods (34 years) with the highest earnings, and we apply the Social Security benefits formula to these average earnings.

5.2.3 The optimization problem

At the beginning of the period, working-age individuals of type (c,e,y) and age t learn their cash in hand x_t, the persistent component of productivity ξ_t, and health state h_t. All these variables form the state of the individual: x_t is payoff relevant in the current period, and the other variables serve to predict next-period outcomes. Based on this information, individuals decide on consumption c_t and savings k_{t+1}. At the end of the period, there are new realizations of the shocks for survival, health, labor force participation, productivity (persistent and transitory), and medical expenses.

The optimization problem for working-age individuals ($t < R$) as:

$$V_t^{\mathrm{c,e,y}}(x,h,\xi) = \max_{c,k'} \left\{ u(c) + \beta s_t^{\mathrm{e,y}}(h) \sum_{h'} \Gamma_t^{\mathrm{e,y}}(h)\, \mathbb{E}_{\ell,\xi,\zeta,\epsilon}[V_{t+1}^{\mathrm{c,e,y}}(x',h',\xi')] + \beta^{T-t}\left(1 - s_t^{\mathrm{e,y}}(h)\right) v(k') \right\}$$

$$\begin{aligned}
&\text{s.t.} \\
&c + k' = x \\
&x' = \min\left\{ w_{t+1}^{\mathrm{c,e,y}}(h',\xi',\epsilon')\ell' - Tax + (1+r)k' - m_{t+1}^{\mathrm{e}}(h',\zeta'),\ \underline{x} \right\} \\
&Tax = T\left(w_{t+1}^{\mathrm{c,e,y}}(h',\xi',\epsilon')\ell'\right) + \tau_{MCR} w_{t+1}^{\mathrm{c,e,y}}(h',\xi',\epsilon')\ell' + \tau_{ss}\min\{w_{t+1}^{\mathrm{e,y}}(h',\xi',\epsilon')\ell', w_{ss}\}
\end{aligned}$$

The optimization problem for retired individuals ($t \geq R$) is analogous, with the constant ξ_{R-1} instead of ξ in the state space, no social security taxes, and a deterministic pension $p^{\mathrm{c,e,y}}(\xi_{R-1})$ instead of stochastic labor earnings.

6 Calibration

The calibration of our model has three distinct parts: the parameters related to the life cycle model of adults (Section 6.1), the parameter $\bar{b}$ driving the value of life, which does not affect the outcomes in the life cycle model (Section 6.2), and the parameters shaping unobserved heterogeneity in the early life stage (Section 6.3).

6.1 Stage 2: Adult life

The stochastic processes for survival and health transitions, $s_t^{\mathrm{e,y}}(h)$ and $\Gamma_t^{\mathrm{e,y}}(h)$, are taken from Section 3.2. We fix the interest rate at 4%, the risk aversion parameter σ at 1, and the discount factor β at 0.98. The calibration of taxes and social security parameters is standard and explained in Appendix D.

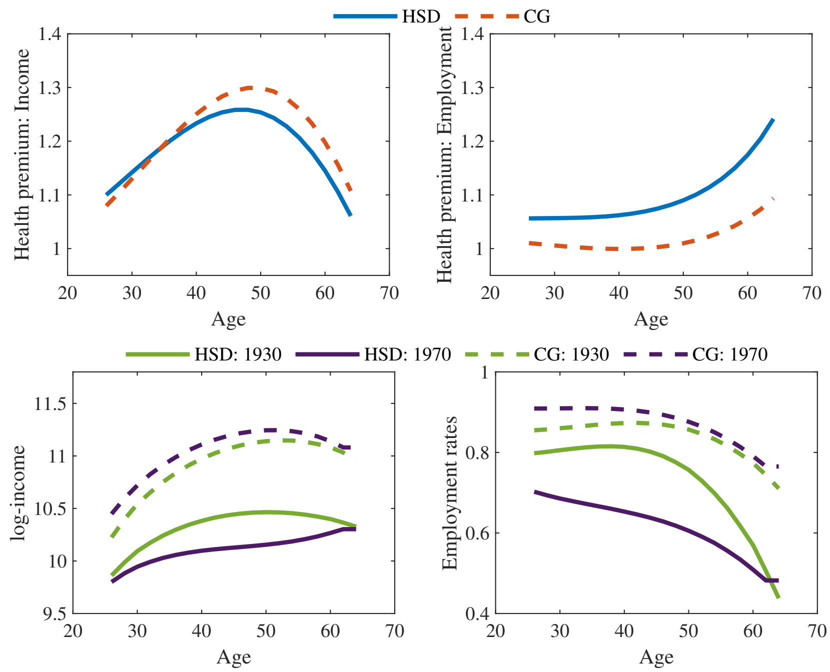
6.1.1 Income process

We use decennial census data spanning from 1940 to 2020 and PSID data covering the years 1999 to 2019 to parameterize the wage $w_t^{\mathrm{c,e,y}}(h,\xi,\epsilon)$ and employment $\mathrm{Prob}\left(\ell_t = 1 | \mathrm{c}, \mathrm{e}, t, h_t\right)$ processes. The advantage of the census data is that it allows us to observe the labor market trajectories of individuals in different cohorts. However, the census data does not report information on health and is not a panel, which is where the PSID becomes useful.

We model the deterministic component of wages, $\omega_t^{\mathrm{c,e}}(h_t)$, as a function of a health dummy and a cubic polynomial in age, both fully interacted with education and cohort dummies. Then, for employed individuals aged between 25 and 60, we use the Census data to regress income against a cubic polynomial on age fully interacted with cohort and education dummies.[13] Since health status

[13] We define as employed those individuals reporting annual labor earnings above 3,770 USD per year.

Figure 4: Deterministic Income Component and Employment Rates



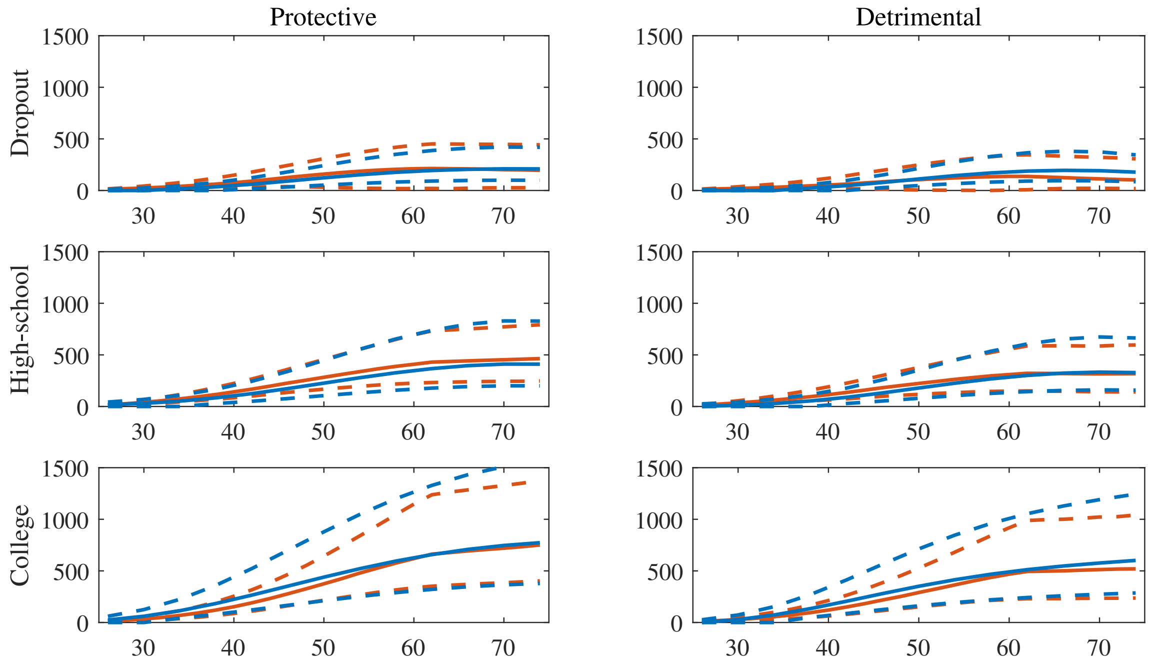
Notes: The top two panels report the ratios of the income and employment rate between good and bad health for two education groups (CG and HSD), data from PSID. The bottom two panels report (log) income and employment rates for two education groups (CG and HSD) in two different cohorts (1930 and 1970), data from Census.

is not directly observable in Census data, we need to impute its effect on wages using PSID data. To do so, we use the PSID data on employed male individuals aged between 25 and 60, interviewed between 1999 and 2019, to regress log income on a health dummy and a cubic polynomial of age, both fully interacted with education dummies. Then, we assume that the health effect on wages observed in the PSID coincides with the one in the census data for all the cohorts. Hence, our assumption here is that, while the education effect on wages varies across cohorts, the health effect does not. Finally, we use the PSID data to estimate the stochastic components of income, ξ_t and ϵ_t. For this purpose, we use Bayesian methods. The assumption here is that the stochastic component of wages has not changed across cohorts.

We model the employment probability $\text{Prob}\,(\ell_t = 1|\text{c}, \text{e}, t, h_t)$ as a linear probability model, which allows us to follow the same steps as in the estimation of wages above.

The two upper panels of Figure 4 report the ratio of income (left panel) and employment rates (right panel) of individuals in good health versus bad health for college and high school dropouts. The figure shows that, when employed, individuals in good health earn on average 20% higher wages than individuals in bad health and that there is little variation across education groups. The figure also shows that, for college graduates, the employment rates are not very different across health states until age 50, but that an employment health premium of around 5% develops after that age. Instead, the employment health premium is large for high school dropouts, reaching 20% at age 60. Overall, this evidence uncovers larger labor market losses of bad health for the less

Figure 5: Model fit: wealth trajectories



Notes: The solid blue lines represent the median wealth in the data, by age in each education and lifestyle group. The solid red lines are the model predictions. The dashed lines represent the 25th and 75th percentiles of wealth for each group, with again blue being data and red model.

educated.

The bottom-left panel of Figure 4 shows the average age profile of wages for the least and most educated groups and for individuals born in 1930 and 1970. The figure reveals a large increase in the college wage premium across cohorts, which is driven by both an increase in college wages and a decrease in the wages of high school dropouts. The bottom-right panel of the figure shows a large increase in the college employment premium across cohorts, which is mainly driven by the well-known decline in employment rates of the least educated individuals. All in all, between the 1930 and 1970 cohorts, the average lifetime labor earnings have declined 18 percent for the high school dropouts, while they have increased 12 percent for college graduates.

Finally, Table E.1 in Appendix E presents the estimated parameters of the stochastic component of the income process. It shows that, among individuals who are employed, those with higher education face larger and more persistent shocks. Additionally, a greater proportion of the variance in their earnings is explained by the persistent component.

6.1.2 Matching wealth trajectories

Given the parameters discussed above, we estimate the remaining model parameters using the simulated method of moments. We do so by minimizing the sum of squared differences between median assets by education e, lifestyle y, and age t for individuals born in 1930s in the data and in

Table 4: Calibrated parameters

Parameter	Description	Value	
		Benchmark	Perceptions
Panel A: Adult Life			
$\underline{x}$	Income floor	17.40	17.35
b_0	Bequest motive: marginal utility	8.86	9.23
b_1	Bequest motive: non-homoteticity	376.64	455.35
$\bar{b}$	Value of life	0.94	1.12
Panel B: Early Life			
μ_{HSG}	Average cost of HSG education	5.39	6.19
μ_{CG}	Average cost of CG education	27.37	26.05
μ_{PRO}	Average. cost of PRO lifestyle	2.58	-2.21
σ_{HSG}	Sd. cost of HSG education	3.00	2.00
σ_{CG}	Sd. cost of CG education	24.50	18.92
σ_{PRO}	Sd. cost of PRO lifestyle	6.07	12.00
λ_{1930}	Probability of right expectation	—	0.55

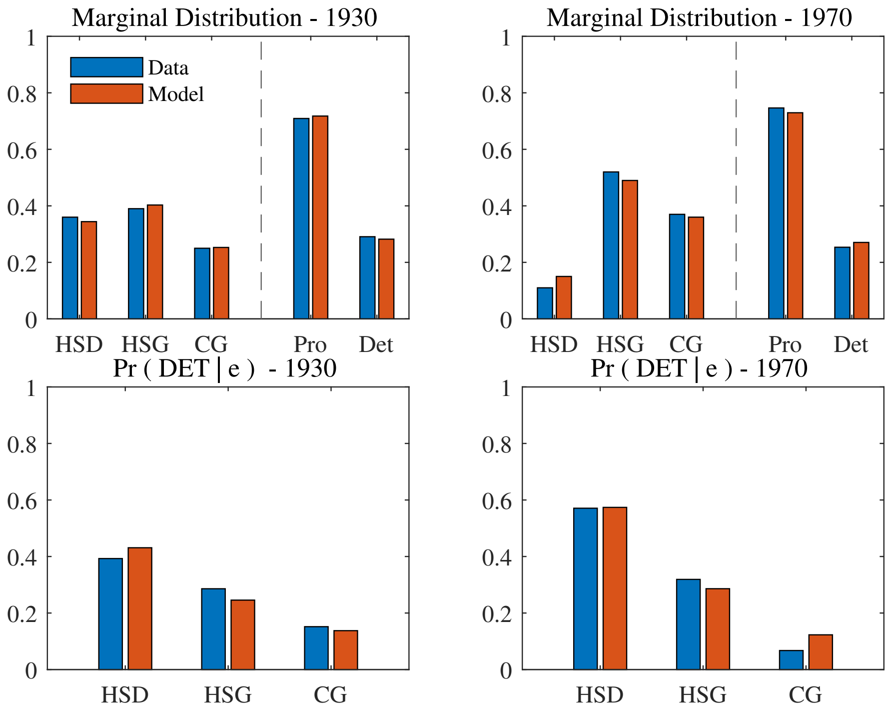
Notes. Panel A: internally calibrated parameters for the adult life stage. Panel B: calibrated parameters for the distributions of the utility costs τ_{PRO}, τ_{CG}, and τ_{HSG} in the early life stage of the model. Column "Benchmark" refers to the main model. Column "Perceptions" refers to the model extension discussed in Section 7.3.

the model. The set of parameters to be estimated is $\{\underline{x}, b_0, b_1\}$.

