Education, Lifestyles, and Health Inequality*

Jesús Bueren
EUI

Josep Pijoan-Mas
CEMFI and CEPR

Dante Amengual
CEMFI

March 2025

Abstract

We study the effect of lifestyles on the education gradient of life expectancy. We use panel data on health behavior and health outcomes to estimate latent lifestyle types and their impact on health dynamics. We find that the higher frequency of health-protective lifestyles among the more educated individuals explains almost 1/2 of the education gradient in life expectancy. To understand lifestyle formation, we build a life cycle model where lifestyles and education are jointly chosen early in life. These two investments are complementary because of the more educated's higher income and the higher yield of their health-protective behavior. Importantly, with these complementarities, individuals with lower costs of healthier lifestyles self-select into higher education. Quantitatively, we find the three mechanisms similarly important in explaining the correlation between education and healthy lifestyles. We also find that the increase in the college wage premium over the last decades has widened the education gradient in lifestyles, resulting in a 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 and 60% by the induced changes in the composition of college graduates and high school dropouts.

JEL classification: E21, D15, I12, I14, I24, C38
Keywords: Health Inequality, Healthy Lifestyles, Education, Latent Types

*The authors thank for their valuable feedback Alexander Bick, Raquel Fonseca, Elena Manresa, Pierre-Carl Michaud, Pedro Mira, Mariacristina de Nardi, Florian Pelgrin, Eugenio Zucchelli and the participants from presentations held at Bristol University, Católica Lisbon, CEMFI, CUNEF, EUI, HEC Montreal, Princeton University, St Louis FED, Stony Brook University, UAB, UQAM, the 2024 BSE Summer Forum in Barcelona, and the 2024 EEA Annual Meetings in Rotterdam. Emails: jesus.bueren@eui.eu, pijoan@cemfi.es, amengual@cemfi.es. Josep Pijoan-Mas acknowledges financial support from Grant PID2022-142790NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU, and from the Mobility Program of the MICIU, which has financed the visit to Princeton University where part of this research took place.

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 health behaviors and lifestyle factors.

Health behaviors —like exercise habits, dietary patterns, or substance abuse— 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 and lower risk of mortality (e.g. 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 (e.g. Lantz et al., 1998; Cutler and Lleras-Muney, 2010).

Our paper tackles three main questions. First, we measure the impact of different lifestyles —understood as the collection of different health behaviors— on health dynamics and health inequalities. Second, we study the joint determination of education and lifestyles early in life, the role of selection, and why there is an education gradient in lifestyles. Finally, we quantify the effect of raising labor earnings inequality across education groups on the observed increase in the education gradients of lifestyles and health outcomes, with a particular focus on the life expectancy decline of the less educated.

To address the first question, we use data from the Health and Retirement Study (HRS) and the Panel Study of Income Dynamics (PSID), which contain a rich array of health behavior indicators —including preventive tests, substance abuse, and exercise habits. The use of these data presents three challenges. First, the observed health behaviors are imperfectly correlated across individuals and over time, which suggests the existence of a noisy relationship between individual incentives and different health behaviors. 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 are likely to appear in the long run. Finally, it is unfeasible to consider so many different variables in a dynamic model that tries to understand the determinants of health behaviors. 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 health behavior by assigning individuals to permanent types based on both their health behaviors and their health dynamics. These types, which we call lifestyles, summarize the propensity of individuals to engage in different health behaviors over time as well as the resulting health trajectories. Consequently, our novel econometric procedure allows us to estimate the long-run effect of lifestyles 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 8.5-year life expectancy gap between individuals classified into the *protective* lifestyle and individuals classified into the *detrimental* one. Second, there is a strong correlation between lifestyle and education, with *protective* individuals being more frequent among the more educated. Indeed, the education gradient in health behaviors explains 48% of the education gradient in life expectancy. Third, the life expectancy differences across lifestyle 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 9.4 years among college graduates. This uncovers an important complementarity between education and lifestyle choices, which we discuss below. Finally, we find a larger education gradient in lifestyles for individuals born in more recent cohorts, with the more educated individuals adopting more *protective* behaviors and the less educated ones engaging in more *detrimental* ones.

To understand type formation and the joint determination of education and lifestyles, 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 their lifestyle. In this stage, individuals exhibit heterogeneity in both their utility costs associated with education and their utility costs associated with acquiring a health-protective lifestyle. 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 interactions between education and lifestyle investments for three reasons. First, as it is standard in models of health investments (e.g. Becker, 2007), an extra year of life is more valuable with higher consumption possibilities. This means that the value of a *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 returns of a *protective* lifestyle in terms of better health transitions are larger for higher-educated individuals. And third, as highlighted in previous work (e.g. Hosseini et al., 2025), the labor earnings losses due to bad health are larger for the less educated. The first two ingredients generate a complementarity between education and lifestyle choices, while the third one goes in the opposite direction. When the choices of education and lifestyle are complementary, the model generates an education gradient of lifestyles. Importantly, complementarity also shapes the endogenous selection of individuals with lower costs of health investments into higher education, which further increases the education gradient of lifestyles.

We feed our model with the health and earnings dynamics estimated from the data and calibrate the adult life stage to replicate the wealth accumulation over the life cycle of individuals with different education and lifestyles. This delivers a value of *protective* lifestyle (compared to *detrimental*) that is large and increasing with education, which implies that the two investments are complementary. In a second step, we calibrate the early life stage of the model to match the joint distribution of education and lifestyles for cohorts born in the 1930s and 1970s, where cohorts differ in their education-specific life cycle paths of labor earnings.

Next, we quantify the relative importance of the different mechanisms driving the education gradient in lifestyle behaviors for the 1970s cohort. We find that the income advantage plus the higher yield of health-protective lifestyles of the more educated explain 2/3 of the education gradient in health behavior and 2.4 years of the education gradient in life expectancy, with the income advantage being somewhat more important. The endogenous selection into college of individuals with lower costs of investing in their health explains the remaining 1/3 of the education gradient in health behavior and 1 year of the education gradient in life expectancy.

Finally, in our third question, we study how the increase in education inequality in labor earnings between individuals in the 1930s and 1970s cohorts has shaped the increase in the education gradient of life expectancy. We find that the direct effect of worse (better) economic prospects for the high school dropouts (college graduates) in the 1970s cohort generates a widening of the education gradient of health behaviors, which leads to an increase in the education gradient of life expectancy of about 5 months. Through an increase in complementarities, this sets in motion a selection mechanism that generates a further 7-month increase in the education gradient of life expectancy, as the smaller group of high school dropouts in the 1970s contains a larger fraction of individuals for whom it is more costly to invest in their health with no worsening of the pool of college graduates. All in all, the increase in income inequality across education groups generates an increase in the education gradient of life expectancy of 1 year (50% of the observed increase), mostly driven by the decline in the life expectancy of the less educated. In a model extension where earlier cohorts are allowed to have less accurate knowledge about the long-term consequences of health-related behaviors, we find that the improvement of information across cohorts may itself generate an increase in the educational gradient of life expectancy, as the perceived complementarities between lifestyles and education increase. In that case, the role of selection becomes more prominent, while the direct role of changing income inequalities decreases.

1.1 Related literature

Our estimation of latent types in health behavior relates to the recent literature estimating unobserved fixed effects on health dynamics. Hosseini et al. (2022) and De Nardi et al. (2025) were the first to show that unobserved fixed effects are needed to explain long-run health dynamics. Contemporaneous work by Borella et al. (2024) and Hong et al. (2025) estimates unobserved fixed effects by k-means clustering of health trajectories. In contrast, we classify 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 socioeconomic 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), Bueren (2023), or Nakajima and Telyukova (2025)— and welfare —see Capatina (2015), Braun et al. (2019), De Nardi et al. (2025), or Hosseini et al. (2025). 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 lifestyles by individuals making different education choices.

The paper is also related to previous work featuring endogenous health dynamics. Fonseca et al. (2021) and Hong et al. (2025) consider monetary health investments, but they struggle to find a strong causal effect of medical spending on health outcomes. Ozkan (2023) does find a stronger role for medical spending, but mostly due to the preventive health care component, which aligns well with our notion of lifestyle. 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 lifestyle choices as a potential driver of the educational gradient of health outcomes.[1] 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, but education choices and the costs of good health behavior in each education group are exogeneously given. Instead, our model allows for the sorting of individuals with different costs of health behavior into different education groups and leverages on the persistence of health behavior —often established early in life— to model health and education investments jointly. This allows us to study type formation and assess the extent to which selection drives differences in health outcomes across education groups. This makes our approach more similar to Hai and Heckman (2022), who study the interaction between smoking, preferences, and education.

Our results on the long-run changes in health inequalities connect 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,

[1]The empirical evidence on the effect of medical spending on health outcomes in the US is mixed. While Card et al. (2009) find that access to Medicare at the 65 years of age discontinuity leads to better health outcomes for emergency room patients, quasi-experimental evidence extending the use of Medicaid to low-income households, like Aron-Dine et al. (2013), Finkelstein et al. (2012), and Baicker et al. (2013) find very small effects.

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

	Mean					AC	
	HSD	HSG	CG	65-70	75-80	65-70	75-80
Cancer test	0.60	0.71	0.79	0.77	0.72	0.40	0.41
Cholesterol	0.78	0.85	0.89	0.84	0.85	0.30	0.27
Flu shot	0.67	0.74	0.81	0.66	0.76	0.59	0.59
Drinking	0.01	0.02	0.02	0.02	0.02	0.43	0.44
Smoking	0.18	0.15	0.07	0.14	0.08	0.70	0.64
Exercise	0.27	0.38	0.54	0.41	0.38	0.41	0.37

Notes: Data from the HRS. HSD: high school dropout; HSG: high school graduate; CG: college graduate; 65-70: sub-sample of individuals aged 65 to 70; 75-80: sub-sample of individuals aged 75 to 80. The last two columns show the autocorrelation (AC) of each health behavior with a 4-year lag.

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 Novosad et al. (2022), who find that the worsening of the health outcomes of non-college-educated males is partly due to selection.

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 (reporting more than two alcoholic beverages every day), 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). We sacrifice some information by restricting to binary variables, but this facilitates the estimation of our complex econometric model. 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 65 and 70 years of age, the 4-year autocorrelations of heavy drinking or a flu shot are 0.43 and 0.59, respectively (see Column 6 in Table

Table 2: Cross correlation health behaviors

	Drinking	Smoking	Cancer test	Cholesterol	Flu shot	Exercise
Drinking	1.00	0.04	0.01	-0.00	-0.02	0.02
Smoking	0.07	1.00	-0.08	-0.07	-0.08	-0.06
Cancer test	-0.02	-0.15	1.00	0.28	0.19	0.10
Cholesterol	-0.02	-0.12	0.36	1.00	0.22	0.07
Flu shot	-0.05	-0.09	0.21	0.24	1.00	0.03
Exercise	-0.01	-0.09	0.11	0.03	0.01	1.00

Notes: Data from the HRS. Lower diagonal: individuals aged 65 to 70. Upper diagonal: individuals aged 75 to 80.

1). This weak persistence is partly the result of a life-cycle pattern, as the incidence of healthier behavior changes with age. For instance, the incidence of smoking falls from 0.14 of the population aged 65 to 70 to 0.08 of the population aged 75 to 80, while the incidence of flu shots rises from 0.66 to 0.76 (see Columns 4 and 5 in Table 1). And third, these different measures of health behavior are correlated across individuals as one would expect, although the correlations are relatively small (see Table 2). For instance, the correlations across individuals aged 65 to 70 of cholesterol tests with flu shots or cancer tests are fairly large, 0.24 and 0.36, respectively, while the correlations of cholesterol tests with drinking, smoking, and exercise habits are much smaller, -0.02, -0.12, and 0.03, respectively.[2]

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. We present such a factor model in the next Section.

3 An econometric model of health dynamics with latent types

We combine data from the HRS and the PSID to create an unbalanced panel of individuals $i = 1, \dots, N$ followed for $t = 0, \dots, T_i$ 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 $\mathrm{e}_i \in \{\textsc{hsd}, \textsc{hsg}, \textsc{cg}\}$ (high school dropout, high school graduate, college graduate), 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 obvioulsy an absorbing state. We build this variable using

[2]The correlation between healthy behavior and education has been documented before by Lantz et al. (1998) and Cutler and Lleras-Muney (2010), while the imperfect correlation of health behaviors in the cross-section has been highlighted by Cutler and Glaeser (2005) and Bairoliya et al. (2024).

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 $\mathbf{z}_{it} = \{z_{1,it}, z_{2,it}, \dots, z_{N_z,it}\}$ contains information on N_z different categorical variables $z_{m,it} \in \{0, 1\}$ describing whether individual i in period t takes some particular action. These actions are the health behavior variables described in Section 2.[3]

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 $\mathrm{y}_i \in Y \equiv \{y_1, y_2, \dots, y_{N_y}\}$. We interpret the latent variable as the individual lifestyle, which captures the idea that individuals differ in their propensity to undertake actions that are good for their health (see Section 5.1 for a discussion on the assumption of lifestyles being fixed).

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 of distributions. That is, both observed health behavior and observed health transitions help inform the classification of individuals into types.[4] We can write the likelihood of the data as a mixture model:

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

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}.[5] 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

[3]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 preventive health 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.

[4]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.

[5]Given a_{i0} and t, a_{it} is obtained as $a_{it} = a_{i0} + 2t$ (interviews in the HRS and PSID are biennial).

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 the same in all demographic groups.

Next, we assume that conditional on $y_i = y$, 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 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}, but not on individual health behaviors $z_{m,it}$. This simplifying assumption is part of the dimension-reduction exercise.[6] The dependence of health dynamics on gender and education captures 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 $y_i = y$, and age a_{it}, the evolution of health outcomes is markovian, that is, it only depends on one lag of health outcomes:

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

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

[6] 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. First, a positive but low correlation of different health-enhancing behaviors over time and across individuals complicates the isolation of singular behavior effects. Second, 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.

3.3 Distribution of health types

The final element in the likelihood function is the fraction of individuals of each type y in each demographic group, 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 (Dempster et al., 1977). However, in our model with health dynamics that are heterogenous by lifestyle types y, the mixing weights will 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 allows us 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, as well as the formal description of the above argument, are given in Appendix A.3.

3.4 Estimation

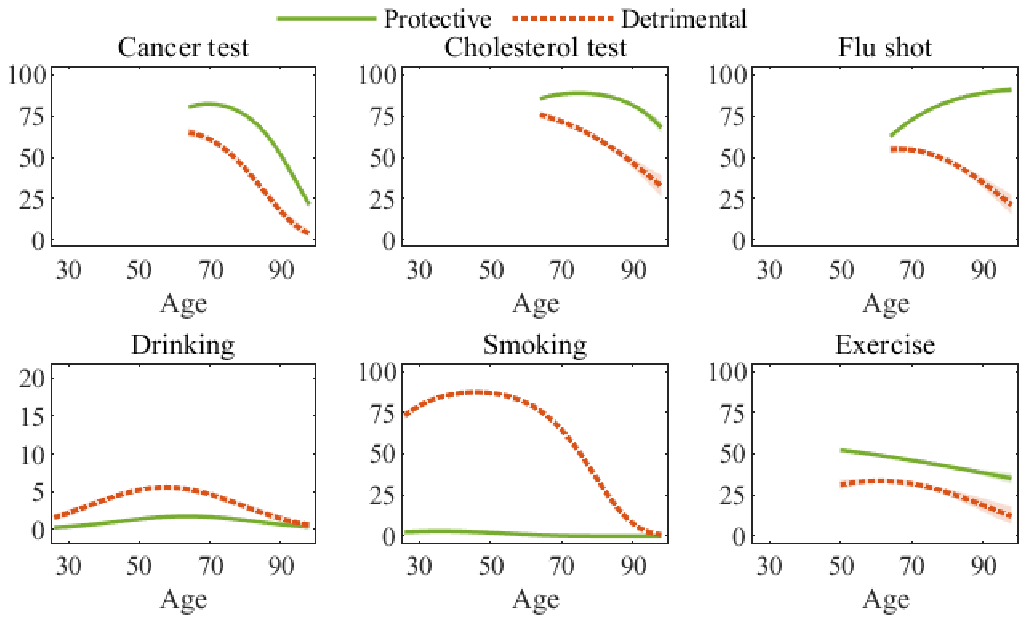
We need to estimate a large vector of parameters $\boldsymbol{\theta}$ plus the probability of each individual of belonging to each latent type. We resort to Bayesian methods, which are particularly well-suited for models with latent structures. Specifically, we implement a Gibbs sampler to break down 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 our sample for estimation, we set to missing all the preventive health behavior variables for individuals under 65 years of age.[7] The results presented below are for males only, which includes 16,296 individuals (interviewed an average of 6.6 times each) in the HRS and 4,738 individuals (interviewed an average of 5.6 times) in the PSID. In favor of parsimony, we choose $N_y = 2$ for our main estimation and discuss the cases with $N_y > 2$ in Appendix B.

[7] Specifically, we set to missing "cancer test", "cholesterol test", and "flu shot". The reason is that access to these tests is typically cheaper for individuals with good health insurance, which may raise concerns regarding their correlation with education. However, from age 65 onwards, these tests are free for all through Medicare. See footnote 3 for our treatment of missings.

Figure 1: Health behaviors and lifestyles



Notes: Estimation results. Probability of engaging in each health behavior by age and type, for male individuals in good health. The shaded areas represent the 95% credible intervals.

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 $h_{it} = h_g$.[8] Individuals in the group that 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). For individuals in the other group, which we label *detrimental* (red dashed line), the probability of reporting all the health-enhancing behaviors 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 65 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 lifestyles and health dynamics. We summarize this information by looking at the gradient in age-50 life expectancy across types.[9] In Table 3 (second column) we show that, on average, *protective* types live for 8.5 more years than *detrimental* types.

[8]The case $h_{it} = h_b$ is not too different, see Figure F.2 in Appendix F

[9]For more details, Figure F.1 in the Appendix shows the actual health transitions and survival functions by lifestyle.

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	29.4	100.0	24.8	100.0	28.0	100.0	32.8	8.0	4.3	3.8
PRO	74.6	31.6	42.9	28.6	68.1	30.1	93.3	33.4	4.9		
DET	25.4	23.0	57.1	22.0	31.9	23.3	6.7	24.1	2.1		
Δ_y	49.3	8.5	-14.2	6.6	36.2	6.8	86.5	9.4	2.8		

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 education gradient in LE when the distribution of behavior types within high school dropouts is the same as for college-educated individuals. Column (b) is the difference between the actual education gradient in LE and the counterfactual in column (a), that is, it corresponds to the gradient explained by differences in lifestyles across education groups for given health dynamics.

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 9.4 more years among males with a college degree (see columns 4, 6, and 8). It is important to note that the lifestyle gradient in life expectancy is increasing in education, that is, the health yield of a *protective* lifestyle increases with education.[10] This is the result of the estimated $p\left(\mathbf{h}^T|\mathrm{s}, \mathrm{e}, y, a_0, h_0; \boldsymbol{\theta}_y\right)$, which allows for interactions between education and lifestyle.

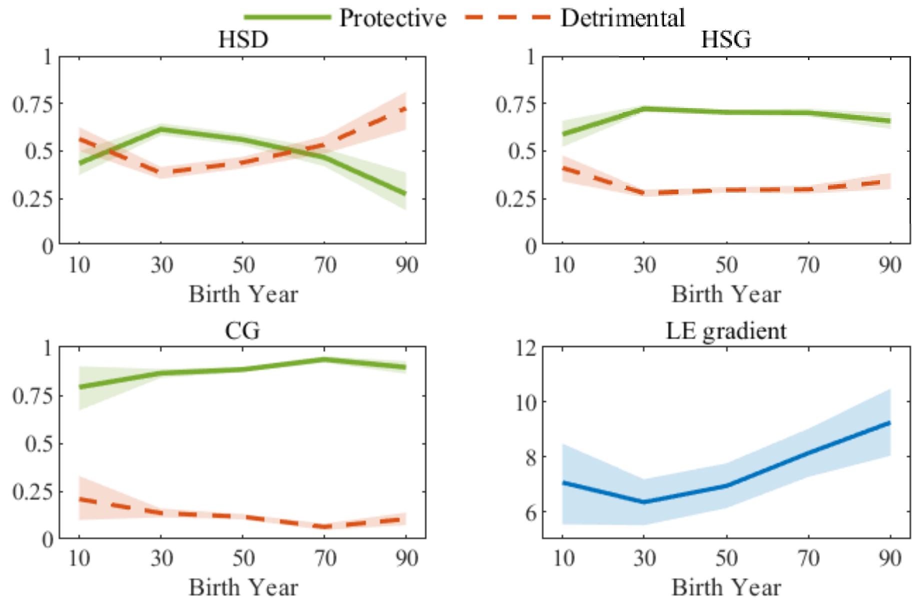
4.3 Distribution of health types

Our estimation assigns a different fraction $p\left(\mathrm{y}|\mathrm{c}, \mathrm{s}, \mathrm{e}, 25, h; \boldsymbol{\theta}_y\right)$ of individuals to each type y depending on cohort c, gender s, and education e. For the 1970s cohort, 74.6% of males are classified as *protective* and 25.4% as *detrimental* (see first column in Table 3). Importantly, there is a large educational gradient of lifestyles: the share of *protective* individuals grows from 42.9% to 93.3% 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 and drinking declines with education, and Table 1 before shows how in our data the share of individuals taking cancer tests, cholesterol tests, and flu shots also increases with education. It is therefore natural that our classification of individuals into types according to the observed health-related behavior retains the education 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.