The first thing we need to do is to recover the wealth distribution across the unobserved health behavior types. For this purpose, we model the observed wealth distribution as a mixture model, see Appendix A.5 for details. We present selected moments of the estimated wealth distribution conditional on age, education, and type in Figure 5. In particular, the solid blue lines represent the median wealth, while the dashed blue lines represent the 25th and 75th percentiles. As it is well known, we see how wealth accumulation is positively correlated with education. In addition to that, we highlight two important results. First, conditional on education, wealth accumulation is stronger for the *protective* type. Second, the difference in wealth accumulation across types is especially apparent within college-educated individuals and smaller within the other two education categories. Overall, our calibrated model can reproduce these facts well. It is worth noting that although we only target the medians, the model is able to match the 25th and 75th quantiles too.

Finally, Panel A in Table 4 reports our estimated parameters (see "Benchmark" column). The estimated bequest parameters b_0 and b_1 are 8.86 and 376.64, respectively. In a model with risk aversion equal to 1, these values imply that in the period before certain death, the bequest motives become active at \$42,500, and the marginal propensity to bequeath is 90 cents out of every additional dollar for bequests. The income floor is estimated at \$17,400 dollar. These parameters fall in the range of parameters estimated by the previous literature.

Figure 6: Model Fit: First-Stage



Notes: The top two panels report the marginal distribution of education and lifestyle types for two different cohorts. The bottom panel reports the distribution of lifestyle types conditional on education choices for the same two cohorts.

6.2 Value of a Statistical Life (VSL)

To identify the parameter $\bar{b}$, we use a measure of the value of a statistical life (VSL). The use of VSL to calibrate the marginal rate of substitution between life years and consumption in macro models follows Rosen (1988) and is today relatively common.[14] The VSL comes from the estimated wage premium for a given probability of a fatal accident in risky jobs. This literature delivers numbers in the range of \$1 to \$7 million to save one life. Because the empirical estimates of a VSL typically come from blue-collar jobs, we want the model to deliver a VSL for the average high school dropout of 35 years of age. We target a VSL of 1,000,000 for a high school graduate aged 40 born in the 1930s. See Appendix D.4 for details.

6.3 Stage 1: Early life

The cost of *protective* behavior is assumed to be normally distributed: $\tau_{\text{PRO}} \sim N\left(\mu_{\text{PRO}}, \sigma^2_{\text{PRO}}\right)$. The cost of graduating from high school and the cost of college education are also assumed to be normally distributed, with $\tau_{\text{HSG}} = \mu_{\text{HSG}} + \sigma_{\text{HSG}}\epsilon_e$ and $\tau_{\text{CG}} = \mu_{\text{CG}} + \sigma_{\text{CG}}\epsilon_e$, respectively with $\epsilon_e \sim N(0, 1)$. In order to calibrate the parameters of these distributions, we require the model to match the joint

[14] See for instance Murphy and Topel (2006), Hall and Jones (2007), Jones and Klenow (2016), De Nardi et al. (2024), Mahler and Yum (2022), or Hong et al. (2024) among others.

distribution of education and lifestyles for two different cohorts: 1930 and 1970. The value functions $V_0^{\mathrm{c,e,y}}$ for each cohort c vary differently in terms of education e and lifestyle y due to the different paths of wages and employment rates over the life cycle, with the expected real labor earnings of high school dropouts being slightly lower for the 1970s cohort, the expected real labor earnings of college graduates being larger for the 1970 cohort, and hence the college premium being larger for the 1970 cohort (see Section 6.1.1).

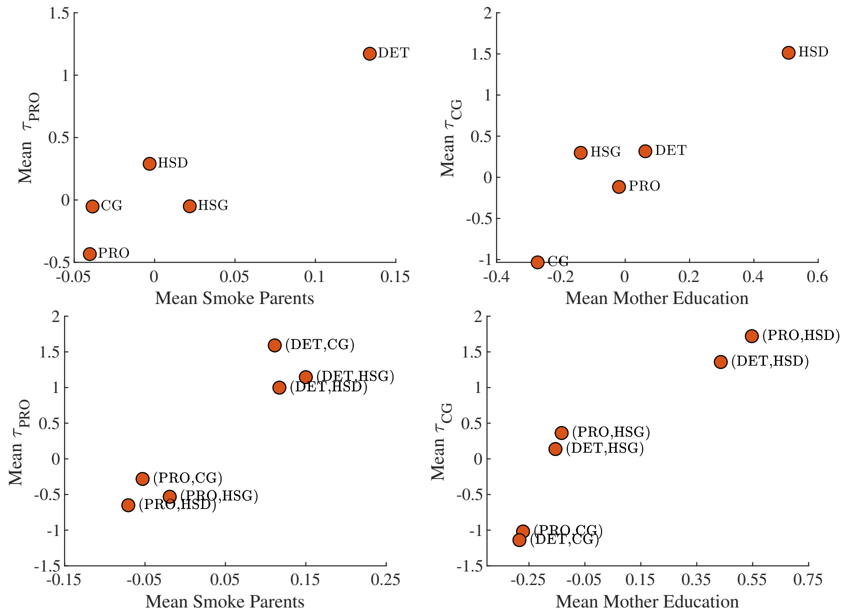
6.3.1 Identification and results

We have 6 parameters to estimate: $(\mu_{\mathrm{CG}}, \mu_{\mathrm{HSG}}, \mu_{\mathrm{PRO}}, \sigma_{\mathrm{CG}}, \sigma_{\mathrm{HSG}}, \sigma_{\mathrm{PRO}})$. Given that we require the model to match the joint distribution of (e,y) for two different cohorts we have 10 moment conditions ($3 \times 2 - 1$ targets per year). In general, one may think that the average cost parameters $(\mu_{\mathrm{CG}}, \mu_{\mathrm{HSG}}, \mu_{\mathrm{PRO}})$ are identified by the marginal distributions of education and lifestyle in the 1930 cohort, while the dispersion parameters $(\sigma_{\mathrm{CG}}, \sigma_{\mathrm{HSG}}, \sigma_{\mathrm{PRO}})$ are identified by the changes of these distributions as the wage paths change between the 1930 and 1970 cohorts.[15] However, by targeting the *joint distribution* of education and lifestyles, these parameters also shape the level and changes in the distribution of lifestyles conditional on education.

The upper panels of Figure 6 show that the model matches well the marginal distributions of education and health behavior in the 1930 cohort as well as their changes across cohorts due to changes in the wage structure. The lower panels of Figure 6 show that the model is also able to approximate well the level and changes in the education gradient of lifestyles. In particular, in the 1930s cohort, the proportion of individuals with *detrimental* lifestyles within the high school dropout category is 39% in the data and 43% in the model, while for college graduates these figures are 15.1% in the data and 13.8% in the model. Furthermore, given the observed changes in wage and employment trajectories by education, the model reproduces well the worsening of health-related behavior by the less educated between the 1930 and 1970 cohorts, with the proportion of individuals with *detrimental* lifestyles among high school dropouts increasing by 17.8 percentage points in the data and 14.3 percentage points in the model. However, the model is less successful at matching the improvements in health behavior of the higher-educated, as the proportion of individuals with *detrimental* lifestyles among college graduates has declined by 8.5 percentage points in the data and 1.5 percentage points in the model. This implies that quantitatively, the model accounts for 15.8 out of the 26.3 percentage points of the observed increase in the education gradient in lifestyles. Consequently, the increase in the life-expectancy gradient is estimated at 1.1 years in the model and 1.9 years in the data. Hence, the calibrated model, with only wage and employment changes, accounts for around 60% of the overall increase in the life-expectancy gradient between college graduates and high school dropouts between the 1930 to 1970 cohorts.

[15] This extends the identification strategy of Heathcote et al. (2010), whose first stage contains a college education choice but does not consider a lifestyle choice.

Figure 7: Shocks and Parental Characteristics



Notes: The top two panels report the average τ_{PRO} (top left) and the average τ_{CG} (top right) for each education and lifestyle category in the calibrated model (Y-axis) against the average share of smoking parents (top left) and average share of high school dropout mothers (top right) for the same groups in the data (X-axis). The bottom two panels report the same information for each education-lifesytle group.

6.3.2 Validation

A relevant question to validate the calibration is whether the recovered distribution of the unobserved τ_{PRO} and τ_{CG} resembles relevant proxies in the data. To do so, we use the rich data in the HRS. In particular, in the HRS we have data on whether the main carer smoked when the respondent was a kid and on parental education (whether the mother did not finish high school). Taking these two variables as imperfect proxies for τ_{PRO} and τ_{CG} respectively, we compare the pattern of selection of τ_{PRO} and τ_{CG} within each education and lifestyle group in the model with the one observed in the data.

First, the model predicts that individuals who choose a *protective* lifestyle have on average lower τ_{PRO} and individuals who choose college education have on average lower τ_{CG}. We can see this in the y-axes on the top two panels in Figure 7. Consistently, in the data, individuals classified as *protective* are less likely to have smoking parents than individuals classified as *detrimental*, and more educated individuals are less likely to have a high school dropout mother, see the x-axes on the top two panels in Figure 7.

Second, the model predicts that individuals who choose a *protective* lifestyle also have on average lower τ_{CG} and individuals who study college education have on average lower τ_{PRO}. This is due to the complementarity of the two investments, and it is clearly apparent in the y-axes of the top two panels in Figure 7. In the data, we also find that individuals classified as *protective* are less likely to have a high school dropout mother than those classified as *detrimental*, and that college-educated individuals are less likely to have smoking parents than high school dropouts. Only the high school graduates do not conform with the model predictions, as the share of them with smoking parents, while being above the one of college graduates, is also above the one of high school dropouts.

Third, within each lifestyle choice, the least educated individuals have lower τ_{PRO}, that is, they are better selected in terms of the utility cost of good lifestyle, see the y-axis in the bottom-left panel of Figure 7. The reason is that, due to the complementarity of investments, what leads a low-educated individual to choose a *protective* lifestyle is a low cost of *protective* lifestyle compared to the more educated individuals. Analogously, what leads a highly educated individual to choose a *detrimental* lifestyle is the high cost of *protective* lifestyle compared to the less educated individuals. In the data, we see that these patterns are also present: among the individuals classified as *protective*, the parents of those who go to college are more likely to smoke than the parents of those who drop out of high school. Among the individuals classified as *detrimental*, the parents of those who go to college are also more likely to smoke than the parents of those who drop out from high school, see the x-axis in the bottom-left panel of Figure 7.

And fourth, within education choice, the individuals with *detrimental* lifestyle have lower τ_{CG} than individuals with *protective* lifestyle, see the y-axis in the bottom-right panel of Figure 7. That is, due to the complementarity of investments, *detrimental* individuals are better selected in terms of education costs. The pattern in the data is similar. Both within high school dropouts and within high school graduates, the mothers of individuals classified as *detrimental* are less likely to be high school dropouts, see the x-axis in the right panel of Figure 7. Among college graduates, the likelihood of high school dropout mothers is virtually the same across lifestyle groups.

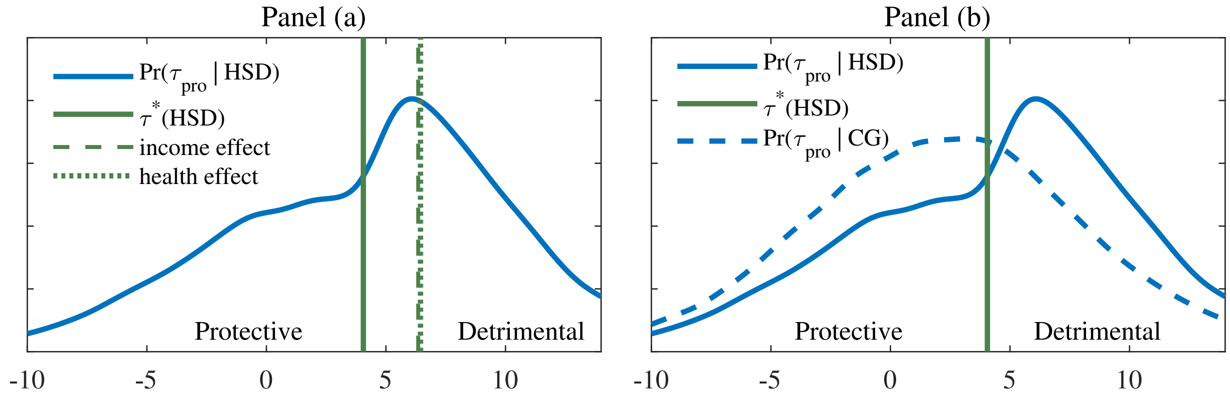
7 Counterfactual results

Given our calibrated model, we aim to answer two questions. First, why is there an education gradient of lifestyles (Section 7.1), and second, what has been the effect of the rise in the education wage premium on the increase in education gradient of life expectancy across cohorts (Section 7.2). Finally, we examine a model extension in Section 7.3.

7.1 The education gradient of lifestyles

Our model incorporates three mechanisms that can generate an educational gradient of lifestyles. Firstly, higher expected income among the more educated encourages healthier behavior as life becomes more valuable with higher consumption possibilities. Secondly, as detailed in Section 3, the estimated health transitions imply a higher yield of a *protective* lifestyle investment for the more educated (the *protective* lifestyle has a larger effect on health outcomes for the more

Figure 8: Decomposition of lifestyle choices conditional on education (1970 cohort)



Notes: In Panel (a) the blue line describes the density function of τ_{PRO} for individuals who choose HSD education. The solid thick green vertical line reports the threshold $\tau^*(\text{HSD})$ that separates HSD individuals into *detrimental* and *protective*. The dashed and dotted green vertical lines report this same threshold if HSD individuals had, respectively, the income prospects and health yield of *protective* lifestyle as CG individuals. The dashed blue line in Panel (b) describes the density function of τ_{PRO} for individuals who choose CG education.

educated). Finally, given the two previous complementarities, individuals facing lower costs of *protective* behavior (lower τ_{PRO}) are more likely to pursue higher education. This means that low τ_{PRO} individuals will be more frequent among the highly educated, as stated by Proposition 1.