[10] Our econometric model is silent about the reasons behind this complementarity. One reason may be mechanical: less-educated individuals have lower life expectancy, which means there is less space for health-protective behavior to have an impact. Another reason could be the higher ability 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).

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. The shaded area represents the 95% credible intervals.

In Panels (a) to (c) of Figure 2 we report the distribution of types by education group from the 1910s to 1990 cohorts. Within the high school dropouts, the *detrimental* types increased monotonically from 40% in the 1930s 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 1910s and 1930s 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 1930s 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 almost 1/2 of the educational gradient of life expectancy, which is 8.0 years for the 1970s 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 lifestyles as the males with a college degree. In this counterfactual case, we find that their life expectancy would rise by 3.8 years, reducing the education gradient of life expectancy from 8.0 to 4.3 years. That is, the different health dynamics across education groups for fixed distribution of lifestyle explains 4.3 years or 53% of the education gradient in life expectancy. The difference, 3.8 years or 47% of the observed gradient, is explained by the different distribution of lifestyles across education groups (see the last two columns in Table 3). We emphasize that this result is obtained with only two types; with more types, the share of the education gradient of life expectancy explained by lifestyles is larger, see Appendix B.

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 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 6.3 years in the 1930s cohort to 9.3 years in the 1990 cohort, which follows an 0.8-year decline in the gradient between the 1910s and the 1930s cohorts.

5 An economic model

Our economic model considers two distinct life stages. The *early life* stage (Section 5.1) is a static problem where young individuals belonging to a cohort c choose their education $\mathrm{e} \in \{\textsc{hsd}, \textsc{hsg}, \textsc{cg}\}$ and their lifestyle $\mathrm{y} \in \{\textsc{det}, \textsc{pro}\}$. This problem is a stand-in for choices and investments made by parents in early childhood or young adults before entering the labor market. This stage serves to account for the observed correlation between education and lifestyles. The *adult life* stage (Section 5.2) 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 provides the endogenous 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 (see Section 5.2). 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 utility costs of choosing an education e and a lifestyle y, respectively. These education and health behavior costs are heterogeneous in the population.[11]

We want to comment on several of the assumptions embedded into equation (4). First, our

[11] 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 lifestyle choices. Early life access to healthy food or sports facilities, as well as parental characteristics, may shape this heterogeneity.

formulation imposes that education and lifestyle choices are taken together at a young age and never change.[12] The main reason for this assumption is the empirical evidence on the early-age adoption of health-related habits and on the fact that, once adopted, lifestyles tend to be very persistent.[13] Furthermore, given that our study focuses on differences in lifestyles across education groups and cohorts, understanding the initial formation of persistent lifestyles is likely the most relevant margin. Second, we abstract from considering potentially *higher prices* of adopting healthy behavior, which would favor the more educated. While this factor could play a role in behaviors like diet quality (not in our data), it is important to note that harmful behaviors like smoking and drinking are expensive and, hence, go in the opposite direction. Furthermore, in our analysis, we exclude the information on preventive health behavior —which may be linked to health insurance— for individuals below the age of 65 and hence not covered by Medicare. Third, we abstract from heterogeneity in discount factors (which could trigger correlated choices of education and lifestyle) because our empirical evidence on wealth accumulation is not consistent with the idea that more patient individuals largely self-select into both higher education and better lifestyles.[14] Futhermore, direct empirical tests find no impact of discount rate heterogeneity on the education gradient of health behavior, see for instance Fuchs (1982) or Cutler and Lleras-Muney (2010). And fourth, we abstract from intergenerational linkages as recent research shows genetic inheritance to play a very minor role in health dynamics.[15]

Distributional assumptions. Going into details, we normalize to 0 the cost of a *detrimental* lifestyle and the cost of not finishing high school ($\tau_{\text{DET}} = 0$ and $\tau_{\text{HSD}} = 0$). Next, we assume that the cost of *protective* behavior τ_{PRO} is heterogeneous in the population and described by some CDF, $F_y(\tau_{\text{PRO}})$, while the costs of graduating from college and high school $(\tau_{\text{CG}}, \tau_{\text{HSG}})$ are also heterogeneous in the population and described by another CDF, $F_e(\tau_{\text{CG}}, \tau_{\text{HSG}})$. Throughout the paper, we assume that these two distributions are independent of 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 endogenously through our model mechanisms.[16]

Independent choices. In this framework, the choices of education and lifestyle are independent of each other if and only if $V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} = V_0^{\text{HSG,PRO}} - V_0^{\text{HSG,DET}} = V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$; that is, if

[12] Recall 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.

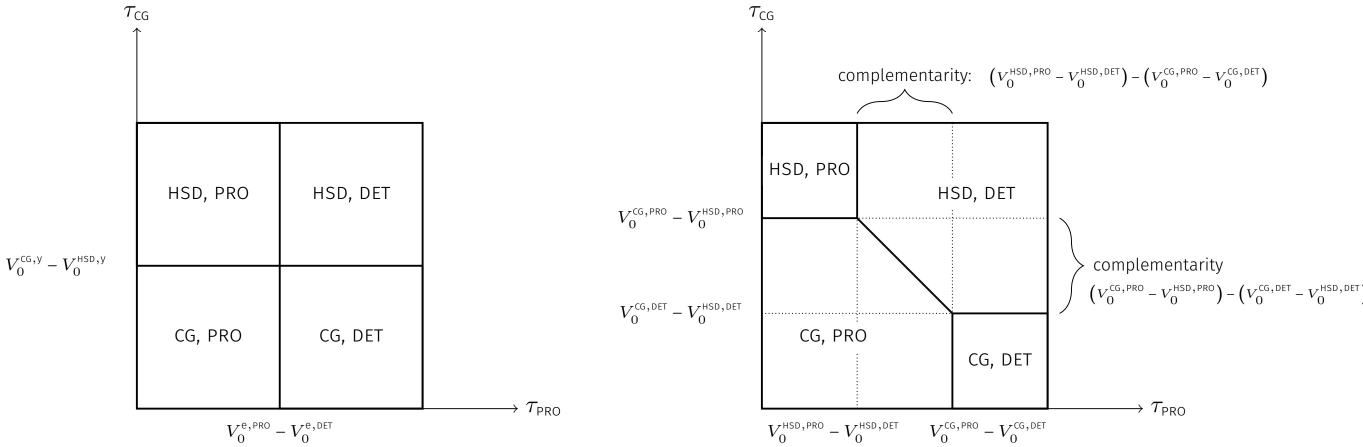
[13] See Farrell and Fuchs (1982) or Hai and Heckman (2022) for the early adoption of smoking and how it correlates with later decisions of college education. Conner and Norman (2017) —and references therein— document the small effects of interventions to change lifestyles during adulthood.

[14] Wealth accumulation in the data is indeed larger for the more educated and the *protective*, but the model with homogeneous discount factors is able to account for it very well, see Section 6.1.2.

[15] See Ruby et al. (2018) or Kaplanis et al. (2018). Furthermore, research conducted with the UK Biobank shows environmental exposure (including lifestyle factors and education) to be much more important for longevity than genetics, see Argentieri et al. (2025).

[16] Alternatively, one could assume that individuals with a low cost of education also tend face a low cost of healthy lifestyles. See Appendix E for a robustness exercise allowing for correlation between these two distributions.

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.

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 categories (we drop e = HSG).[17] 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 difference 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.

Complementarity. 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, as 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 case with 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 studying (τ_{CG}) choose a college education and a *protective* lifestyle when the cost of *protective* lifestyle (τ_{PRO}) is low. However, when the cost of *protective* lifestyle is high, they choose a lower level of education and a *detrimental* lifestyle, as the value of college is lower with a *detrimental* lifestyle. The complementarity between the two investments will be the relevant case of our calibrated model, and it is consistent with the empirical evidence showing that increases in compulsory education lead to reductions in adult mortality (e.g. Lleras-Muney, 2005) or that increasing the cost of harmful behavior leads to better

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

school outcomes (e.g. Bagues and Villa, 2025).

Selection. 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, $\Pr[\tau_{\text{PRO}} \leq \tau | \text{HSD}] \leq \Pr[\tau_{\text{PRO}} \leq \tau | \text{CG}]$ $\forall \tau$, with strict inequality 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 have lower costs τ_{PRO} than individuals without a high school degree. This makes them more likely to choose a *protective* lifestyle just because of selection.

5.2 Stage 2: adult life

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

Demographics and preferences. The model period corresponds to two years. Individuals live for at most T periods, but survival is stochastic every period. People are exposed to health shocks, medical expenditure shocks and, up to period $R-1$, 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 degree of non-homotheticity.

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

Labor earnings and medical expenses. At every period of their working life, individuals receive an exogenous income, which we model in two components. First, there is an employment shock $\ell_t \in \{0,1\}$, with the probablity of employment $\text{Prob}\left(\ell_t = 1|\text{c},\text{e},t,h_t\right)$ being an *i.i.d* shock conditional on cohort, education, age, and health. This component aims to capture the higher non-employment probability of individuals in bad health —which is at the core of the health cost of labor income— and the differences in employment across education groups. Second, conditional on working, individuals receive labor income

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

where $\omega_t^{\text{c,e}}(h)$ is a deterministic component depending on cohort, education, age, and health, while ξ_t and ϵ_t are persistent and transitory shocks, respectively, with the initial value of the persistent component ξ_0 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, while the transitory component ϵ_t is a Gaussian white noise with zero mean and variance σ^2_ϵ. Finally, medical expenses are given by $\log m_t^{\text{e}}(h_t, \zeta_t) = \mu_t^{\text{e}}(h_t) + \zeta_t$, where $\mu_t^{\text{e}}(h_t)$ is a deterministic component depending on education e, age t, and health h_t, while ζ_t is an i.i.d gaussian white noise process with zero mean and variance $\sigma^{\text{e}}_{\zeta,t}(h)$.

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 as follows. 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 lifestyle. 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.

5.2.1 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 ($t < R$) is:

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

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

Different from standard models, we discount the flow utility of bequests by β^{T-t} instead of β. This implies that, from the point of view of the newborn, the value of a bequest in our formulation depends on the bequeathed amount but not on the age of death (as it does in the standard warm-glow formulation). We do this for two reasons. First, a strong bequest motive may produce a value of life $V_0^{\mathrm{c,e,y}}$ that is lower for PRO than for DET individuals just because DET individuals get to enjoy the bequest motive earlier in life, which distorts the endogenous choice of lifestyle. Second, with heterogeneous survival probabilities, the incentives to save with the standard bequest motive are distorted, inducing individuals with lower survival probabilities to save more to enjoy the bequest, see Foltyn and Olsson (2024). However, evidence from genetic testing shows this not to be the case, see Karpati (2023). The optimization problem for retired individuals ($t \geq R$) is analogous, with a 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.