To gauge the importance of each mechanism, we start by graphically showing the choice of lifestyle conditional on education for the calibrated model. The solid blue line in Panel (a) of Figure 8 represents the distribution of *protective* behavior costs (τ_{PRO}) for individuals who choose to drop out of high school in the 1970 cohort. The vertical solid green line represents $\tau^*(\text{HSD}) \equiv V^{70,\text{HSD},\text{PRO}} - V^{70,\text{HSD},\text{DET}}$, the value of choosing a *protective* lifestyle over a *detrimental* lifestyle for those choosing to drop out of high school. Individuals with $\tau_{\text{PRO}} < \tau^*(\text{HSD})$ opt for *protective* behavior, while the rest choose *detrimental* behavior. The integral of the distribution of τ_{PRO} between minus infinity and $\tau^*(\text{HSD})$ represents the fraction of dropouts adopting *protective* behavior.

Next, we conduct a series of counterfactual experiments. In these experiments, we keep individuals' education choices fixed and observe how their lifestyle type investments would differ under various scenarios. In the first scenario, we simulate the behavior of high school dropouts if they were to have the income prospects of college graduates. Due to the higher consumption possibilities, the increase in expected earnings leads to higher values for both y types, which we denote as $\widetilde{V}^{70,\text{HSD},\text{PRO}}$ and $\widetilde{V}^{70,\text{HSD},\text{DET}}$. However, this gain is higher for y = PRO, as life expectancy is higher for this type, and the higher consumption flow is enjoyed for more years. Thus, we have that $\tilde{\tau}^*(\text{HSD}) \equiv \widetilde{V}^{70,\text{HSD},\text{PRO}} - \widetilde{V}^{70,\text{HSD},\text{DET}} > \tau^*(\text{HSD})$, which means that the threshold value that prompts individuals to adopt *protective* health behavior shifts to the right, see the dashed vertical line. This indicates that income largely influences why high school dropouts tend to adopt more *detrimental* behavior. If faced with the same expected income as college graduates, the proportion of high school dropouts choosing *detrimental* behavior would decrease from 57% to 40%, reducing the education gradient of life-expectancy gap in 1.3 years (from 7.3 to 6.0 years, an 18% reduction),

Table 5: Decomposition: 1970s cohort

	Pr(PRO|HSD)	Pr(PRO|CG)	ΔPr(PRO)	LE(HSD)	LE(CG)	ΔLE
Model	0.43	0.88	0.45	24.6	31.9	7.3
Same lifestyle	0.88	0.88	0.00	27.7	31.9	4.2
Income	0.60	0.88	0.28	25.9	31.9	6.0
Health	0.60	0.88	0.28	25.9	31.9	6.0
Selection	0.62	0.88	0.26	26.5	31.9	5.8

Notes: The first row reports results from the calibrated model. The second row reports the results of a model where the lifestyle distribution is the same in all education groups. The 3rd to 5th rows perform the counterfactuals where high school dropouts are given, respectively, the same income prospects, the same health gains of *protective* behavior, and the same underlying distribution of τ_{PRO} as college graduates. See Section 7.1 for details.

see 3rd row in Table 5.

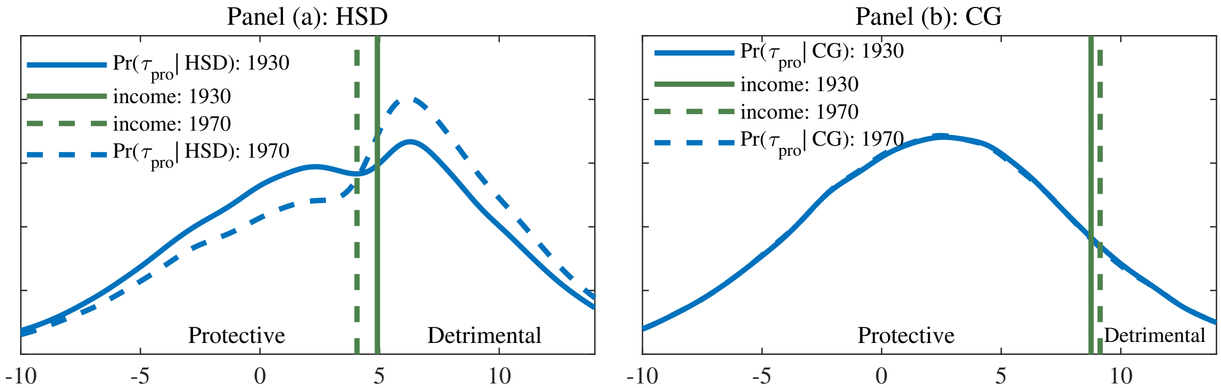
In the second scenario, we simulate the behavior of high school dropouts if they were to face the same relative gains from *protective* health transitions as college graduates. For this purpose, we first compute a counterfactual health transition model for high school dropouts with *protective* behavior that matches the relative gains from *protective* behavior (compared to *detrimental* behavior) that we observed for the college graduates. The improved health transitions for dropout *protective* result in higher values of $\widehat{V}^{70,\text{HSD,PRO}}$, which implies that $\hat{\tau}^*(\text{HSD}) \equiv \widehat{V}^{70,\text{HSD,PRO}} - V^{70,\text{HSD,DET}} > \tau^*$, see the vertical dotted line. If faced with the relative gains from *protective* behavior as college graduates, the proportion of high school dropouts choosing *detrimental* behavior would decrease from 57% to 40%, reducing the life-expectancy gap as much as the income effect, see 4th row in Table 5.

Lastly, we look at selection. In the third scenario, we compute the fraction of individuals that choose *detrimental* type among high school dropouts in the case that high school dropouts had the same distribution of τ_{PRO} as college graduates. Recall that Proposition 1 establishes that the pool of less educated individuals has a worse distribution of τ_{PRO} than the pool of more educated individuals, and hence they are more prone to choose a *detrimental* lifestyle just because of that. The right panel in Figure 8 displays the distribution of the cost of *protective* behavior for both high school dropouts and college (dashed line). It shows that the mass of individuals below the $\tau^*(\text{HSD})$ threshold (vertical line) is larger for the distribution of *protective* costs faced by the college individuals. If faced with the distribution of costs τ_{PRO} of college-educated individuals, the proportion of *detrimental* individuals would decrease from 57% to 38%, reducing the education gradient of life-expectancy gap in 1.5 years (from 7.3 to 5.8 years, a 20% reduction), see 5th row in Table 5.

7.2 Changes over time

The econometric model estimated in Section 4 uncovers an increase in the life expectancy gap between college graduates and high school dropouts of 1.9 years across the cohorts from the 1930s to the 1970s. This increase is driven by lower (higher) health investments among high school dropouts (college graduates). Our calibrated model, as discussed above, replicates 60% of this

Figure 9: Cost of *protective* lifestyle across cohorts



Notes: In Panel (a) the solid blue line describes the density function of τ_{PRO} for individuals in the 1930 cohort who chose HSD education. The solid green vertical line reports the threshold $\tau^*(\text{HSD})$ that separates HSD individuals into *detrimental* and *protective* in the 1930 cohort. The dashed green vertical line reports this same threshold if the income prospects of HSD individuals were as in the 1970 cohort. The dashed blue line describes the density function of τ_{PRO} for individuals who choose HSG education in the 1970 cohort. Panel (b) reports the same information for CG individuals.

observed increase. The question here is how much of this increase is due to the direct effect of changes in individuals' income prospects and how much is due to changes in the composition of individuals in each education group.

To quantify the income effect while controlling for selection, we fix the distribution of health behavior costs faced by individuals in different education categories to the one for the 1930 cohort. We then analyze how investments in lifestyle types would have changed solely due to changes in income. In Figure 9, the solid blue line represents the distribution of the cost of *protective* behavior τ_{PRO} faced by individuals in 1930 for high school dropouts (Panel a) and college graduates (Panel b). These two distributions are different due to selection, see Proposition 1. The solid green vertical line indicates the threshold value at which individuals of that cohort decided to switch from a *protective* to a *detrimental* lifestyle conditional on each education choice. The dashed green vertical line in the left panel illustrates that for high school dropouts, the threshold value decreased across cohorts due to declines in income prospects, and consequently, individuals dropping out of high school are less willing to invest in their health for a given distribution of τ_{PRO}. The quantitative results are reported in Table 6: the share of *protective* individuals would drop by 5 percentage points in this counterfactual (compared to 14 percentage points predicted by the model). The opposite happens for college graduates, with the share of *protective* individuals increasing by 1 percentage point in this counterfactual (which is the full increase predicted by the model). Overall, we find that the direct effect of increasing divergence in income prospects across education groups accounts for 42% of the increase in the life expectancy gradient between the 1930s and 1970s cohorts as predicted by the model (and 23% of the increase compared to data).

Next, we look at changes in selection. The dashed green line in Figure 9 illustrates the distribution of behavior costs conditional on education in the 1970s. Driven by changes in income across cohorts the cost distribution worsened for high school dropouts born in the 1970s but did not

Table 6: Decomposition: changes across cohorts

	$\Delta_c\Pr(\text{CG})$	$\Delta_c\Pr(\text{HSD})$	$\Delta_c\Pr(\text{PRO})$	$\Delta_c\Pr(\text{PRO}\|\text{CG})$	$\Delta_c\Pr(\text{PRO}\|\text{HSD})$	$\Delta_c\Delta_e\text{LE}$
Data	0.12	-0.25	0.04	0.08	-0.18	1.92
Model	0.11	-0.19	0.02	0.01	-0.14	1.06
Income				0.01	-0.05	0.44
Selection				0.00	-0.09	0.63

Notes: The first row reports statistics from the data, comparing the cohorts of 1930 and 1970. The second row reports the results from the calibrated model. The 3rd performs the counterfactual exercise where only income prospects change between 1930 and 1970; the 4th row performs the counterfactual exercise where only the distribution of τ_{PRO} conditional on education changes between 1930 and 1970. See Section 7.2 for details.

quantitatively change for the college graduates. To quantify the selection effect while controlling for income, we fix the income prospects faced by individuals in different education categories to the one in 1930. We then analyze how their lifestyle choices would have changed solely due to changes in the distribution of health behavior costs. We find that if only the distribution of health behavior cost conditional on education had changed, the life-expectancy gradient would have increased by 0.63 years or 59% of the increase in the life expectancy gradient predicted by the model and 33% of the increase in the data.

7.3 Changes in perceptions

One possible concern with our exercise is the assumption that individuals have perfect information regarding the effects of lifestyles on health dynamics. In fact, numerous studies suggest that the impact of different health-related behaviors on health outcomes —especially smoking— was not fully understood for the cohorts born during the first half of the 20th century. To address this issue, we extend our model to allow for imperfect knowledge about the health dynamics associated with each behavior, and we quantify to which extent changes in the life-expectancy gradient between the 1930 and 1970 cohorts could be explained by changes in perceptions.

Modeling. To do so, we introduce a cohort-specific parameter $\lambda^{\text{c}} \in [0, 1]$, which captures the degree to which individuals born in cohort c understand the link between a given lifestyle and its associated health dynamics. The objective function in the second-stage worker's problem of Section 5.2.3 is now expressed as:

$$V_t^{\text{c,e,y}}(x, h, \xi) = \max_{c,k'} \left\{ u(c) + \beta \left[\lambda^{\text{c}} W_t^{\text{c,e,y}}(x, h, \xi) + (1 - \lambda^{\text{c}})\, \widehat{W}_t^{\text{c,e,y}}(x, h, \xi) \right] \right\} \tag{6}$$

where

$$W_t^{\mathrm{c,e,y}}(x,h,\xi) = s_t^{\mathrm{ey}}(h) \sum_{h'} \Gamma_t^{\mathrm{ey}}\left(h'|h\right) \mathbb{E}\left[V_{t+1}^{\mathrm{c,e,y}}(x',h',\xi')\right] + \beta^{T-t-1}\left(1 - s_t^{\mathrm{e,y}}(h)\right) v_{t+1}(k') \quad (7)$$

$$\widehat{W}_t^{\mathrm{c,e,y}}(x,h,\xi) = s_t^{\mathrm{e}}(h) \sum_{h'} \Gamma_t^{\mathrm{e}}\left(h'|h\right) \mathbb{E}\left[V_{t+1}^{\mathrm{c,e,y}}(x',h',\xi')\right] + \beta^{T-t-1}\left(1 - s_t(h)\right) v_{t+1}(k') \quad (8)$$

As shown in equations (6) to (8), λ^{c} represents the probability associated with the continuation value $W^{\mathrm{c,e,y}}$, which is built with perfect information on health dynamics conditional on lifestyle y. The complementary probability, $1-\lambda^{\mathrm{c}}$, corresponds to the continuation value $\widehat{W}^{\mathrm{c,e,y}}$ for the scenario where the individual incorrectly perceives that health dynamics depend only on education and not on lifestyle. In this manner, λ^{c} measures the degree of knowledge about the effects of y on health dynamics.

Calibration. For our quantitative exercise, we set $\lambda^{1970} = 1$ (full knowledge for the 1970 cohort) and re-estimate the whole model for the 1930 cohort, including the extra parameter λ^{1970}. This implies re-estimating the second-stage parameters $(\underline{x}, b_0, b_1, \bar{b})$ to match the life cycle wealth profiles and all the first-stage parameters to match the joint distribution of education and lifestyles for the 1930 and 1970 cohorts.[16] We find $\lambda^{1930} = 0.55$, which states that young individuals in the 1930 cohort only understood about 1/2 of the true effects of their lifestyle choices on future health outcomes. The second-stage parameters and model fit (see last column in Table 4 and Figure E.2) do not change much compared to the benchmark model. Regarding the first-stage parameters, there are two significant changes (see also last column in Table 4). First, there is a decline of μ_{PRO} from 2.58 in the benchmark model to -2.21 in the model with imperfect information. This happens because, given that the 1930 cohort only gives 55% of probability to the poor (good) health outcomes associated with a *detrimental* (*protective*) lifestyle, the actual difference in the values between the two choices narrows, which reduces the perceived value of choosing a (*protective*) lifestyle. Hence, the cost of a *protective* lifestyle has to decline as compared to the full information model to hit the same share of individuals choosing a *protective* lifestyle. Second, there is an increase in σ_{PRO} from 6.07 to 12.0. This happens because, with a larger σ_{PRO}, the increase over time in the share of *protective* individuals following the wage changes across cohorts is smaller. With the increase in λ^{c} the model introduces a new driver for the increase in the share of *protective* individuals, so the strength of the changes in income has to decline (σ_{PRO} has to increase) to hit the same target. Finally, we note that the model fit for the first stage (see Figure E.3) is also good in the imperfect information model, and it improves the benchmark model in its ability to match the change across cohorts in the education gradient of lifestyles. In particular, comparing the first two rows of Table 6 and 7, we see that the education gradient of *protective* lifestyles increases by 26 percentage points

[16] To perform this exercise, we set an outer loop in the calibration of the benchmark model where we search for the value of λ^{1930} that best matches the observed increase in the life-expectancy gradient between 1930 and 1970. Furthermore, we calibrate $\Gamma_t^e(h'|h)$ and $s_t^e(h)$ as the average health transitions and survival for individuals in the 1930 cohort weighted by the frequency of health behavior types at each age for each education level for that cohort.