In this problem, wealth accumulation will be larger for the more educated and for the *protective* lifestyle for two different reasons. First, both groups earn more labor income net of medical expenses (either directly or through the lower incidence of bad health). Second, both groups feature higher survival, which translates into larger effective discount factors and larger willingness to save.

6 Calibration

The calibration has two parts: the parameters related to the life cycle model of adults (Section 6.1) and the parameters shaping unobserved heterogeneity in the early life stage (Section 6.2).

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% and the risk aversion parameter σ at 1. 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 $w_t^{\mathrm{c,e}}(h, \xi, \epsilon)$ and employment $\mathrm{Prob}\,(\ell_t = 1|\mathrm{c}, \mathrm{e}, t, h_t)$ processes. The census data allows us to observe the average life cycle labor market trajectories of individuals of different education levels in several cohorts. However, the census data does not report information on health, which is where the PSID becomes useful. Furthermore, we will also exploit the panel dimension of the PSID to estimate the stochastic component of wages.

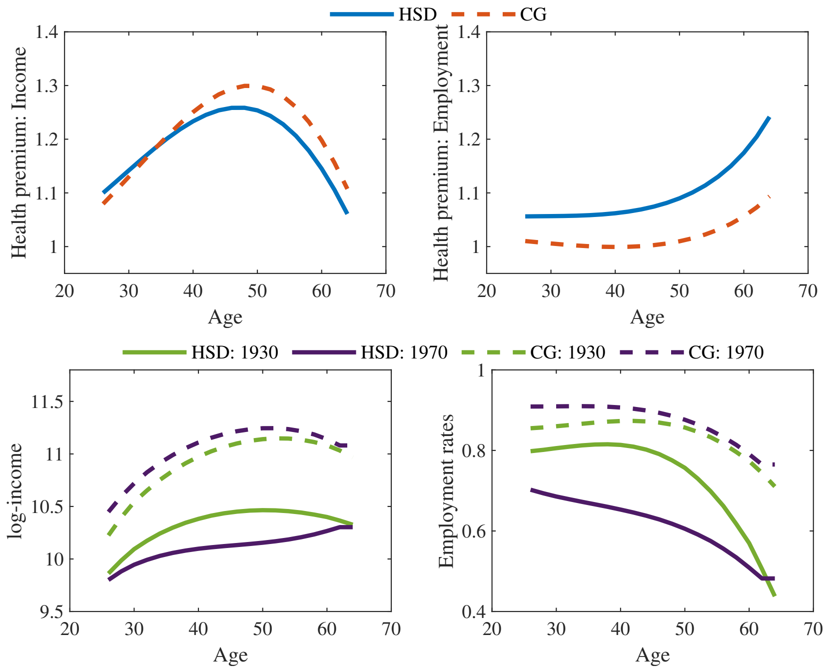
We model the deterministic component of wages, $\omega_t^{\mathrm{c,e}}(h_t)$, as the sum of two terms: a polynomial in age fully interacted with education and cohort dummies plus another polynomial in age fully interacted with education and health dummies. This specification allows for the education paths of wages to be different across cohorts and for the wage losses due to bad health to be different across education groups. However, due to data limitations, this specification restricts the education-specific wage losses due to bad health to be the same for all cohorts. For the estimation, we use employed males between 25 and 60 years of age.[18] We use the Census data to estimate the first component and the PSID to estimate the second one. Finally, we use the PSID data to estimate the stochastic components of income, ξ_t and ϵ_t, under the assumption that the stochastic component of wages has not changed across cohorts. We model the employment probability $\mathrm{Prob}\,(\ell_t = 1|\mathrm{c}, \mathrm{e}, t, h_t)$ 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 educated, which is consistent with recent findings (e.g. Hosseini et al., 2025).

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 the 1930s and 1970s. 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 1930s and 1970s cohorts, the average lifetime labor earnings declined 18 percent for the high school dropouts, while they increased 12 percent for college graduates.

[18] We define as employed those individuals reporting annual labor earnings above $3,770 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 (1930s and 1970), data from Census.

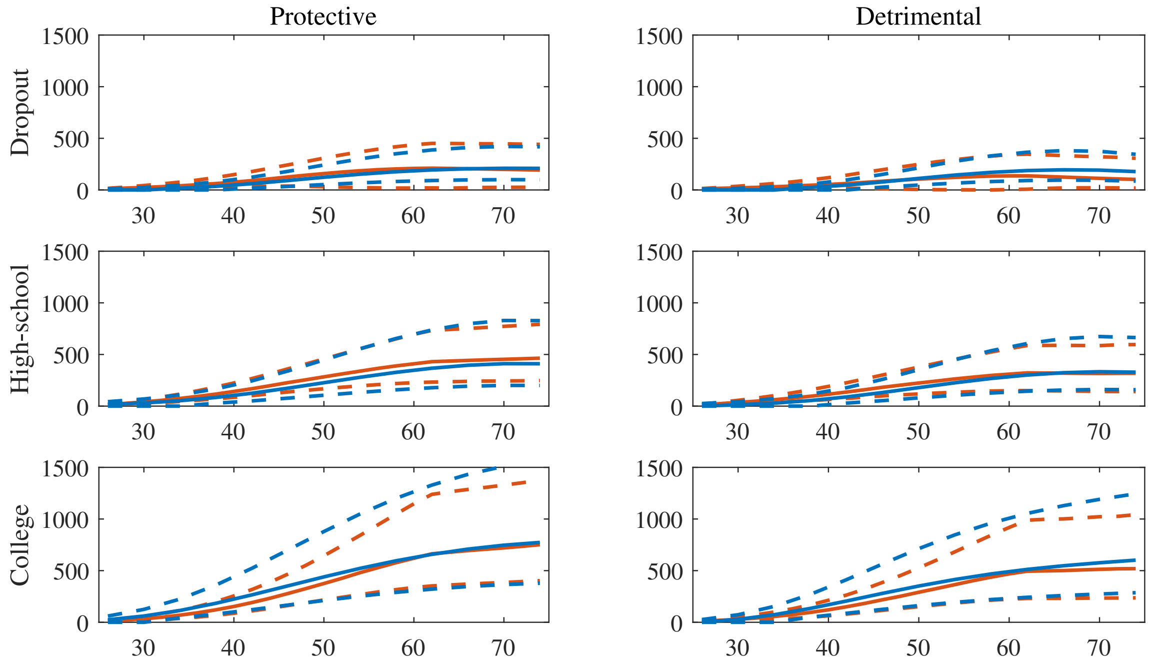
Finally, the estimated parameters of the stochastic component of the income process show that more educated individuals face larger and more persistent shocks and a greater proportion of the variance in their earnings coming from the persistent component, see Table F.1 in Appendix F.

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

We first need to recover from the data the wealth distribution by age, education, and lifestyle. This is not straightforward, as individuals are probabilistically assigned to 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 lifestyle 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

Figure 5: Model fit. Wealth trajectories



Notes: The solid blue lines represent the median wealth by age, education, and lifestyle obtained from the data, see Appendix A.5 for details. 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.

Table 4: Calibrated parameters

Parameter	Description	Value	
		Benchmark	Perceptions
Panel A: Adult Life			
β	Discount factor	0.952	0.953
$\underline{x}$	Income floor	17.97	17.35
b_0	Bequest motive: marginal utility	15.11	16.11
b_1	Bequest motive: non-homoteticity	296.36	417.77
$\bar{b}$	Value of life	3.44	4.38
Panel B: Early Life			
μ_{HSG}	Average cost of HSG education	4.99	5.48
μ_{CG}	Average cost of CG education	22.19	21.28
μ_{PRO}	Average. cost of PRO lifestyle	2.23	-0.77
σ_{HSG}	Sd. cost of HSG education	1.85	1.22
σ_{CG}	Sd. cost of CG education	18.56	14.81
σ_{PRO}	Sd. cost of PRO lifestyle	2.86	7.18
λ_{1930}	Probability of right expectation	—	0.60

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.

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 biennial discount factor is 0.952 (0.976 annual) and the bequest parameters b_0 and b_1 are 15.11 and 296.36, 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 \$20,602, and the marginal propensity to bequeath is 94 cents out of every additional dollar. The biennial income floor is estimated at \$17,970. These parameters fall in the range of parameters estimated by the previous literature, see discussion in Ameriks et al. (2020).

6.1.3 Value of life

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.[19] 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 40 years of age, see Appendix D.4 for details. We target a VSL of \$1,000,000 and in Appendix E we discuss the case with a VSL of \$4,000,000.

6.1.4 Taking stock

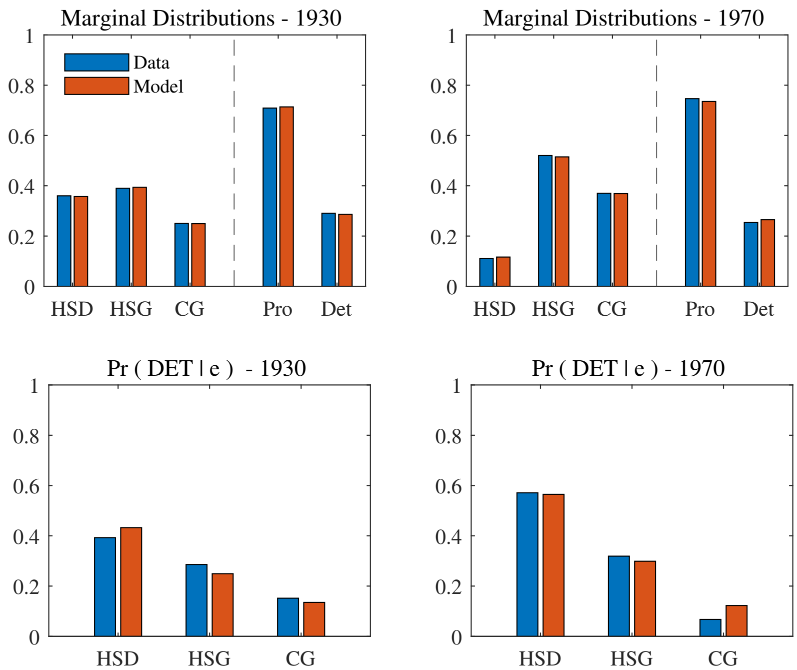
With this calibration, the model delivers a value of *protective* lifestyle, $V_0^{\mathrm{e,PRO}} - V_0^{\mathrm{e,DET}}$, that is large and increasing in education e. In particular, when starting his working life, a *detrimental* high school dropout would be willing to pay 22% of his consumptiom flow in all periods and states to become *protective*. This figure grows to 28% and 41% for high school graduates and college graduates. Therefore, the calibrated model delivers a strong complementarity between lifestyle and education investments. We note that this complementarity has been informed by the adult life stage of the model, and that it does not use any information on the observed correlation between education and lifestyles.