Table 7: Decomposition: changes across cohorts (changes in perceptions)

	Δ_cPr(CG)	Δ_cPr(HSD)	Δ_cPr(PRO)	Δ_cPr(PRO|CG)	Δ_cPr(PRO|HSD)	$\Delta_c\Delta_e$LE
Data	0.12	-0.25	0.04	0.08	-0.18	1.92
Model	0.13	-0.23	0.09	0.10	-0.15	1.94
Income				0.01	-0.01	0.19
Selection				0.01	-0.17	1.23
Perceptions				0.08	0.07	0.28

Notes: The first row reports statistics from the data, comparing the cohorts of 1930 and 1970. The second row reports the results from the calibrated model. The 3rd performs the counterfactual exercise where only income prospects change between 1930 and 1970; the 4th row performs the counterfactual exercise where only the distribution of τ_{PRO} conditional on education changes between 1930 and 1970. See Section 7.2 for details.

in the data, 15 percentage points in the benchmark model, and 25 percentage points in the model with a change in perceptions.

Results. Table 7 shows that the model incorporating changes in perceptions across cohorts successfully matches the increase in the life-expectancy gradient. This is largely due to the fact that changes in perceptions contribute an additional 15 percentage points to the increase in the education gradient of *protective* lifestyles out of the 25 percentage points predicted by the model (see the last row in Table 7). Compared to the benchmark model, the increase in the lifestyle gradient due to changes in the income prospects falls from 6 to 2 percentage points. This happens because the model with imperfect information requires a larger σ_{PRO} as discussed above. In contrast, the effect of selection is stronger in the model with changing perceptions: in the benchmark model, selection accounts for a 9 percentage point increase in the education gradient of protective lifestyles, whereas in the model with changing perceptions, it accounts for an 18 percentage point increase. This is because changes in perceptions increase the complementarities between education and lifestyle investments, which, as Proposition 1 shows, are central to selection. One of the key reasons complementarities emerge in our model is that protective lifestyles improve health transitions more significantly for individuals with higher education. When individuals believe that lifestyles have little impact on future health outcomes, these complementarities are reduced. On the other hand, when the 1970 cohort has a better understanding of the effects of lifestyles on health, complementarities are reinforced.

8 Conclusions

In this paper, we propose a latent variable model to characterize how health behavior influences health dynamics across different education groups. Our findings indicate that health behavior can be parsimoniously summarized into two lifestyles: *protective* and detrimental. We observe that individuals with higher levels of education tend to more frequently choose *protective* behavior, and differences in lifestyles explain 40% of the education gradient in life expectancy. We also find an increasing education gradient in life expectancy gradient between the 1930s and 1970s cohorts, which is driven by worsening lifestyles among the less educated and improved lifestyles among the more educated.

To understand the correlation between lifestyles and education (and the changes across cohorts), we introduce a heterogeneous agents model comprising two distinct stages. Initially, individuals make a one-time decision regarding education and lifestyle, with the value of each choice given by the second stage of the model. In the second stage, agents address a consumption-savings problem subject to income and health risks, as modeled in the econometric framework.

This model enables us to explain the connection between income and health inequality. Health and education decisions are shown to be complementary due to two key factors. Firstly, higher income increases the value of life, leading to greater returns from investing in health. Secondly, as reflected in our calibration process, where we integrate the health dynamics from the econometric model, we observe greater returns to health investment for college-educated individuals. Driven by these complementarities, individuals facing higher costs of maintaining *protective* health behaviors are more likely to self-select into lower education categories.

We calibrate the model to match savings, education, and lifestyle choices across cohorts, and then we use it to understand why lower-educated individuals tend to choose unhealthier lives. Our analysis reveals that lower income and diminished returns in health outcomes from *protective* behavior largely account for the disparities observed across education groups. However, selection also matters.

Finally, the model is able to explain 50% of the increase in health inequality across the 1930s and the 1970s cohorts. 40% of the deterioration in life expectancy among the less educated is driven by worsening economic conditions, and 60% is attributed to selection effects. All improvements in lifestyle among college graduates are explained by improvements in economic conditions. The importance of selection compared to the direct effect of income changes has clear policy implications. To reduce health inequalities, our results suggest that early childhood interventions to foster the adoption of better lifestyles (like healthy food at school or exercise habits) may be more effective than labor market interventions to reduce wage inequality.

The benchmark model cannot fully account for the increase in the life-expectancy gradient observed in the data. Factors such as peer influence, segregation, genetic predispositions, and variations in intergenerational mobility across cohorts are likely significant drivers of health behavior choices made by individuals, which we abstract from in our current analysis. Also, as our main extension shows, improvements in perceptions about the future health consequences of lifestyle choices can also close the gap between the model and the data. All these avenues offer promising directions for future research.

References

Allcott, H., R. Diamond, J.-P. Dubé, J. Handbury, I. Rahkovsky, and M. Schnell (2019): "Food Deserts and the Causes of Nutritional Inequality," *Quarterly Journal of Economics*, 134, 1793–1844. (Cited on page 3.)

Ameriks, J., J. Briggs, A. Caplin, M. Shapiro, and C. Tonetti (2020): "Long-Term Care Utility and Late in Life Saving," *Journal of Political Economy*, 128, 2375–2451. (Cited on page 5.)

Aron-Dine, A., L. Einav, and A. Finkelstein (2013): "The RAND Health Insurance Experiment, Three Decades Later," *Journal of Economic Perspectives*, 27, 197–222. (Cited on page 5.)

Baicker, K., S. Taubman, H. Allen, M. Bernstein, J. Gruber, J. P. Newhouse, E. C. Schneider, B. Wright, A. M. Zaslavsky, and A. Finkelstein (2013): "The Oregon Experiment - Effects of Medicaid on Clinical Outcomes," *New England Journal of Medicine*, 368, 1713–1722. (Cited on page 5.)

Bairoliya, N., R. Miller, and V. M. Nygaard (2024): "Exercise or Extra Fries? How Behaviors Impact Health Over the Life Cycle," Mimeo. (Cited on pages 5 and 6.)

Boháček, R., J. Bueren, L. Crespo, P. Mira, and J. Pijoan-Mas (2021): "Inequality in Life Expectancies across Europe and the US," *Health Economics*, 30, 1871–1885. (Cited on page 5.)

Borella, M., F. Bullano, M. De Nardi, B. Krueger, and E. Manresa (2024): "Health Inequality and Health Types," Mimeo. (Cited on page 4.)

Braun, A., K. Kopecky, and T. Koreshkova (2019): "Old, Frail, and Uninsured: Accounting for Puzzles in the US Long-Term Care Insurance Market," *Econometrica*, 87, 981–1019. (Cited on page 5.)

Bueren, J. (2023): "Long-Term Care Needs and Savings in Retirement," *Review of Economic Dynamics*, 49, 201–224. (Cited on page 5.)

Capatina, E. (2015): "Life-cycle Effects of Health Risk," *Journal of Monetary Economics*, 74, 67–88. (Cited on page 5.)

Card, D. (1995): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, ed. by L. N. Christofides, E. K. Grant, and R. Swidinsky, Toronto: University of Toronto Press, 201–222. (Cited on page 14.)

Case, A. and A. Deaton (2015): "Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century," *Proceedings of the National Academy of Sciences*, 112, 15078–15083. (Cited on page 1.)

——— (2017): "Mortality and Morbidity in the 21st Century," *Brooking Papers on Economic Activity*, Spring, 397–443. (Cited on page 4.)

Chetty, R., M. Stepner, S. Abraham, S. Lin, B. Scuderi, N. Turner, A. Bergeron, and D. Cutler (2016): "The association between income and life expectancy in the United States, 2001-2014," *JAMA*, 315, 1750–1766. (Cited on page 1.)

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. (Cited on page 5.)

Conner, M. and P. Norman (2017): "Health Behaviour: Current Issues and Challenges," *Psychology & Health*, 32, 895–906. (Cited on page 14.)

Cutler, D. M. and A. Lleras-Muney (2010): "Understanding Differences in Health Behaviors by Education," *Journal of Health Economics*, 29, 1–28. (Cited on page 1.)

De Nardi, M., E. French, and J. Jones (2010): "Why Do The Elderly Save? The Role Of Medical Expenses," *Journal of Political Economy*, 118, 39–75. (Cited on page 5.)

——— (2016): "Medicaid Insurance in Old Age," *American Economic Review*, 106, 3480–3520. (Cited on page 5.)

De Nardi, M., S. Pashchenko, and P. Porapakkarm (2024): "The Lifetime Costs of Bad Health," NBER Working Paper 23963. (Cited on pages 4, 5, 22, and A.14.)

Farrell, P. and V. Fuchs (1982): "Schooling and Health: the Cigarette Connection," *Journal of Health Economics*, 1, 217–230. (Cited on page 14.)

Finkelstein, A., S. Taubman, B. Wright, M. Bernstein, J. Gruber, J. P. Newhouse, H. Allen, and K. Baicker (2012): "The Oregon Health Insurance Experiment: Evidence from the First Year," *Quarterly Journal of Economics*, 127, 1057–1106. (Cited on page 5.)

Fonseca, R., P.-C. Michaud, T. Galama, and A. Kapteyn (2021): "Accounting for the rise of health spending and longevity," *Journal of the European Economic Association*, 19, 536–579. (Cited on page 5.)

French, E. and J. Jones (2011): "The Effects of Health Insurance and Self-Insurance on Retirement Behavior," *Econometrica*, 79, 693–732. (Cited on page 5.)

Gold, D. T. and B. McClung (2006): "Approaches to Patient Education: Emphasizing the Long-Term Value of Compliance and Persistence," *The American Journal of Medicine*, 119, 32S–37S. (Cited on page 3.)

Hai, R. and J. Heckman (2022): "The Causal Effects of Youth cigarette Addiction and Education," NBER Working Paper 30304. (Cited on pages 5 and 14.)

Hall, R. and C. Jones (2007): "The Value of Life and the Rise in Health Spending," *Quarterly Journal of Economics*, 122, 39–72. (Cited on page 22.)

Hauser, R. M. and D. L. Featherman (1976): "Equality of Schooling: Trends and Prospects," *Sociology of Education*, 49, 99–120. (Cited on page 14.)

Heathcote, J., K. Storesletten, and G. Violante (2010): "The Macroeconomic Implications of Rising Wage Inequality in the United States," *Journal of Political Economy*, 118, 681–722. (Cited on page 23.)

Heathcote, J., K. Storesletten, and G. L. Violante (2020): "Optimal progressivity with age-dependent taxation," *Journal of Public Economics*, 189, 104074. (Cited on page 17.)

Holter, H. A., D. Krueger, and S. Stepanchuk (2019): "How do tax progressivity and household heterogeneity affect Laffer curves?" *Quantitative Economics*, 10, 1317–1356. (Cited on page A.15.)

Hong, J., J. Pijoan-Mas, and V. Ríos-Rull (2024): "Health, Consumption, and Inequality," Mimeo. (Cited on pages 4, 5, and 22.)

Hosseini, R., K. A. Kopecky, and K. Zhao (2021): "How important is health inequality for lifetime earnings inequality?" FRB Atlanta Working Paper. (Cited on page 5.)

Jones, C. and P. Klenow (2016): "Beyond GDP? Welfare across Countries and Time," *American Economic Review*, 106, 2426–2457. (Cited on page 22.)

Keane, M. and K. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105, 473–522. (Cited on page 14.)

Kitagawa, E. M. and P. M. Hauser (1973): *Differential mortality in the United States: A study in socioeconomic epidemiology*, Harvard University Press. (Cited on page 1.)

Lantz, P. M., J. S. House, J. M. Lepkowski, D. R. Williams, R. P. Mero, and J. Chen (1998): "Socioeconomic factors, health behaviors, and mortality: results from a nationally representative prospective study of US adults," *JAMA*, 279, 1703–1708. (Cited on page 1.)

Li, Y., A. Pan, D. D. Wang, X. Liu, K. Dhana, O. H. Franco, S. Kaptoge, E. Di Angelantonio, M. Stampfer, W. C. Willett, et al. (2018): "Impact of healthy lifestyle factors on life expectancies in the US population," *Circulation*, 138, 345–355. (Cited on page 1.)

Lleras-Muney, A. and F. R. Lichtenberg (2005): "Are the More Educated More Likely to Use New Drugs?" *Annales d'Economie et de Statistique*, 79/80, 671–696. (Cited on page 3.)

Mahler, L. and M. Yum (2022): "Lifestyle behaviors and wealth-health gaps in Germany," *Available at SSRN 4034661.* (Cited on pages 5 and 22.)

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*. (Cited on page 5.)

MCGINNIS, J. M. AND W. H. FOEGE (1993): "Actual Causes of Death in the United States," *JAMA*, 270, 2207–2212. (Cited on page 1.)

MEARA, E. R., S. RICHARDS, AND D. M. CUTLER (2008): "The gap gets bigger: changes in mortality and life expectancy, by education, 1981–2000," *Health affairs*, 27, 350–360. (Cited on page 1.)

MURPHY, K. M. AND R. H. TOPEL (2006): "The Value of Health and Longevity," *Journal of Political Economy*, 114, 871–904. (Cited on page 22.)

NAKAJIMA, M. AND I. A. TELYUKOVA (2023): "Medical Expenses and Saving in Retirement: The Case of U.S. and Sweden," Mimeo. (Cited on page 5.)