6.2 Stage 1: Early life

We parameterize the cost of *protective* behavior as $\tau_{\mathrm{PRO}} = \mu_{\mathrm{PRO}} + \sigma_{\mathrm{PRO}}\epsilon_{\mathrm{PRO}}$ with $\epsilon_{\mathrm{PRO}} \sim N(0,1)$. The cost of graduating from high school and the cost of college education are parameterized similary,

[19] See for instance Murphy and Topel (2006), Hall and Jones (2007), Jones and Klenow (2016), De Nardi et al. (2025), Mahler and Yum (2022), or Hong et al. (2025) among others.

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.

with $\tau_{\text{HSG}} = \mu_{\text{HSG}} + \sigma_{\text{HSG}}\epsilon_e$ and $\tau_{\text{CG}} = \mu_{\text{CG}} + \sigma_{\text{CG}}\epsilon_e$, with $\epsilon_e \sim N(0,1)$.[20] In order to calibrate the parameters of these distributions, we require the model to match the joint distribution of education and lifestyles for two different cohorts: the 1930s and 1970s. The value functions $V_0^{\text{c,e,y}}$ vary by cohort c due to the different paths of wages and employment rates over the life cycle, with the expected lifetime labor income of the 1970s cohort being 18% lower for the high school dropouts and 12% larger for college graduates (see Section 6.1.1).

6.2.1 Identification and results

We have 6 parameters to estimate: $(\mu_{\text{CG}}, \mu_{\text{HSG}}, \mu_{\text{PRO}}, \sigma_{\text{CG}}, \sigma_{\text{HSG}}, \sigma_{\text{PRO}})$. 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 principle, one may think that the average cost parameters $(\mu_{\text{CG}}, \mu_{\text{HSG}}, \mu_{\text{PRO}})$ are identified by the marginal distributions of education and lifestyle in the 1930s cohort, while the dispersion parameters $(\sigma_{\text{CG}}, \sigma_{\text{HSG}}, \sigma_{\text{PRO}})$ are identified by the changes of these distributions as the paths of labor earnings change between the 1930s and 1970s cohorts.[21] 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.

[20] As discussed in Section 5.1, we assume ϵ_{PRO} and ϵ_e to be uncorrelated. See Appendix E for a robustness exercise with $\text{corr}(\epsilon_{\text{PRO}}, \epsilon_e) > 0$.

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

The upper panels of Figure 6 show that the model matches well the marginal distributions of education and health behavior in the 1930s cohort as well as their changes across cohorts. 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 15.2% in the data and 13.5% 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 1930s and 1970s cohorts, with the proportion of among high school dropouts with *detrimental* lifestyles increasing by 17.8 percentage points in the data and 13.3 percentage points in the model. The model is less successful at matching the improvement in health behavior of the college educated, as the proportion of individuals with *detrimental* lifestyles among college graduates has declined by 8.5 percentage points in the data and 1.2 percentage points in the model. This implies that quantitatively, the model accounts for 14.5 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 1.0 years in the model and 2.0 years in the data. Hence, the calibrated model, with only wage and employment changes, accounts for around 50% of the overall increase in the life expectancy gradient between college graduates and high school dropouts between the 1930s to 1970s cohorts.

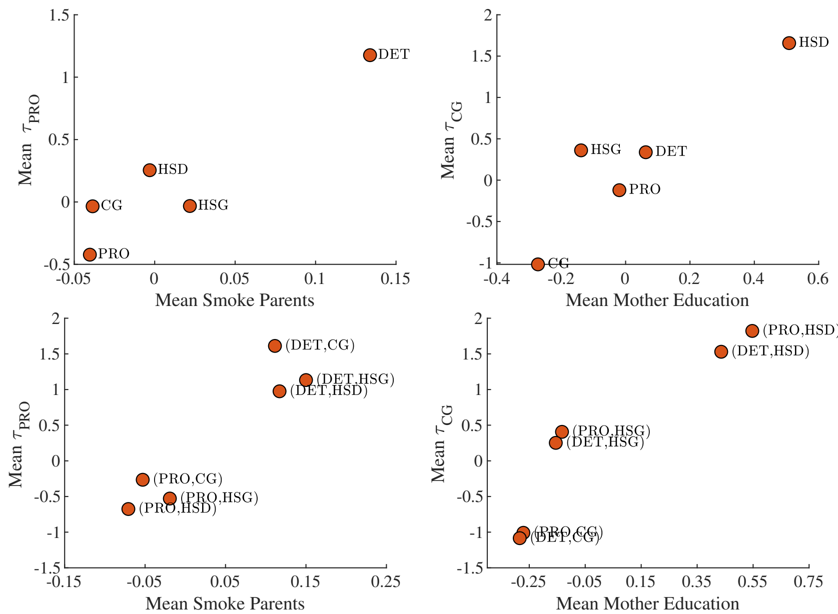
6.2.2 Validation

A relevant question to validate the calibration is whether the recovered distribution of unobserved heterogeneity in τ_{PRO} and τ_{CG} resembles relevant proxies in the data. The HRS reports 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} than individuals that choose a *detrimental* lifestyle. We can see this in the y-axis on the top-left panel of Figure 7. Consistently, in the data, individuals classified as *protective* are less likely to have smoking parents, see the x-axis on the same panel. That is, the dot DET is northeast of the dot PRO. At the same time, the model predicts that individuals who choose college education have on average lower τ_{CG} than individuals with lower education levels. Consistently, in the data, individuals classified as more educated are less likely to have a high school dropout mother, see the top-right panel in Figure 7. In particular, the dots HSD, HSG, and CG are northeast of each other.

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

Figure 7: Shocks and Parental Characteristics



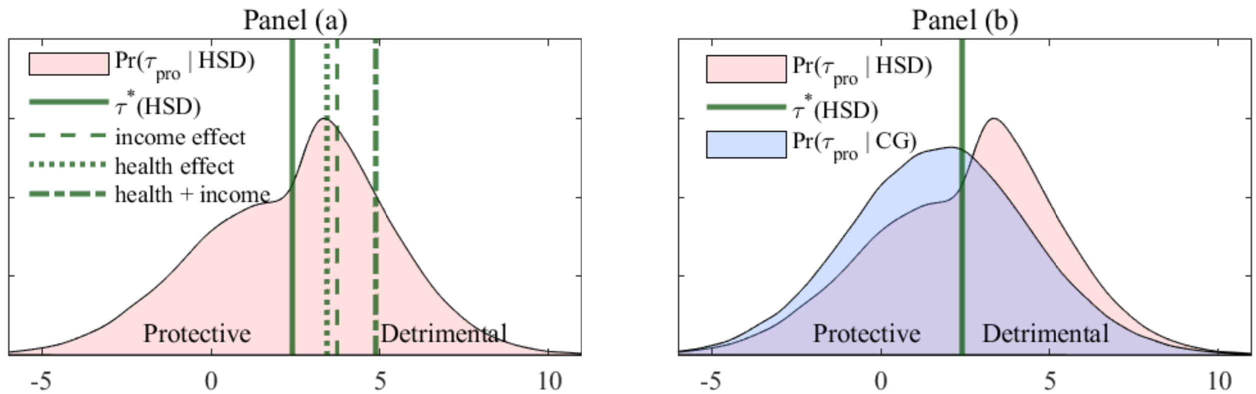
Notes: The top two panels report the average τ_{PRO} (top left) and the average τ_{CG} (top right) for each education and lifestyle category in the calibrated model (Y-axis) against the average share of smoking parents (top left) and average share of high school dropout mothers (top right) for the same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

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 fully 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, the model predicts that, 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 a 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 triggers a low-educated individual to choose a *protective* lifestyle is a low cost of *protective* lifestyle, as the gains for him are lower. Analogously, what leads a highly educated individual to choose a *detrimental* lifestyle is the high cost of *protective* lifestyle, as the gains for him are higher. The data largely supports these patterns. Among the individuals classified as *protective*, the parents of high school dropouts are less like to smoke than the parents of those who go to college, see dots (PRO,CG) and (PRO,HSD) in the bottom-left panel of Figure 7. Similarly, among the individuals classified as *detrimental*, the parents of those who finished high-school are more likely to smoke than the parents of those who drop out which lines up well with our model see dots (PRO,HSG) and (PRO,HSD) in the bottom-left panel of Figure 7.[22]

[22]In both cases there is, however, a discrepancy: among *detrimental* types, the parents of high-school graduates smoke more than parents of high-school dropouts, as predicted by the model, but less than the parents of college

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



Notes: In Panel (a), the shaded red area describes the density function of τ_{PRO} for individuals who choose HSD education. The solid thick green vertical line reports the threshold τ^*(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 semi-dotted green line reports this same threshold if HSD individuals had both the income prospects and health yield of *protective* lifestyle as CG individuals. The shaded blue area in Panel (b) describes the density function of τ_{PRO} for individuals who choose CG education.

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, a pattern similar in the data. 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 dots (PRO,HSD) and (DET,HSD) and the dots (PRO,HSG) and (DET,HSG) 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: why is there an education gradient of lifestyles (Section 7.1), and what has been the effect of the rising education premium of labor earnings on the increase in education gradient of life expectancy (Section 7.2 and 7.3).

7.1 The education gradient of lifestyles

Our calibrated model delivers a complementarity between education and lifestyle investments, which in turn generates the observed correlation between education and lifestyles. There are three reasons for this. First, higher expected income among the more educated encourages healthier behavior as life becomes more valuable with higher consumption possibilities. This is the standard mechanism in models of health investments through health behavior, see Becker (2007) and recent work by

graduates, while among the *detrimental* types, the parents of college graduates do not smoke more than the rest.

Cole et al. (2019), Mahler and Yum (2022), or Margaris and Wallenius (2023). Second, 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 educated). Finally, given these two complementarities, individuals facing lower costs of *protective* behavior (lower τ_{PRO}) are more likely to pursue higher education, which 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 shaded red area in Panel (a) of Figure 8 plots the distribution of *protective* lifestyle costs (τ_{PRO}) for individuals who choose to drop out of high school in the 1970s 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 individuals 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} conditional on HSD between minus infinity and $\tau^*(\text{HSD})$ gives the fraction of dropouts adopting a *protective* lifestyle.