NOVOSAD, P., C. RAFKIN, AND S. ASHER (2022): "Mortality Change among Less Educated Americans," *American Economic Journal: Applied Economics*, 14, 1–34. (Cited on page 4.)

OZKAN, S. (2023): "Preventive vs. Curative Medicine: A Macroeconomic Analysis of Health Care over the Life Cycle," St Louis FED Working Paper # 2023-025B. (Cited on page 5.)

PIJOAN-MAS, J. AND J.-V. RÍOS-RULL (2014): "Heterogeneity in expected longevities," *Demography*, 51, 2075–2102. (Cited on pages 1 and 9.)

POLVINEN, A., R. GOULD, E. LAHELMA, AND P. MARTIKAINEN (2013): "Socioeconomic differences in disability retirement in Finland: the contribution of ill-health, health behaviours and working conditions," *Scandinavian Journal of Public Health*, 41, 470–478. (Cited on page 1.)

PRESTON, S. H. AND I. T. ELO (1995): "Are educational differentials in adult mortality increasing in the United States?" *Journal of Aging and Health*, 7, 476–496. (Cited on page 1.)

ROSEN, S. (1988): "The Value of Changes in Life Expectancy," *Journal of Risk and Uncertainty*, 1, 285–304. (Cited on page 22.)

TAYLOR, D. H., V. HASSELBLAD, S. J. HENLEY, M. J. THUN, AND F. A. SLOAN (2002): "Benefits of Smoking Cessation for Longevity," *American Journal of Public Health*, 92, 990–996. (Cited on page 1.)

WILLIS, R. AND S. ROSEN (1979): "Education and Self-Selection," *Journal of Political Economy*, 87, s7–s36, pt. 2. (Cited on page 14.)

XIANG, S., T. JIA, C. XIE, W. CHENG, B. CHAARANI, T. BANASCHEWSKI, G. J. BARKER, A. L. W. BOKDE, C. BÜCHEL, S. DESRIVIÈRES, H. FLOR, A. GRIGIS, P. A. GOWLAND, R. BRÜHL, J.-L. MARTINOT, M.-L. P. MARTINOT, F. NEES, D. P. ORFANOS, L. POUSTKA,

S. Hohmann, J. H. Fröhner, M. N. Smolka, N. Vaidya, H. Walter, R. Whelan, H. Garavan, G. Schumann, B. J. Sahakian, T. W. Robbins, and J. Feng (2023): "Association between vmPFC Gray Matter Volume and Smoking Initiation in Adolescents," *Nature Communications*, 14, 4684. (Cited on page 14.)

Zaninotto, P., J. Head, and A. Steptoe (2020): "Behavioural risk factors and healthy life expectancy: evidence from two longitudinal studies of ageing in England and the US," *Scientific Reports*, 10, 6955. (Cited on page 1.)

Appendix

Table of Contents

Appendix A: Estimation details of the health model

In this appendix, we explain the details of the estimation of the health model in Section 4. Sections A.1 to A.3 give details on how we model the three different elements of the likelihood function in equation (1), while Section A.4 explains the estimation procedure.

In terms of notation, we break the parameter space into three different types of parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}\}$, where $\boldsymbol{\beta}$ denotes the parameters associated with health transitions, $\boldsymbol{\gamma}$ denotes the parameters associated with health behaviors, and $\boldsymbol{\delta}$ denotes the parameters associated with the probability of having lifestyle y at the initial age.

A.1 Health Behaviors

Given lifestyle $\mathrm{y}_i = y$, age a_{it}, and health h_{it}, we model the probability of an individual i at period t displaying health behavior m ($z_{m,it} = 1$) as the probability that $z^{\star}_{m,it} > 0$, where

$$z^{\star}_{m,it} \equiv \gamma_{1,m,y} + \gamma_{2,m,y} a_{it} + \gamma_{3,m,y} a_{it}^2 + \gamma_{4,m,y} I(h_{it} = h_b) + \epsilon_{m,it}, \quad \epsilon_{m,it} \sim N(0,1)$$

Hence,

$$p(z^{*}_{m,it} > 0 | y, a_{it}, h_{it}) = \Phi(\gamma_{1,m,y} + \gamma_{2,m,y} a_{it} + \gamma_{3,m,y} a_{it}^2 + \gamma_{4,m,y} I(h_{it} = h_b)) \equiv \alpha(y, a_{it}, h_{it}; \boldsymbol{\gamma_{y,m}})$$

where Φ is the cdf of the normal distribution and $\boldsymbol{\gamma_{y,m}} \subset \boldsymbol{\gamma}$ is the relevant subset of parameters. Then, the term $p(z_{m,it}|y, a_{it}, h_{it}; \boldsymbol{\theta}_y)$ in equation (2) is given by,

$$p(z_{m,it}|y, a_{it}, h_{it}; \boldsymbol{\gamma_{y,m}}) = \alpha(y, a_{it}, h_{it}; \boldsymbol{\gamma_{y,m}})^{z_{m,it}} \Big(1 - \alpha(y, a_{it}, h_{it}; \boldsymbol{\gamma_{y,m}})\Big)^{1-z_{m,it}}$$

A.2 Health Dynamics

We model the term $p(h_{it}|\mathrm{s}_i, \mathrm{e}_i, y, a_{i,t-1}, h_{i,t-1}; \boldsymbol{\theta}_y)$ in equation (3) reflecting health dynamics of an individual i at period t as a nested probit model. In the first nest, individuals are exposed to a survival/mortality shock. Then, conditional on surviving, individuals suffer a good/bad health shock. We thus partition the set of parameters health transitions $\boldsymbol{\beta} = \{\boldsymbol{\beta}^s, \boldsymbol{\beta}^h\}$ where $\boldsymbol{\beta}^s$ denotes parameters driving survival probabilities and $\boldsymbol{\beta}^h$ drive health transition probabilities given survival.

In the first nest, an individual i with gender $\mathrm{s}_i = s$, education $\mathrm{e}_i = e$, lifestyle type $\mathrm{y}_i = y$, health $h_{it} = h$, and age a_{it} at time t will survive ($h_{it} \neq h_d$) if $h^{*}_{s,it} > 0$:

$$h^{\star}_{s,it} = \beta^s_{1,h,s,e,y} + \beta^s_{2,h,s,e,y}\, a_{it} + \epsilon_{s,it}, \quad \epsilon_{s,it} \sim N(0,1)$$

In the second nest, conditional on survival, an individual can transition into good or bad health.

An individual will transition to good health ($h_{it} = h_g$) if $h^*_{g,it} > 0$:

$$h^{\star}_{g,it} = \beta^h_{1,h,s,e,y} + \beta^h_{2,h,s,e,y}\, a_{it} + \epsilon_{g,it}, \quad \epsilon_{g,it} \sim N(0,1)$$

Therefore we can write each probability $p(h_{it}|\mathrm{s}_i, \mathrm{e}_i, y, a_{i,t-1}, h_{i,t-1}; \boldsymbol{\theta}_y)$ in equation (3) as:

$$\begin{aligned}
p(h_{it}|\mathrm{s}_i, \mathrm{e}_i, y, a_{i,t-1}, h_{i,t-1}; \boldsymbol{\theta}_y) \quad &= \quad \Phi\left(f(\mathrm{s}_i, \mathrm{e}_i, y, a_{it}, h_{it}; \boldsymbol{\beta}^h)\right)^{I(h_{i,t+1}=h_g)} \\
&\times \quad \left[1 - \Phi\left(f(\mathrm{s}_i, \mathrm{e}_i, y, a_{it}, h_{it}; \boldsymbol{\beta}^h)\right)\right]^{I(h_{i,t+1}=h_b)} \\
&\times \quad \Phi\Big(f(\mathrm{s}_i, \mathrm{e}_i, y, a_{it}, h_{it}; \boldsymbol{\beta}^s)\Big)^{I(h_{i,t+1}\neq h_d)} \\
&\times \quad \Big[1 - \Phi\Big(f(\mathrm{s}_i, \mathrm{e}_i, y, a_{it}, h_{it}; \boldsymbol{\beta}^s)\Big)\Big]^{I(h_{i,t+1}=h_d)}
\end{aligned}$$

where $f(\mathrm{s}, \mathrm{e}, y, a, h; \boldsymbol{\beta}) = \beta_{1,h,s,e,y} + \beta_{2,h,s,e,y}\, a$.

A.3 Mixture Weights

As discussed in Section 3.3, the mixture weights $p\,(\mathrm{y}|\mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, a_{i0}, h_{i0}; \boldsymbol{\beta}, \boldsymbol{\delta})$ differ across entry age a_{i0} given the differential health transitions and mortality conditional on type y. The way they do is determined by the model of health dynamics in equation (3). In particular, let's write,

$$p\,(\mathrm{y}|\mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, a_{i0}, h_{i0}; \boldsymbol{\beta}, \boldsymbol{\delta}) = \frac{p\,(\mathrm{y}, h_{i0}|\mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, a_{i0}; \boldsymbol{\beta}, \boldsymbol{\delta})}{\sum_{y \in N_y} p\,(y, h_{i0}|\mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, a_{i0}; \boldsymbol{\beta}, \boldsymbol{\delta})} \tag{A.1}$$

The joint probability $p\,(\mathrm{y}, h_{i0}|\mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, a_{i0}; \boldsymbol{\beta}, \boldsymbol{\delta})$ can be decomposed as,

$$\begin{aligned}
p\,(\mathrm{y}, h_{i0} \mid \mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, a_{i0}; \boldsymbol{\beta}, \boldsymbol{\delta}) &= \sum_{h_{i(-1)} \in H} p\left(\mathrm{y}, h_{i0} \mid \mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, a_{i0}, h_{i(-1)}; \boldsymbol{\beta}, \boldsymbol{\delta}\right) p\left(h_{i(-1)} \mid \mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, a_{i0}; \boldsymbol{\beta}, \boldsymbol{\delta}\right) \\
&= \sum_{h_{i(-1)} \in H} p\left(h_{i0} \mid h_{i(-1)}, \mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, \mathrm{y}, a_{i0}; \boldsymbol{\beta}\right) p\left(\mathrm{y}, h_{i(-1)} \mid \mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, a_{i0}; \boldsymbol{\beta}, \boldsymbol{\delta}\right),
\end{aligned} \tag{A.2}$$

where we integrate over all possible health states in the period before the individual enters our sample ($h_{i(-1)}$). The first term in the right-hand side of equation (A.2) describes the health dynamics, and it has been discussed in Section 3.2, see equation (3). The second term on the right-hand side is the same as the left-hand side, just one period before. Hence, we can use equation (A.2) recursively up to some common initial age, which we choose to be age 25. This gives us $p\,(\mathrm{y}, h_{i,25}|\mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, 25; \boldsymbol{\delta})$, the joint probability of an individual i in cohort c_i, of gender s_i and education e_i of being of health $h_{i,25}$ and type y at age 25. We decompose this probability in two pieces:

$$p\,(\mathrm{y}, h_{i,25}|\mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, 25; \boldsymbol{\delta}) = p\,(\mathrm{y}|\mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, 25, h_{i,25}; \boldsymbol{\delta})\, p\,(h_{i,25}|\mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, 25)$$

The second term in the right-hand side describes the share of individuals of age 25 of given cohort c_i, gender s_i, and education e_i that have health $h_{i,25}$. This can be measured directly in the PSID for many but not all cohorts. We thus assume that conditional on education and gender, the probability of good and bad health at age 25 does not change across cohorts and we set it equal to the one that we observe in the PSID.[17] The first term in the right-hand side describes the fraction of individuals of age 25, cohort c_i, gender s_i, education e_i, and health $h_{i,25}$ who are of type y_i. This is the probability that we have to estimate, and we model it through a multinomial probit.

In order to compute $p(\mathrm{y}|\mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, 25, h_{i,25}; \boldsymbol{\delta}, \boldsymbol{\beta})$ we need to specify the probability of the lifestyle at the initial age. For this purpose we define:

$$
\begin{aligned}
y^*_{1,i} &= \delta^1_{0,s,e} + \delta^1_{1,s,e}\mathbb{1}_{h=h_g} + \sum_c \delta^1_{2,s,e,c}\mathbb{1}_{\mathrm{c}_i=c} + \epsilon_{1,i} \\
&\vdots \\
y^*_{N_y,i} &= \delta^{N_y}_{0,s,e} + \delta^{N_y}_{1,s,e}\mathbb{1}_{h=h_g} + \sum_c \delta^{N_y}_{2,s,e,c}\mathbb{1}_{\mathrm{c}_i=c} + \epsilon_{N_y,i}
\end{aligned}
$$

where $\epsilon_{1,i}, ..., \epsilon_{N_y,i}$ are independently normally distributed across lifestyles with mean 0 and variance 1.

$$
y_i = \begin{cases} 1 \text{ if } y^*_{1,i} > y^*_{2,i}, \dots, y^*_{N_y,i} \\ 2 \text{ if } y^*_{2,i} > y^*_{1,i}, \dots, y^*_{N_y,i} \\ \vdots \\ N_y \text{ otherwise} \end{cases}
$$

A.4 Gibbs sampler

We can write the complete data likelihood as:

$$
\begin{aligned}
p\left(\mathbf{y}, \mathbf{h}^T, \mathbf{z}^T|\mathbf{c}, \mathbf{s}, \mathbf{e}, \boldsymbol{a}_0, \boldsymbol{h}_0; \boldsymbol{\theta}\right) = \prod_{i=1}^{N}\prod_{y=1}^{N_y} \Big[& p\left(\mathbf{z}_i^T|y, a_{i0}, \mathbf{h}_i^T; \boldsymbol{\gamma}_y\right) \\
& p\left(\mathbf{h}_i^T|\mathrm{s}_i, \mathrm{e}_i, y, a_{i0}, h_{i0}; \boldsymbol{\beta}_y\right) \\
& p\left(y|\mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, a_{i0}, h_{i0}; \boldsymbol{\beta}_y, \boldsymbol{\delta}_y\right) \Big]^{\mathbb{1}_{\mathrm{y}_i=y}}
\end{aligned}
$$

where y_i denotes the lifestyle assigned to individual i and $\boldsymbol{h}^T$, $\boldsymbol{z}^T$ and $\mathbf{y}$ denote health status, health behaviors, and assigned lifestyles for every individual in the sample, respectively. We can now write

[17] The probability of being in good health at age 25 varies between 77% for dropout females to 98% for male college graduates.

the joint posterior distribution as:

$$
\begin{aligned}
p\left(\boldsymbol{\theta}|\mathbf{c},\mathbf{s},\mathbf{e},\mathbf{y},\boldsymbol{a}_{i0},\mathbf{h}_0,\mathbf{h}^T,\mathbf{z}^T\right) \propto \prod_{i=1}^{N}\prod_{y=1}^{N_y} \Big[& p\left(\mathbf{z}_i^T|a_{i0},y,h_{i0},\mathbf{h}_i^T;\boldsymbol{\gamma}_y\right) \\
& p\left(\mathbf{h}_i^T|\mathrm{s}_i,\mathrm{e}_i,y,a_{i0},h_{i0};\boldsymbol{\beta}_y\right) \\
& p\left(y|\mathrm{c}_i,\mathrm{s}_i,\mathrm{e}_i,a_{i0},h_{i0};\boldsymbol{\beta}_y,\boldsymbol{\delta}_y\right)\Big]^{\mathbb{1}_{\mathrm{y}_i=y}} \; p(\boldsymbol{\theta}).
\end{aligned}
$$

Block 1: transition parameters The first block is given by

$$
p\left(\boldsymbol{\beta}|\mathbf{y},\mathbf{c},\mathbf{s},\mathbf{e},\boldsymbol{a}_0,\boldsymbol{h}_0,\mathbf{h}^T,\mathbf{z}^T;\boldsymbol{\delta},\boldsymbol{\gamma}\right) \propto \prod_{y=1}^{N_y}\prod_{i:\mathrm{y}_i=y} p\left(\mathbf{h}_i^T|\mathrm{s}_i,\mathrm{e}_i,y,a_{i0},h_{i0};\boldsymbol{\beta}_y\right) p\left(y|\mathrm{c}_i,\mathrm{s}_i,\mathrm{e}_i,a_{i0},h_{i0};\boldsymbol{\beta}_y,\boldsymbol{\delta}_y\right) p(\boldsymbol{\beta}_y)
$$

Using flat prior in $\boldsymbol{\beta}$, we sample from the posterior distribution using a Metropolis algorithm.

Block 2: health behaviors parameters The second block is given by:

$$
p\left(\boldsymbol{\gamma}|\mathbf{c},\mathbf{s},\mathbf{e},\mathbf{y},\boldsymbol{a}_0,\mathbf{h}^T,\mathbf{z}^T;\boldsymbol{\delta},\boldsymbol{\beta}\right) \propto \prod_{y=1}^{N_y}\prod_{i:\mathrm{y}_i=y}\prod_{t=0}^{T_i} p\left(\mathbf{z}_{it}|y,a_{i0},h_{it};\boldsymbol{\gamma}_y\right) p(\boldsymbol{\gamma}_y)
$$

Using data augmentation for z_m^*, the posterior of γ is normally distributed (probit model).

Block 3: mixture weights parameters The third blocks is given by:

$$
p\left(\boldsymbol{\delta}|\mathbf{c},\mathbf{s},\mathbf{e},\mathbf{y},\boldsymbol{a}_0,\mathbf{h}^T,\mathbf{z}^T;\boldsymbol{\gamma},\boldsymbol{\beta}\right) \propto \prod_{y=1}^{N_y}\prod_{i:\mathrm{y}_i=y} p\left(y|\mathrm{c}_i,\mathrm{s}_i,\mathrm{e}_i,a_{i0},h_{i0};\boldsymbol{\beta}_y,\boldsymbol{\delta}_y\right) p(\boldsymbol{\delta}_y)
$$

we sample δ using a Metropolis algorithm with flat priors in $p(\boldsymbol{\delta})$

Block 4: latent lifestyles The posterior distribution of the latent lifestyles with flat priors is given by:

$$
\begin{aligned}
p\left(\mathrm{y}_i|\mathrm{c}_i,\mathrm{s}_i,\mathrm{e}_i,a_{i0},h_{i0},\mathbf{h}_i^T,\mathbf{z}_i^T;\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\delta}\right) & \propto p\left(\mathbf{z}_i^T|\mathrm{y}_i,a_{i0},\mathbf{h}_i^T;\boldsymbol{\gamma}_y\right) \\
& \times p\left(\mathbf{h}_i^T|\mathrm{s}_i,\mathrm{e}_i,\mathrm{y}_i,a_{i0},h_{i0};\boldsymbol{\beta}_y\right) \\
& \times p\left(\mathrm{y}_i|\mathrm{c}_i,\mathrm{s}_i,\mathrm{e}_i,a_{i0},h_{i0};\boldsymbol{\beta}_y,\boldsymbol{\delta}_y\right)
\end{aligned}
$$

Thus, we can directly sample the latent lifestyle from:

$$p\left(\mathrm{y}_i = y|\mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, a_{i0}, h_{i0}, \mathbf{h}^T, \mathbf{z}^T; \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\delta}\right) = \frac{p\left(\mathbf{z}_i^T|y, a_{i0}, h_{i0}, \mathbf{h}_i^T; \boldsymbol{\gamma}_y\right) p\left(\mathbf{h}_i^T|\mathrm{s}_i, \mathrm{e}_i, y, a_{i0}, h_{i0}; \boldsymbol{\beta}_y\right) p\left(y|\mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, a_{i0}, h_{i0}; \boldsymbol{\beta}_y, \boldsymbol{\delta}_y\right)}{\sum_{\tilde{y}=1}^{N_y} p\left(\mathbf{z}_i^T|\tilde{y}, a_{i0}, h_{i0}, \mathbf{h}_i^T; \boldsymbol{\gamma}_y\right) p\left(\mathbf{h}_i^T|\mathrm{s}_i, \mathrm{e}_i, \tilde{y}, a_{i0}, h_{i0}; \boldsymbol{\beta}_y\right) p\left(\tilde{y}|\mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, a_{i0}, h_{i0}; \boldsymbol{\beta}_y, \boldsymbol{\delta}_y\right)}$$

A.5 Obtaining the wealth distribution conditional on age, education, and type

We model the observed wealth distribution as a mixture model. In order to separate the mass point at zero wealth and the distribution of wealth conditional on positive wealth we proceed in two stages. First, we write the distribution of (positive) wealth conditional on observables as:

$$p(w_{it}|\mathrm{e}_i, a_{it}, z_i^T, h_i^T, w_{it} > 0) = \sum_{y \in Y} p(w_{it}|y, \mathrm{e}_i, a_{it}, z_i^T, h_i^T, w_{it} > 0) p(y|\mathrm{e}_i, z_i^T, h_i^T),$$

The first term in the right-hand side is the conditional probability of observing wealth w_{it}. We assume that wealth conditional on age a, education e, and type y is lognormally distributed, that is, $\log p(w|y, e, a, w > 0) \sim N(\mu^1(y, e, a), \sigma^1(y, e))$. This implies that we are imposing that given a, e, and y, wealth is independent from h_i^T and z_i^T.

The second term on the right-hand side gives the conditional distribution of types, which we have estimated above, see Section 3.3. Hence, we only need to estimate $\mu^1(y, e, a)$ and $\sigma^1(y, e)$ for the sample of male individuals with positive asset holdings.[18]

Second, we can similarly write the probability of reporting zero (or negative) assets conditional on observables as

$$p(w_{it} = 0|\mathrm{e}_i, a_{it}, z_i^T, h_i^T) = \sum_{y \in Y} p(w_{it} = 0|y, \mathrm{e}_i, a_{it}, z_i^T, h_i^T) p(y|\mathrm{e}_i, z_i^T, h_i^T),$$

where the first term on the right-hand side is modeled as a probit, that is, it is given by $\Phi\left(w^*\left(y, \mathrm{e}_i, a_{it}\right)\right)$, where the threshold $w^*\left(y, \mathrm{e}_i, a_{it}\right)$ is modeled as a flexible low order polynomial on age. As above, we assume that this probability does not depend on h_i^T or z_i^T.

[18]We model $\mu^1(y, e, a)$ as a flexible low-order polynomial on age and $\sigma^1(y, e)$ non-parametrically.

Appendix B: More than 2 groups

In our main specification, we estimate the econometric model with only two latent types, which we label as *protective* and *detrimental*. In this estimation, the different composition of types across education groups explains 40% of the education gradient of life expectancy. In this Appendix, we extend the model by estimating it with three and four groups.

When estimating it with three groups, the estimation splits the *protective* lifestyle into two subgroups. One of these groups includes individuals who drink alcohol and are less likely to engage in health-enhancing behaviors like taking a flu shot or doing cholesterol or cancer tests, see Figure B.1. The life expectancy of the *protective* group drinking alcohol is one year lower. The *detrimental* lifestyle group remains intact but with even lower life expectancy compared to the two-group model. In this specification, lifestyle differences explain approximately 50% of the education gradient of life expectancy, see Table B.1.

When we estimate the model with four groups, the results become more complex. The *detrimental* lifestyle group still remains stable. However, the *protective* group is further split into three subgroups: (i) individuals who neither smoke nor drink, (ii) individuals who drink but engage in all other preventive activities at levels similar to the healthiest group, and (iii) individuals who drink and engage in fewer preventive activities than the healthiest group, see Figure B.2. Despite this increased granularity, lifestyle differences still explain around 50% of the life expectancy differences across education groups, see Table B.2.

Table B.1: Life expectancies at age 50: males born in 1970s (3 groups)

	HSD		HSG		CG		Δ_eLE (CG-HSD)		
	%	LE	%	LE	%	LE	Data	(a)	(b)
All	100.0	24.6	100.0	27.9	100.0	32.8	8.3	4.1	4.2
Group 1	21.7	29.6	41.3	30.6	67.6	33.7	4.1		
Group 2	26.8	28.3	30.0	29.7	26.9	32.5	4.2		
Group 3	51.5	20.5	28.7	22.2	5.6	23.7	3.2		

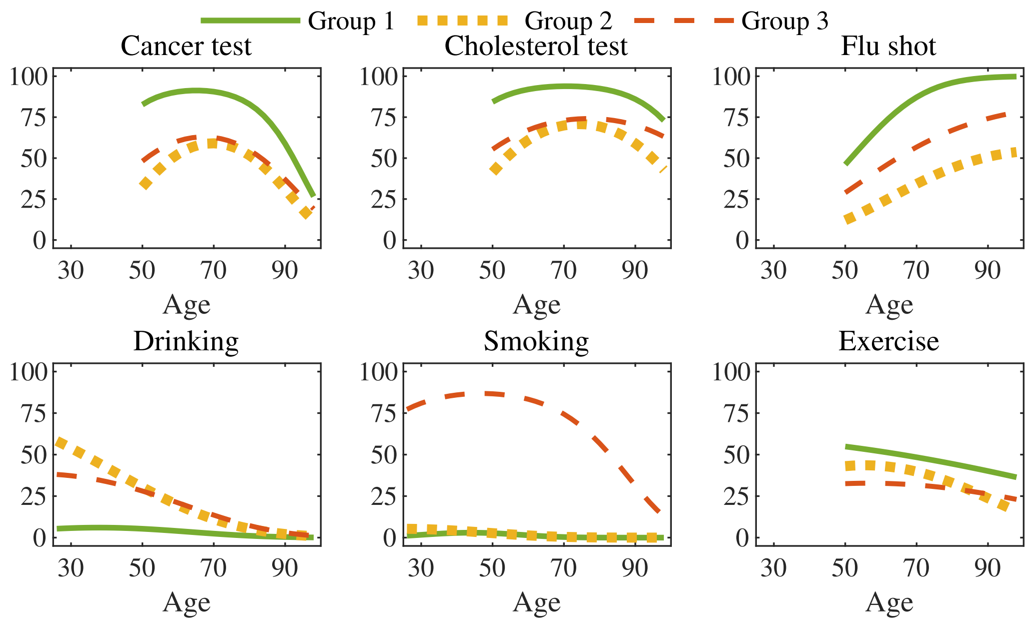
Notes: This table reports the share of male individuals of each lifestyle and the life expectancy (LE) at age 50 for different population groups. Column (a) corresponds to the counterfactual LE gradient when the distribution of behavior types among high school dropouts is the same as for college-educated individuals. Column (b) is the difference between the actual LE gradient and the counterfactual in column (a), that is, it corresponds to the gradient explained by difference in lifestyles across education groups for given health dynamics.

Table B.2: Life expectancies at age 50: males born in 1970s (4 groups)

	HSD		HSG		CG		Δ_eLE (CG-HSD)		
	%	LE	%	LE	%	LE	Data	(a)	(b)
All	100.0	25.1	100	28.4	100.0	33.0	7.8	4.0	3.9
Group 1	19.6	28.9	38.3	30.5	64.0	33.4	4.5		
Group 2	16.1	27.9	13.1	29.0	8.0	32.5	4.6		
Group 3	17.6	31.0	24.7	30.9	24.2	33.4	2.4		
Group 4	46.8	20.4	23.9	22.3	3.8	23.9	3.5		

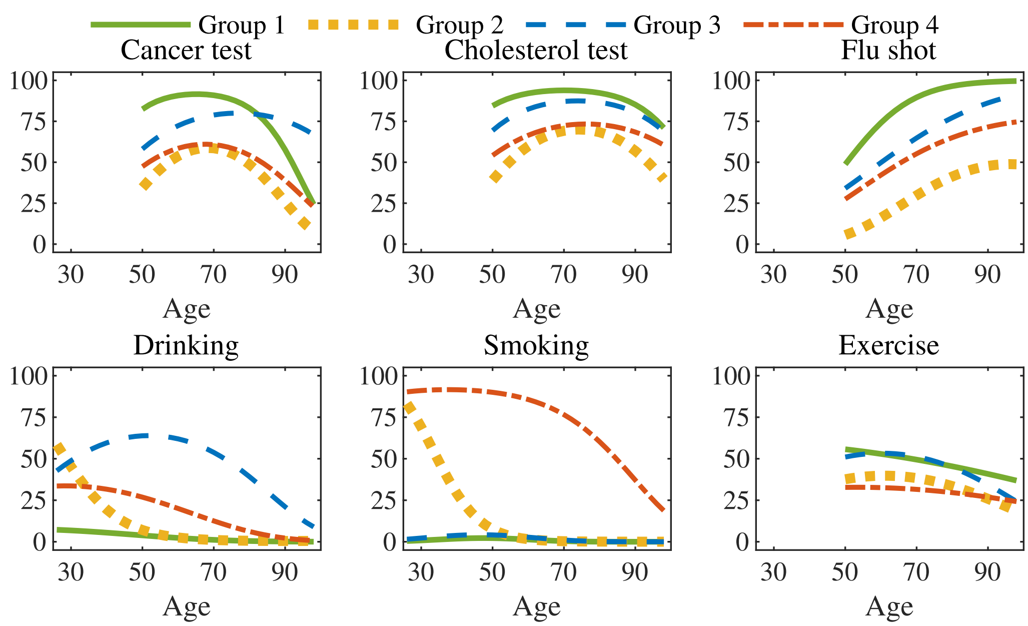
Notes: This table reports the share of male individuals of each lifestyle and the life expectancy (LE) at age 50 for different population groups. Column (a) corresponds to the counterfactual LE gradient when the distribution of behavior types among high school dropouts is the same as for college-educated individuals. Column (b) is the difference between the actual LE gradient and the counterfactual in column (a), that is, it corresponds to the gradient explained by difference in lifestyles across education groups for given health dynamics.