We conduct a series of decomposition exercises, in which we keep individuals' education choices fixed and observe how their lifestyle investments would differ under various scenarios. In the first scenario, we compute the behavior of high school dropouts if they were to face the income prospects of college graduates. Due to the higher consumption possibilities, the increase in expected earnings leads to higher value functions for both lifestyle options, which we denote as $\widetilde{V}^{70,\text{HSD},\text{PRO}}$ and $\widetilde{V}^{70,\text{HSD},\text{DET}}$. We also find that the gain is higher for y = PRO, as life expectancy is higher for this type and the higher consumption flow is enjoyed for more years. The larger gain for y = PRO is despite the fact that the income losses due to bad health are larger for high school dropouts than for college graduates (see Section 6.1.1), which pushes in the direction of larger gains for the *detrimental* lifestyle. 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 shifts to the right the threshold value that prompts individuals to adopt a *protective* lifestyle , see the dashed green vertical line in Figure 8. In this case, the proportion of high school dropouts choosing *protective* behavior would increase from 44% to 65%, reducing the education gradient of life expectancy gap in 1.3 years (from 7.4 to 6.1 years, an 18% reduction), see the 2nd and 4th rows in Table 5.

In the second scenario, we compute the behavior of high school dropouts if they were to face the same health transitions —conditional on lifestyle— as college graduates. The higher life expectancy gives high school dropouts higher counterfactual values $\widehat{V}^{70,\text{HSD},\text{PRO}}$ and $\widehat{V}^{70,\text{HSD},\text{DET}}$, the former increasing more because the health transitions of college graduates feature larger life expectancy gains of a *protective* lifestyle. That is, $\hat{\tau}^*(\text{HSD}) \equiv \widehat{V}^{70,\text{HSD},\text{PRO}} - \widehat{V}^{70,\text{HSD},\text{DET}} > \tau^*(\text{HSD})$, see the vertical dotted line in Panel (a) of Figure 8. In this case, the proportion of high school dropouts choosing a *protective* lifestyle would increase from 44% to 60%, reducing the life expectancy gap by 1 year (from 7.4 to 6.4, a 13% reduction), somewhat less than in the previous exercise, see 5th

Table 5: Decomposition: 1970s cohort

	Pr(e)		Pr(PRO)	Pr(PRO|e)			LE(e)		
	CG	HSD		CG	HSD	CG-HSD	CG	HSD	CG-HSD
Data	0.37	0.11	0.75	0.93	0.43	0.50	32.8	24.8	8.0
Model	0.37	0.12	0.74	0.88	0.44	0.44	32.3	24.9	7.4
Same lifestyle				0.88	0.88	0.00	32.3	27.7	4.5
Income				0.88	0.65	0.23	32.3	26.2	6.1
Health				0.88	0.60	0.28	32.3	25.9	6.4
Income + Health				0.88	0.81	0.07	32.3	27.3	5.0
Selection				0.88	0.59	0.28	32.3	25.9	6.4

Notes: The first two rows report the results from the data and from the calibrated model. The 3rd row reports the results of a model where the lifestyle distribution is the same in all education groups. The 4th to 7th rows perform the counterfactuals where high school dropouts are given, respectively, the same income prospects, the same health gains of *protective* behavior, both things together, and the same underlying distribution of τ_{PRO} as college graduates. See Section 7.1 for details.

row in Table 5.

In the third scenario, we compute the behavior of high school dropouts if they had the same distribution of τ_{PRO} as college graduates. Panel (b) in Figure 8 displays the distributions of the cost of *protective* behavior for both high school dropouts (red shaded area) and college graduates (blue shaded area), with the latter shifted to the left as predicted by Proposition 1. As a consequence, the mass of individuals below the $\tau^*(\text{HSD})$ threshold (green vertical line) is larger for the distribution of τ_{PRO} faced by college graduates. In this case, the proportion of *protective* individuals among high school dropouts would increase from 44% to 59%, reducing the education gradient of life expectancy gap in 1 year (from 7.4 to 6.4 years, a 13% reduction), see 6th row in Table 5.

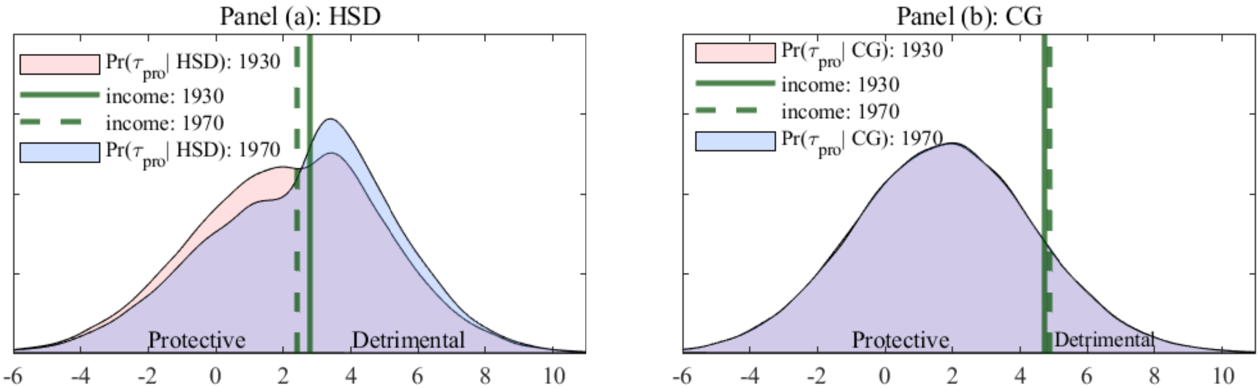
Overall, if we do the three exercises together the distribution of lifestyles across education groups is equalized and the education gradient of life expectancy decreases in 2.9 years, see 3rd row in Table 5. Of those, 1.3 years (45%) correspond to income differences across education groups, 1 year (34%) to health transition differences between education groups, and 1 year (34%) to the selection generated by these two forces.[23]

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 2 years across the 1930s and 1970s cohorts. This increase is driven by worse (better) lifestyles among high school dropouts (college graduates). Our calibrated model, as discussed above, produces 50% of this observed increase. In this Section, we quantify how much of the increase in the education gradient of life expectancy is due to the

[23]The sum of the three channels adds up to 3.3 years, slightly above the total of 2.9. This is because of the negative interaction between the first two channels and selection, which can be seen in Figure 8. The region of the distribution of τ_{PRO} that switches from *detrimental* to *protective* lifestyle when increasing income and/or improving health transitions becomes less populated when giving high school dropouts the distribution of τ_{PRO} of college graduates.

Figure 9: Cost of *protective* lifestyle across cohorts



Notes: In Panel (a), the red shaded area describes the density function of τ_{PRO} for individuals in the 1930s 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 1930s cohort. The dashed green vertical line reports this same threshold if the income prospects of HSD individuals were as in the 1970s cohort. The dashed shaded blue area describes the density function of τ_{PRO} for individuals who choose HSD education in the 1970s cohort. Panel (b) reports the same information for CG individuals.

direct effect of diverging income prospects across education groups and how much is due to the induced changes in the composition of individuals in each education group.

To quantify the direct effect of income changes, we fix the distributions of lifestyle costs conditional on education to the ones for the 1930s cohort. We then analyze how investments in lifestyle would have changed solely due to changes in income prospects across education groups. Recall that in our calibration, expected lifetime earnings increase by 18% for college graduates and decrease by 12% for high school dropouts. In Figure 9, the shaded red areas represent the distributions of the cost of *protective* lifestyle τ_{PRO} faced by individuals in the 1930s cohort 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 lines indicate the threshold values $\tau^*(\text{HSD})$ and $\tau^*(\text{CG})$ that separate individuals of that cohort into *protective* and *detrimental* lifestyles conditional on each education choice. The changes in income prospects are represented by the dashed green vertical lines. For high school dropouts, the threshold value $\tau^*(\text{HSD})$ decreases across cohorts due to the decline in expected labor earnings, 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 among high-school dropouts would drop by 5 percentage points in this counterfactual exercise (compared to a 13 percentage points decline predicted by the model), while their life expectancy would drop by 0.3 years (compared to a 0.9 years decline predicted by the model). The opposite happens for college graduates, with the share of *protective* individuals increasing by 1 percentage point and life expectancy increasing by 0.1 years in this counterfactual (which are the full increases predicted by the model). Overall, the direct effect of increasing differences in income prospects across education groups accounts for 40% of the increase in the education gradient in life expectancy between the 1930s and 1970s cohorts as predicted by the model (and 20% of the increase compared to data).

Table 6: Decomposition: changes across cohorts

	Δ_cPr(e)		Δ_cPr(PRO)	Δ_cPr(PRO|e)			Δ_cLE(e)		
	CG	HSD		CG	HSD	CG-HSD	CG	HSD	CG-HSD
Data	0.12	-0.25	0.04	0.08	-0.18	0.26	0.8	-1.2	2.0
Model	0.12	-0.24	0.02	0.01	-0.13	0.14	0.1	-0.9	1.0
Income				0.01	-0.05	0.06	0.1	-0.3	0.4
Selection				0.00	-0.08	0.08	0.0	-0.6	0.6

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

Next, selection into each education category changes with income changes for two reasons. First, the increase in the education wage premium shrinks the group of high school dropouts and increases the size of the other two. This leads to a worsening of the distribution of τ_{PRO} in all education groups, as the marginal individuals who chose HSD in the 1930s cohort chose higher education in the 1970s. Second, the increase in income inequality across education groups increases the complementarity between education and lifestyles, which leads to a worsening of the distribution of τ_{PRO} among individuals choosing HSD and an improvement among individuals choosing CG. The overall effect on the conditional distributions of τ_{PRO} can be observed in Figure 9, where the shaded blue areas in each panel represent the conditional distributions of τ_{PRO} for the 1970s cohort. The distribution of τ_{PRO} worsened for high school dropouts (it shifted to the right as both effects are pushing in the same direction), while it did not change significatively for the college graduates (as each effect pushes the distribution of τ_{PRO} in opposite directions). Consequently, selection changes imply that the share of *protective* individuals would have fallen by 8 percentage points among high school dropouts (compared to 14 in the model), leading to a drop of 0.6 months in their life expectancy (0.9 in the model). With no changes for college graduates, the education gradient in life expectancy would have increased by the same 0.6 years or 60% of the increase predicted by the model and 30% 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. However, 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 1930s and 1970s cohorts could be explained by changes in perceptions.

Modeling. We introduce a cohort-specific parameter $\lambda^{\mathrm{c}} \in [0,1]$, capturing the degree to which individuals understand the link between a given lifestyle and its associated health dynamics. The objective function in the second-stage worker's problem of Section 5.2.1 is now expressed as:

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

where

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

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

As shown in equations (6)-(8), λ^{c} represents the probability associated with the continuation value $W^{\mathrm{c,e,y}}$, which is built with the correct 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 does not depend 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 1970s cohort) and re-estimate the whole model for the 1930s cohort, including the extra parameter λ^{1970}. This implies re-estimating the second-stage parameters $(\beta, \underline{x}, b_0, b_1, \bar{b})$ to match the life cycle wealth profiles and all the first-stage parameters to match the joint distribution of education and lifestyles for the 1930s and 1970s cohorts.[24] We find $\lambda^{1930} = 0.60$, which states that young individuals in the 1930s cohort only understood a little more than 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 F.3 in the Appendix) 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 2.23 in the benchmark model to -0.77 in the model with imperfect information. This happens because, given that the 1930s cohort only gives 60% 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 2.86 to 7.18. 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

[24] 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 the 1930s and 1970s. Furthermore, we calibrate $\Gamma_t^e(h'|h)$ and $s_t^e(h)$ as the average health transitions and survival for individuals in the 1930s 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(e)		Δ_cPr(PRO)	Δ_cPr(PRO|e)			Δ_cLE(e)		
	CG	HSD		CG	HSD	CG-HSD	CG	HSD	CG-HSD
Data	0.12	-0.25	0.04	0.08	-0.18	0.26	0.80	-1.20	2.00
Model	0.14	-0.31	0.10	0.10	-0.15	0.25	0.90	-1.00	1.90
Income				0.01	-0.01	0.02	0.10	-0.10	0.20
Selection				0.00	-0.18	0.18	0.00	-1.20	1.20
Perceptions				0.09	0.09	0.00	0.80	0.60	0.30

Notes: The first row reports statistics from the data, comparing the cohorts of the 1930s and 1970s. The second row reports the results from the calibrated model. The 3rd performs the counterfactual exercise where only income prospects change between the 1930s and 1970s; the 4th row performs the counterfactual exercise where only the distribution of τ_{PRO} conditional on education changes between the 1930s and 1970s. The last row increases the perceived probability of lifestyles affecting health dynamics. See Section 7.3 for details.

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 F.4 in the Appendix) improves the benchmark model in its ability to match the change in the education gradient of lifestyles across cohorts. 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 in the data, 14 percentage points in the benchmark model, and 25 percentage points in the model with a change in perceptions. This translates into the model being able to produce 1.9 of the 2-year increase in the education gradient of life expectancy seen in the data.

Results. We decompose the increase in the education gradient of life expectancy into three different channels. First, we give to the 1930s cohort full information about the consequences of lifestyle for health dynamics, that is, we set $\lambda^{1930} = \lambda^{1970} = 1$. The change in perceptions improves behavior in all education groups but does not change the education gradient (see the last row in Table 7). However, because of the complementarity of education and lifestyle, there are larger life expectancy gains among the college-educated, leading to an increase in the life expectancy gradient of 0.3 years. Next, compared to the benchmark model, the changes in income prospects have a lower effect on the increases in the lifestyle and life expectancy gradients: 2 percentage points and 0.2 years, respectively (compared to 6 percentage points and 0.4 years). 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: it accounts for an 18-percentage-point increase in the gradient of lifestyles (8-percentage-point increase in the benchmark model) and a 1.2-year increase in the gradient of life expectancy (0.6 in the benchmark model). This is because changes in perceptions increase the complementarities between education and lifestyle investments, which are central to selection. One of the reasons complementarities emerge in our model is that protective lifestyles improve health transitions more for individuals with higher education. When individuals believe that lifestyles have little impact on future health, these complementarities are small, but they increase as newer cohorts have a better understanding of their effects.

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*. More educated individuals choose *protective* behavior more frequently, and differences in lifestyles explain almost 1/2 of the education gradient in life expectancy. We also find an increasing education gradient in life expectancy between the 1930s and 1970s cohorts, which is mostly driven by worsening lifestyles among the less 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 choose their education and lifestyle, with the value of each choice given by the second stage of the model. In the second stage, agents solve a consumption-savings problem subject to income and health risks, as modeled in the econometric framework. 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 by the 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. Quantitatively, we show the three mechanisms to matter.

Finally, the model explains 50% of the increase in health inequalities between the 1930s and the 1970s cohorts, with 40% due to the direct effect of worsening economic conditions for the less educated and 60% due to the induced selection effects. If something, our robustness exercises increase the relative importance of selection. This has clear policy implications: early childhood interventions to foster the adoption of better lifestyles (like healthy food at school or exercise habits) may be more effective to reduce health inequalities than labor market interventions to reduce wage inequality. Furthermore, this type of policy can also increase the education outcomes of the population, see for instance Bagues and Villa (2025).

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 driverss 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. A more detailed treatment of changes in perceptions should take into account their correlation with education, see Cutler and Lleras-Muney (2010). All these avenues offer promising directions for future research.

References

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 pages 4 and 22.)

Argentieri, M. A., N. Amin, A. J. Nevado-Holgado, W. Sproviero, J. A. Collister, S. M. Keestra, M. M. Kuilman, B. N. R. Ginos, M. Ghanbari, A. Doherty, D. J. Hunter, A. Alvergne, and C. M. van Duijn (2025): "Integrating the Environmental and Genetic Architectures of Aging and Mortality," *Nature Medicine*. (Cited on page 14.)

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 4.)

Bagues, M. and C. Villa (2025): "Minimum Legal Drinking Age and Educational Outcomes," CEPR Discussion Paper 19733. (Cited on pages 16 and 33.)

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 4.)

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 4 and 6.)

Becker, G. S. (2007): "Health as Human Capital: Synthesis and Extensions," *Oxford Economic Papers*, 59, 379–410. (Cited on pages 2 and 26.)

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 4.)

Borella, M., F. Bullano, M. De Nardi, B. Krueger, and E. Manresa (2024): "Health Inequality and Health Types," NBER Working Paper 32799. (Cited on page 3.)

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 4.)

Bueren, J. (2023): "Long-Term Care Needs and Savings in Retirement," *Review of Economic Dynamics*, 49, 201–224. (Cited on page 4.)

Capatina, E. (2015): "Life-cycle Effects of Health Risk," *Journal of Monetary Economics*, 74, 67–88. (Cited on page 4.)

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 13.)

Card, D., C. Dobkin, and N. Maestas (2009): "Does Medicare Save Lives?" *Quarterly Journal of Economics*, 124, 597–636. (Cited on page 4.)

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 pages 4 and 27.)

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 E. Glaeser (2005): "What Explains Differences in Smoking, Drinking, and Other Health Related Behaviors?" *American Economic Review*, 95, 238–242. (Cited on page 6.)

Cutler, D. M. and A. Lleras-Muney (2010): "Understanding Differences in Health Behaviors by Education," *Journal of Health Economics*, 29, 1–28. (Cited on pages 1, 6, 14, and 33.)

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 4.)

——— (2016): "Medicaid Insurance in Old Age," *American Economic Review*, 106, 3480–3520. (Cited on page 4.)

De Nardi, M., S. Pashchenko, and P. Porapakkarm (2025): "The Lifetime Costs of Bad Health," Forthcoming *Review of Economic Studies*. (Cited on pages 3, 4, 22, and A.14.)

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society: series B (methodological)*, 39, 1–22. (Cited on page 9.)

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 4.)

Foltyn, R. and J. Olsson (2024): "Subjective life expectancies, time preference heterogeneity, and wealth inequality," *Quantitative Economics*, 15, 699–736. (Cited on page 18.)

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 4.)

French, E. and J. Jones (2011): "The Effects of Health Insurance and Self-Insurance on Retirement Behavior," *Econometrica*, 79, 693–732. (Cited on page 4.)

Fuchs, V. (1982): "Time Preference and Health: an Exploratory Study," in *Economic Aspects of Health*, ed. by V. Fuchs, Chicago: University of Chicago Press, 93–120. (Cited on page 14.)

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 11.)

Hai, R. and J. Heckman (2022): "The Causal Effects of Youth Cigarette Addiction and Education," NBER Working Paper 30304. (Cited on pages 4 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 13.)

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 (2025): "Health, Consumption, and Inequality," Mimeo. (Cited on pages 3, 4, and 22.)

Hosseini, R., K. A. Kopecky, and K. Zhao (2022): "The Evolution of Health over the Life Cycle," *Review of Economic Dynamics*, 45, 237–263. (Cited on page 3.)

——— (2025): "How Important is Health Inequality for Lifetime Earnings Inequality?" Forthcoming *Review of Economic Studies*. (Cited on pages 2, 4, and 19.)

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

Kaplanis, J., A. Gordon, T. Shor, O. Weissbrod, D. Geiger, M. Wahl, M. Gershovits, B. Markus, M. Sheikh, M. Gymrek, G. Bhatia, D. G. MacArthur, A. L. Price, and Y. Erlich (2018): "Quantitative Analysis of Population-Scale Family Trees with Millions of Relatives," *Science*, 360, 171–175. (Cited on page 14.)

Karpati, D. (2023): "Does Life Expectancy Affect Wealth Accumulation? Evidence from Genetic Testing," Mimeo. (Cited on page 18.)

Keane, M. and K. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105, 473–522. (Cited on page 13.)

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 pages 1 and 6.)

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. (2005): "The Relationship Between Education and Adult Mortality in the United States," *Review of Economic Studies*, 72, 189–221. (Cited on page 15.)

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 11.)

Mahler, L. and M. Yum (2022): "Lifestyle Behaviors and Wealth-Health Gaps in Germany," *Available at SSRN 4034661*. (Cited on pages 4, 22, and 27.)

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 pages 4 and 27.)

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 (2025): "Medical Expenses and Saving in Retirement: The Case of U.S. and Sweden," *American Economic Journal: Macroeconomics*, 17, 161–202. (Cited on page 4.)

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 5.)

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 4.)

Pijoan-Mas, J. and J.-V. Ríos-Rull (2014): "Heterogeneity in Expected Longevities," *Demography*, 51, 2075–2102. (Cited on pages 1 and 8.)

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.)

Ruby, J. G., K. M. Wright, K. A. Rand, A. Kermany, K. Noto, D. Curtis, N. Varner, D. Garrigan, D. Slinkov, I. Dorfman, J. M. Granka, J. Byrnes, N. Myres, and C. Ball (2018): "Estimates of the Heritability of Human Longevity Are Substantially Inflated due to Assortative Mating," *Genetics*, 210, 1109–1124. (Cited on page 14.)

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 13.)

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:

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

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 $p\left(\mathrm{y}|\mathrm{c}_i, \mathrm{s}_i, \mathrm{e}_i, a_{i0}, h_{i0}; \boldsymbol{\beta}, \boldsymbol{\delta}\right)$ 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,

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

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

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

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

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

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.[25] The first term on 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.