Figure B.1: Health habits and health behavior types (3 groups)



Notes: Estimation results. Probability of engaging in each health behavior by age and type, for male individuals in good health.

Figure B.2: Health habits and health behavior types (4 groups)



Notes: Estimation results. Probability of engaging in each health behavior by age and type, for male individuals in good health.

Appendix C: Theoretical results of the first stage model

In this appendix, we obtain two theoretical results for the first-stage model in Section 5.1. The results are proved for a slightly simpler version of the model with only two education categories (without loss of generality, we drop e = HSG by imposing $\tau_{\text{HSG}} = \infty$ for all individuals). First, in Lemma 1 and 2 we characterize the choice of $(\mathrm{e}, \mathrm{y}) \in \{(\text{CG,PRO}), (\text{CG,DET}), (\text{HSD,PRO}), (\text{HSD,DET})\}$ described in Figure 3, and second, in Proposition 1 we characterize the pattern of selection of individuals into the different (e, y) groups. All our discussion below focuses on the cases where the choices of education e and lifestyle y are independent or complements, see Definition 1.

Definition 1. *We say that the choices of education e and lifestyle y are independent from each other whenever* $V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} = V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$ *and complements whenever* $V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} > V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$.

Individuals in the model are heterogeneous and characterized by a pair $(\tau_{\text{CG}}, \tau_{\text{PRO}}) \in \mathbb{R}^2$. Given the objective function (4), we can characterize the choice of education for an individual with taste shocks $(\tau_{\text{CG}}, \tau_{\text{PRO}})$ with a threshold function $\tau^*_{\text{CG}}(\tau_{\text{PRO}})$ making the individual indifferent between going to college or not such that the individual will choose e = CG whenever $\tau_{\text{CG}} < \tau^*_{\text{CG}}(\tau_{\text{PRO}})$ and e = HSD with the reverse inequality. Lemma 1 below characterizes this threshold function, which can be seen in Panel (B) of Figure 3 as the piecewise linear function that separates the (CG, PRO) and (CG, DET) regions from the (HSD, PRO) and (HSD, DET) regions. Symmetrically, we can characterize the choice of lifestyle with a threshold function $\tau^*_{\text{PRO}}(\tau_{\text{CG}})$ such that the individual will choose y = PRO whenever $\tau_{\text{PRO}} < \tau^*_{\text{PRO}}(\tau_{\text{CG}})$ and y = DET with the reverse inequality. Lemma 2 below characterizes this threshold function, which can be seen in Panel (B) of Figure 3 as the piecewise linear function that separates the (CG, PRO) and (HSD, PRO) regions from the (CG, DET) and (HSD, DET) regions. Hence, the crossing of these two threshold functions fully characterizes the joint decision on (e, y).

Lemma 1. *Let $\tau^*_{\text{CG}}(\tau_{\text{PRO}})$ be the threshold function making an individual with taste shock τ_{PRO} indifferent between going to college or not such that if $\tau_{\text{CG}} < \tau^*_{\text{CG}}(\tau_{\text{PRO}})$ the individual goes to college. This function is continuous, piecewise linear, and non-increasing in τ_{PRO}. Furthermore,*

a) For the case where education and lifestyle are complements, the function is defined by

$$\tau^*_{\text{CG}}(\tau_{\text{PRO}}) = \begin{cases} V^{\text{CG,PRO}} - V^{\text{HSD,PRO}} & \text{if } \tau_{\text{PRO}} < V^{\text{HSD,PRO}} - V^{\text{HSD,DET}} \\ V^{\text{CG,PRO}} - V^{\text{HSD,DET}} - \tau_{\text{PRO}} & \text{if } \tau_{\text{PRO}} \in [V^{\text{HSD,PRO}} - V^{\text{HSD,DET}}, V^{\text{CG,PRO}} - V^{\text{CG,DET}}] \\ V^{\text{CG,DET}} - V^{\text{HSD,DET}} & \text{if } \tau_{\text{PRO}} > V^{\text{CG,PRO}} - V^{\text{CG,DET}} \end{cases}$$

(b) For the case where education and lifestyle are independent, the function is just defined by

$$\tau^*_{\text{CG}}(\tau_{\text{PRO}}) = V^{\text{CG,PRO}} - V^{\text{HSD,PRO}} = V^{\text{CG,DET}} - V^{\text{HSD,DET}}$$

Proof. An individual with taste shocks $(\tau_{\text{CG}}, \tau_{\text{PRO}})$ chooses to go to college if

$$\max\{V_0^{\text{CG,PRO}} - \tau_{\text{PRO}}, V_0^{\text{CG,DET}}\} - \tau_{\text{CG}} > \max\{V_0^{\text{HSD,PRO}} - \tau_{\text{PRO}}, V_0^{\text{HSD,DET}}\} \tag{C.1}$$

that is, the education decision depends on both τ_{CG} and τ_{PRO}. Let's define the threshold function $\tau^*_{\text{CG}}(\tau_{\text{PRO}})$ making an individual with taste shocks $(\tau_{\text{CG}}, \tau_{\text{PRO}})$ indifferent between going to college or not. To characterize this function, we distinguish three cases according to the value of τ_{PRO}:

Case 1: $\tau_{\text{PRO}} < V^{\text{HSD,PRO}} - V^{\text{HSD,DET}}$

This inequality implies the choice of y = PRO whenever e = HSD. Furthermore, as long as $V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} \geq V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$ (complementary or independent choices) we also have that $\tau_{\text{PRO}} < V^{\text{CG,PRO}} - V^{\text{CG,DET}}$, which also implies the choice of y = PRO whenever e = CG. Therefore, case 1 guarantees the choice of y = PRO for all education levels (in both the complementary and independent choices). In this case, inequality (C.1) states that the individual will choose e = CG iff $\tau_{\text{CG}} < V^{\text{CG,PRO}} - V^{\text{HSD,PRO}}$.

Case 2: $\tau_{\text{PRO}} > V^{\text{CG,PRO}} - V^{\text{CG,DET}}$

This inequality implies the choice of y = DET whenever e = CG. Furthermore, as long as $V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} \geq V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$ (complementary or independent choices) we have that $\tau_{\text{PRO}} > V^{\text{CG,HSD}} - V^{\text{CG,HSD}}$, which also implies the choice of y = DET whenever e = HSD. Therefore, case 2 guarantees the choice of y = DET for all education levels (in both the complementary and independent choices). In this case, inequality (C.1) states that individual will choose e = CG iff $\tau_{\text{CG}} < V^{\text{CG,DET}} - V^{\text{HSD,DET}}$.

Case 3: $V^{\text{CG,PRO}} - V^{\text{CG,DET}} > \tau_{\text{PRO}} > V^{\text{HSD,PRO}} - V^{\text{HSD,DET}}$

This intermediate case for τ_{PRO} is only possible in the case of complementary choices, as in the case of independent choices the interval has length zero. In this case, inequality (C.1) states that the individual will choose e = CG iff $\tau_{\text{CG}} < V^{\text{CG,PRO}} - V^{\text{HSD,DET}} - \tau_{\text{PRO}}$. That is, in this third case, the threshold $\tau^*_{\text{CG}}(\tau_{\text{PRO}})$ making individuals indifferent between e = CG and e = HSD is strictly decreasing in τ_{PRO}. □

Lemma 2. *Let $\tau^*_{\text{PRO}}(\tau_{\text{CG}})$ be the threshold function making an individual with taste shock τ_{CG} indifferent between protective and detrimental lifestyles such that if $\tau_{\text{PRO}} < \tau^*_{\text{PRO}}(\tau_{\text{CG}})$ the individual chooses a protective lifestyle. This function is continuous, piecewise linear, and non-increasing in τ_{CG}. Furthermore,*

a) For the case where education and lifestyle are complements, the function is defined by

$$\tau^*_{\text{PRO}}(\tau_{\text{CG}}) = \begin{cases} V^{\text{CG,PRO}} - V^{\text{CG,DET}} & \textit{if} \;\; \tau_{\text{CG}} < V^{\text{CG,DET}} - V^{\text{HSD,DET}} \\ V^{\text{CG,PRO}} - V^{\text{HSD,DET}} - \tau_{\text{PRO}} & \textit{if} \;\; \tau_{\text{CG}} \in [V^{\text{CG,DET}} - V^{\text{HSD,DET}}, V^{\text{CG,PRO}} - V^{\text{HSD,PRO}}] \\ V^{\text{HSD,PRO}} - V^{\text{HSD,DET}} & \textit{if} \;\; \tau_{\text{CG}} > V^{\text{CG,PRO}} - V^{\text{HSD,PRO}} \end{cases}$$

(b) For the case where education and lifestyle are independent, the function is just defined by

$$\tau^*_{\text{PRO}}(\tau_{\text{CG}}) = V^{\text{CG,PRO}} - V^{\text{CG,DET}} = V^{\text{HSD,PRO}} - V^{\text{HSD,DET}}$$

Proof. The proof is symmetric to Lemma 1 □

Proposition 1. *Let $F(\tau_{\text{PRO}})$ be the CDF of τ_{PRO}, $f(\tau_{\text{PRO}})$ its associated PDF, and $F(\tau_{\text{PRO}}|e)$ be the CDF of τ_{PRO} conditional on a particular educational choice e. Then, in the case where the education and lifestyle choices are complements we have that $F(\tau_{\text{PRO}}|\text{HSD})$ first-order stochastically dominates $F(\tau_{\text{PRO}})$, which in turn first-order stochastically dominates $F(\tau_{\text{PRO}}|\text{CG})$. In the case of independent choices the three distributions are identical.*

Proof. An individual with taste shocks $(\tau_{\text{CG}}, \tau_{\text{PRO}})$ goes to college whenever $\tau_{\text{CG}} < \tau^*_{\text{CG}}(\tau_{\text{PRO}})$. Let us start with the case in which the choices of e and y are complementary. Lemma 1 states that $\tau^*_{\text{CG}}(\tau_{\text{PRO}})$ is non-increasing in τ_{PRO}, and strictly decreasing in τ_{PRO} within an interval of positive length. That is, if $\tau^a_{\text{PRO}} > \tau^b_{\text{PRO}}$ then $\tau^*_{\text{CG}}(\tau^a_{\text{PRO}}) \le \tau^*_{\text{CG}}(\tau^b_{\text{PRO}})$, with strict inequality for some pairs $\left(\tau^a_{\text{PRO}}, \tau^b_{\text{PRO}}\right)$. Therefore $\Pr(\text{CG}|\tau_{\text{PRO}}) \equiv \Pr(\tau_{\text{CG}} \le \tau^*_{\text{CG}}(\tau_{\text{PRO}}))$ is weakly decreasing with τ_{PRO}, and so is $\Pr(\text{CG}|\tau \le \tau_{\text{PRO}}) = \int^{\tau_{\text{PRO}}} \Pr(\text{CG}|\tau) \frac{f(\tau)}{F(\tau_{\text{PRO}})} d\tau$.[19] Using Bayes theorem, we have that

$$F(\tau_{\text{PRO}}|\text{CG}) = \Pr(\tau \le \tau_{\text{PRO}}|\text{CG}) = \frac{\Pr(\text{CG}|\tau \le \tau_{\text{PRO}})\Pr(\tau \le \tau_{\text{PRO}})}{\Pr(\text{CG})} \ge \Pr(\tau \le \tau_{\text{PRO}}) = F(\tau_{\text{PRO}})$$

where the last inequality comes from the fact that, since $\Pr(\text{CG}|\tau \le \tau_{\text{PRO}})$ is non-increasing in τ_{PRO} and strictly decreasing for some τ_{PRO}, it must be that

$$\frac{\Pr(\text{CG}|\tau \le \tau_{\text{PRO}})}{\Pr(\text{CG})} \ge 1$$

For high school dropouts we have equivalently:

$$\begin{aligned} F(\tau_{\text{PRO}}|\text{HSD}) = \Pr(\tau \le \tau_{\text{PRO}}|\text{HSD}) &= \frac{\Pr(\text{HSD}|\tau \le \tau_{\text{PRO}})\Pr(\tau \le \tau_{\text{PRO}})}{\Pr(\text{HSD})} \\ &= \frac{[1 - \Pr(\text{CG}|\tau \le \tau_{\text{PRO}})]\Pr(\tau \le \tau_{\text{PRO}})}{1 - \Pr(\text{CG})} \le \Pr(\tau \le \tau_{\text{PRO}}) = F(\tau_{\text{PRO}}) \end{aligned}$$

thus:

$$F(\tau_{\text{PRO}}|\text{HSD}) \le F(\tau_{\text{PRO}}) \le F(\tau_{\text{PRO}}|\text{CG}) \quad \forall \tau_{\text{PRO}}$$

with strict inequality for some τ_{PRO}.

Finally, note that whenever we are in the case of independent choices, Lemma 1 states that

[19] To see this last part, note that $\frac{\partial \Pr(\text{CG}|\tau \le \tau_{\text{PRO}})}{\partial \tau_{\text{PRO}}} = \Pr(\text{CG}|\tau_{\text{PRO}}) - \Pr(\text{CG}|\tau \le \tau_{\text{PRO}}) < 0$, where the first term in the right-hand side is always smaller than the second one whenever $\Pr(\text{CG}|\tau_{\text{PRO}})$ is decreasing in τ_{PRO}.

$\tau^*_{\text{CG}}(\tau_{\text{PRO}})$ is just a constant independent from τ_{PRO}. Hence, $\Pr(\text{CG}|\tau_{\text{PRO}})$ is also independent from τ_{PRO} and we have $F(\tau_{\text{PRO}}|\text{HSD}) = F(\tau_{\text{PRO}}) = F(\tau_{\text{PRO}}|\text{CG}) \;\; \forall \tau_{\text{PRO}}$. $\square$

Appendix D: First step estimation details

D.1 Income process

The labor income process is modeled as the sum of a deterministic and stochastic component:

$$\begin{aligned}
&\log w_t^{\mathrm{c,e}}(h_t, \xi_t, \epsilon_t) = \omega_t^{\mathrm{c,e}}(h_t) + \xi_t + \epsilon \\
&\xi_{t+1} = \rho_\xi \xi_t + \nu_t, \nu_t \sim N(0, \sigma_\nu^2) \\
&\epsilon \sim N(0, \sigma_\epsilon^2) \\
&\xi_0 \sim (0, \sigma_{\xi,0}^2)
\end{aligned}$$

We propose a Gibbs algorithm to compute the posterior distribution of all parameters using Bayesian methods.

1. Sample parameters of the deterministic component: multivariate normal.
2. Sample persistent shocks: Kalman smoother.
3. Sample persistent component parameters : normal posterior for ρ_{xi} and inverse gamma for σ_ν^2.
4. Sample initial distribution of shocks: Metropolis.
5. Sample variance of the transitory component: inverse gamma.

D.2 Medical shocks

To estimate the mean of health cost distribution $\mu_t^e(h_t)$, we run an OLS regression of log out-of-pocket expenditures in the last two years in HRS on a cubic in age, health, health interacted with age and education interacted with age and individual fixed effects. In order to compute the education fixed effects, we regress the residuals of the previous regression on education dummies. In order to estimate $\sigma_{\zeta,t}^e(h)$, we regress the squared residuals from the previous regression on a cubic in age, health, health interacted with age, education, and education interacted with age.

D.3 Tax system

We follow De Nardi et al. (2024) and set the Medicare and Social Security tax rates to 2.9% and 12.4%, respectively. We use the Social Security rules for 2018, and therefore we set the maximum taxable income for Social Security to w_{ss} =\$113,700.

For the progressive tax labor income tax function, we follow Holter et al. (2019) estimates of the tax progressivity for families without children:

$$T(y) = y - 0.873964 \times y^{1-0.108002}$$

D.4 Value of statistical life

To obtain the model equivalent of the VSL, we proceed as follows. Using the value function expressed in Section 5.2.3 we can obtain the total differential:

$$\frac{\partial V_t^{\text{c,e,y}}(x,h,\xi)}{\partial x} dx + \frac{\partial V_t^{\text{c,e,y}}(x,h,\xi)}{\partial s_t^{\text{e,y}}(h)} ds_t^{\text{e,y}}(h) = 0$$

relating changes in cash-on-hand x and survival probabilities $s_t^{\text{e,y}}(h)$ that leave individuals indifferent. Rearranging we obtain,

$$-\frac{dx}{ds_t^{\text{e,y}}(h)} = \frac{\partial V_t^{\text{c,e,y}}(x,h,\xi)}{\partial s_t^{\text{e,y}}(h)} \left[\frac{\partial V_t^{\text{c,e,y}}(x,h,\xi)}{\partial x}\right]^{-1}$$

Hence, for an individual of type (c,e,y) with state variables (x, h, ξ) at age t to accept an increase in his survival probability in say 1%, he would require $0.01 \times dx/ds_t^{\text{e,y}}(h)$ units of income. Putting 100 identical agents together, we would have one death on average in exchange for $dx/ds_t^{\text{e,y}}(h)$ units of income.

Appendix E: Extra figures and tables

Table E.1: Parameters of the stochastic component of income

	$\sigma^2_{\xi_0}$	ρ_ξ	$\sigma^2_\xi \times 10^{-2}$	σ^2_ϵ
HSD	0.18	0.96	2.37	0.16
HSG	0.16	0.98	2.33	0.14
CG	0.18	0.99	3.76	0.13

Notes: Estimated parameters for the stochastic process of wages.

Figure E.1: Health transitions across education and lifestyles

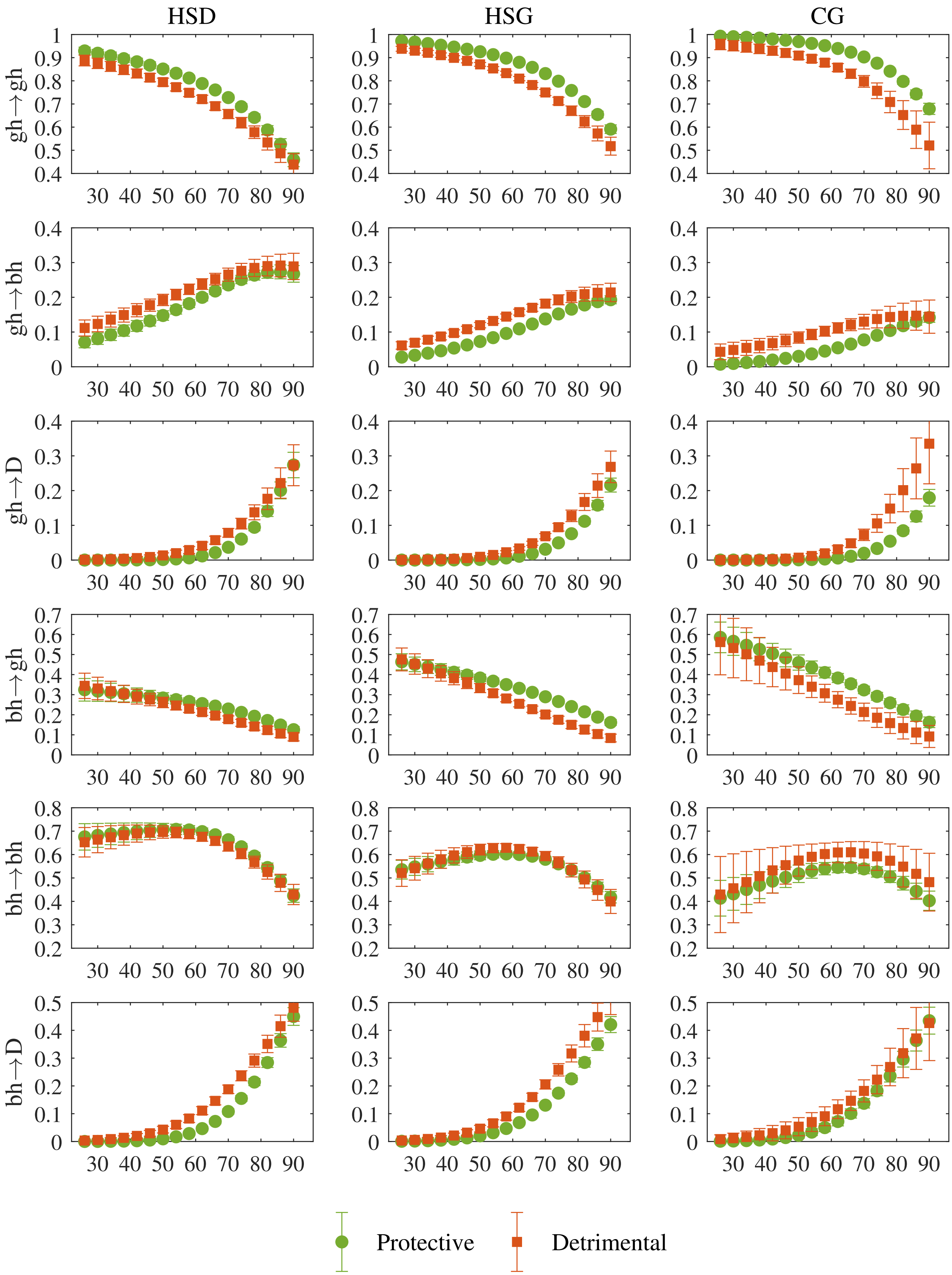
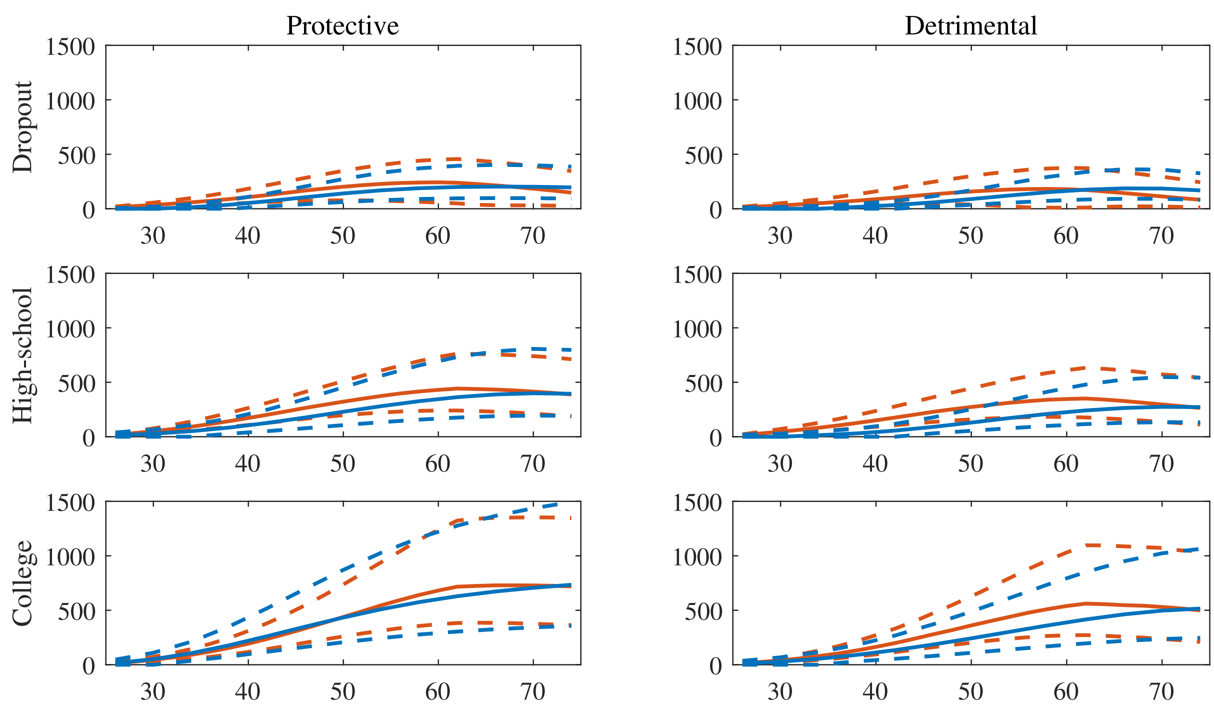
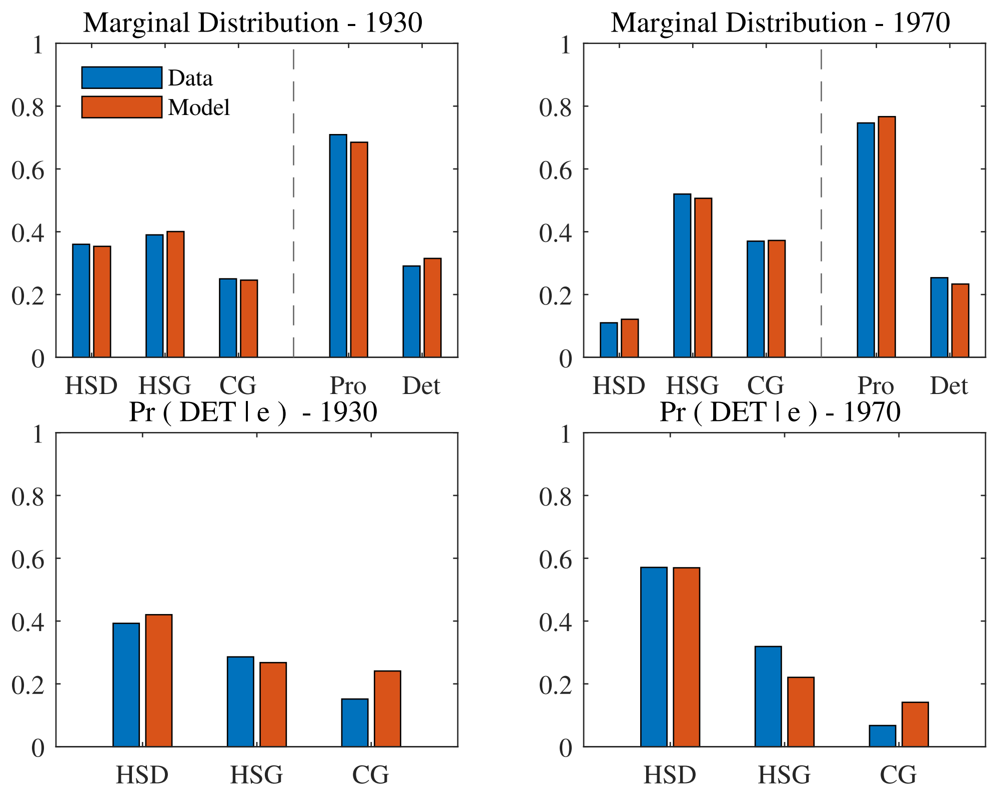


Figure E.2: Model fit: wealth trajectories (changes in perceptions)



Notes: The solid blue lines represent the median wealth in the data, by age in each education and lifestyle group. The solid red lines are the model predictions. The dashed lines represent the 25th and 75th percentiles of wealth for each group, with again blue being data and red model.

Figure E.3: Model Fit: First-Stage (changes in perceptions)



Notes: The top two panels report the marginal distribution of education and lifestyle types for two different cohorts. The bottom panel reports the distribution of lifestyle types conditional on education choices for the same two cohorts.