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

A.4 Gibbs sampler

Estimating the mixture model poses a challenge because the health behavior type, y, which classifies individuals into different lifestyle categories, is unobserved. This means that the likelihood function must account for all possible assignments of y, making direct estimation computationally complex. A particularly effective solution in a Bayesian framework is *data augmentation*, which introduces y as an auxiliary variable and allows us to estimate the model using Markov Chain Monte Carlo (MCMC) methods.

The key idea behind data augmentation is to treat y as missing data and iteratively infer it alongside the model parameters. In a Bayesian setting, we sample from the posterior distribution of y given the observed health behaviors and health outcomes. This imputed dataset is then used to update the posterior distributions of the model parameters. Specifically, given prior distributions for the parameters, we construct a Gibbs sampler where we alternate between:

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

i) drawing lifestyle assignments y_i from their conditional posterior distribution, and

ii) updating the parameters given the imputed values of y

We can write the complete data likelihood as:

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

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 the joint posterior distribution as:

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

To estimate the parameters of the model, we implement a Gibbs sampler that sequentially updates different sets of parameters by drawing from their respective conditional posterior distributions. The sampler consists of four blocks: (i) *Transition parameters*, which govern the evolution of health status over time and are updated based on the observed health trajectories given the lifestyle; (ii) *Health behavior parameters*, which capture the relationship between lifestyles and observed health behaviors; (iii) *Mixture weights*, which determine the prior probability of assignment to each lifestyle type given individual characteristics; and (iv) *Latent lifestyle*, where we impute the unobserved health behavior type y for each individual based on the current parameter estimates. By iterating over these blocks, the algorithm ensures convergence to the joint posterior distribution of the model parameters and latent variables.

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:

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

A.5 Wealth distribution conditional on age, education, and type

In estimating the wealth distribution across different lifestyle types, we face the challenge that the lifestyle type, y, is unobserved. This makes it impossible to directly model the wealth distribution conditional on lifestyle. However, we can circumvent this issue by using the estimated mixture model. In our econometric model, we recovered the posterior probabilities (mixture weights) for each individual's membership in a particular lifestyle category. These mixture weights represent the likelihood that an individual belongs to each lifestyle type, given their observed characteristics. We use these mixture weights to model the wealth distribution across lifestyles and separate it into two components: the mass at zero wealth and the distribution of wealth conditional on positive wealth.

We, thus 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.[26]

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.

[26]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 48% of the education gradient of life expectancy. In this Appendix, we extend the model by estimating it with three and four groups.

When estimating the model with three groups, the estimation keeps the *protective* group more or less intact (*group 1*) and splits the *detrimental* lifestyle into two subgroups. These groups are very similar to each other in terms of behavior, including the incidence of smoking while young. However, as they age, individuals in the group that we call *group 2* tend to quit smoking, while individuals in the group that we call *group 3* do not, see Figure B.1. This classification increases the difference in life expectancy between the best and worst lifestyles (*group 1* and *group 3*) to 10.2 years (well above the 8.5 in the two-group classification), see Table B.1. Furthermore, the share of the education gradient in life expectancy explained by the different incidence of lifestyles across education groups rises to 63% (above the 48% in the two-group partition).

Next, we estimate the model with four groups. In this case, it is the *protective* lifestyle that gets split into two different subgroups. The main difference between the two subgroups is that one of them, *group 1*, does more preventive health care (cancer tests, cholesterol tests, flu shots) while the other, *group 2*, tends to do more exercise when young (age 50), see Figure B.2. This generates 6 years of life expectancy difference between the two groups, and 13.6 years between *group 1* and *group 4*, see Table B.2. This large impact of lifestyles, together with their education gradient, implies that 3/4 of the education gradient of life expectancy is explained by differences in lifestyles.

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

	All		HSD		HSG		CG		Δ_eLE (CG-HSD)		
	%	LE	%	LE	%	LE	%	LE	Data	(a)	(b)
All	100.0	29.9	100.0	25.3	100.0	28.5	100.0	33.2	7.9	2.9	5.0
Group 1	69.2	32.2	36.7	31.0	61.1	30.8	90.3	33.7			
Group 2	11.0	29.4	14.4	26.8	14.5	29.2	5.0	32.5			
Group 3	19.8	22.0	48.9	20.6	24.4	22.3	4.7	23.8			

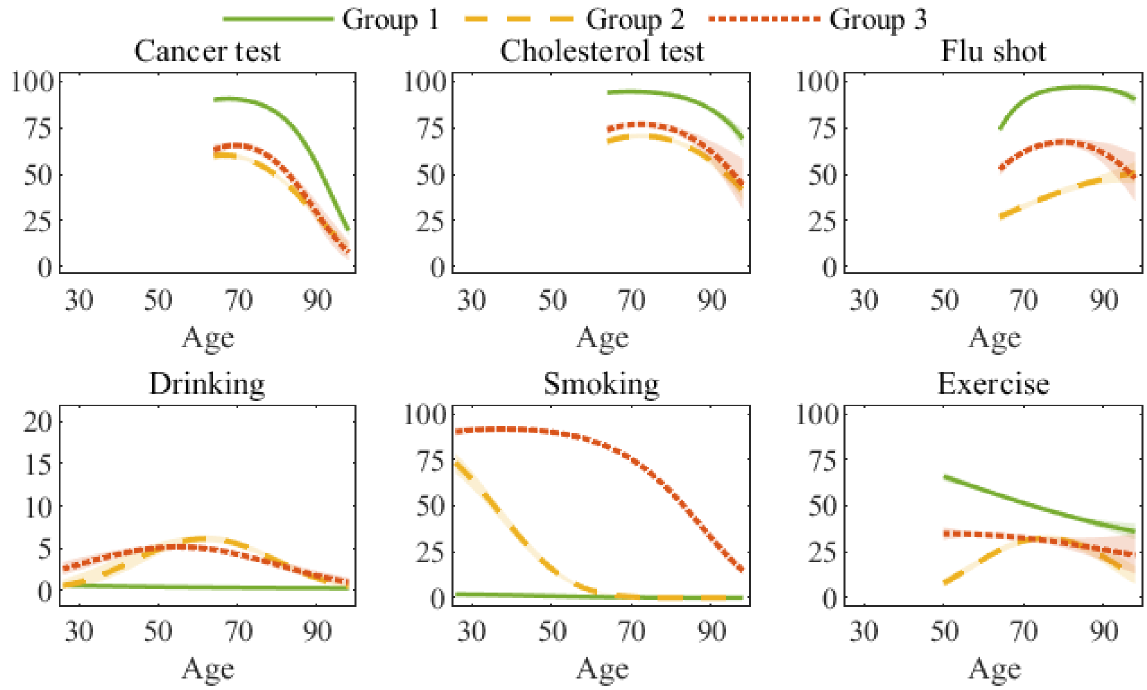
Notes: See footnote in Table 3.

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

	All		HSD		HSG		CG		Δ_eLE (CG-HSD)		
	%	LE	%	LE	%	LE	%	LE	Data	(a)	(b)
All	100.0	30.3	100.0	25.1	100.0	28.7	100.0	34.1	9.0	2.1	6.9
Group 1	47.8	35.5	17.2	34.1	39.2	34.6	69.0	36.2			
Group 2	19.9	29.0	18.1	29.8	20.6	27.4	19.5	31.1			
Group 3	12.4	26.2	14.8	24.4	15.6	26.0	7.2	27.9			
Group 4	19.9	21.9	49.9	20.5	24.6	22.2	4.3	23.8			

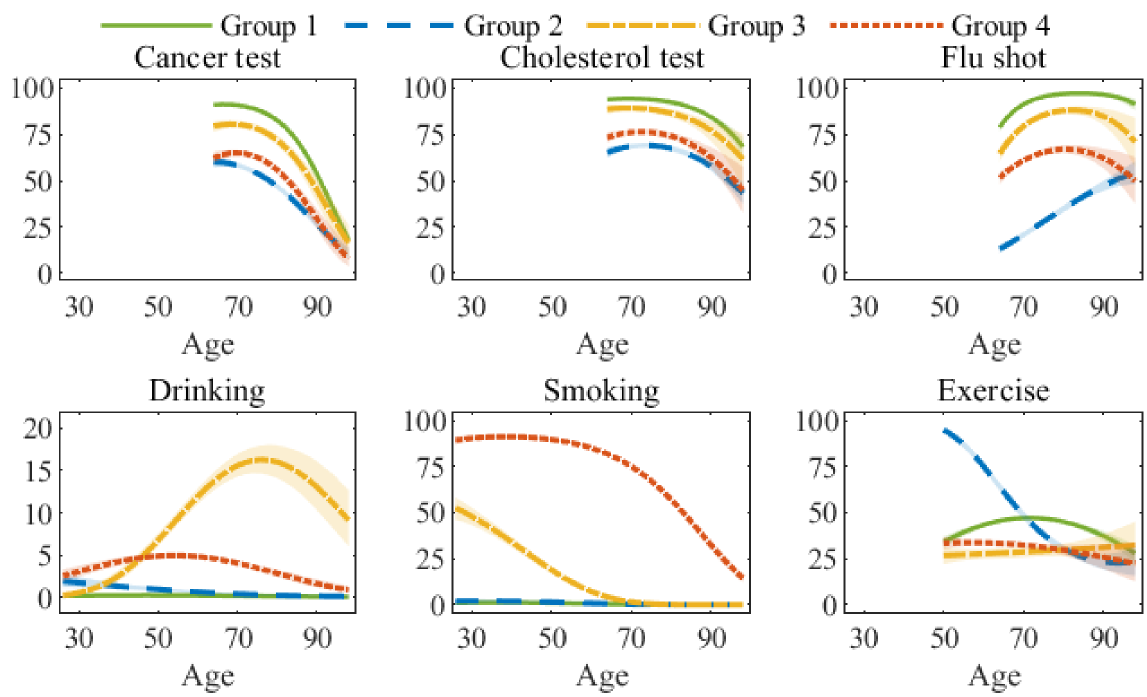
Notes: See footnote in Table 3.

Figure B.1: Health behaviors and lifestyles (3 groups)



Notes: Estimation results. Probability of engaging in each health behavior by age and type, for male individuals in good health. The shaded areas represent the 95% credible intervals.

Figure B.2: Health behaviors and lifestyles (4 groups)



Notes: Estimation results. Probability of engaging in each health behavior by age and type, for male individuals in good health. The shaded areas represent the 95% credible intervals.

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

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

(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 $\square$

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 $\tau^*_{\text{CG}}(\tau^a_{\text{PRO}}) \leq \tau^*_{\text{CG}}(\tau^b_{\text{PRO}})$, with strict inequality for some pairs $\left(\tau^a_{\text{PRO}}, \tau^b_{\text{PRO}}\right)$. Therefore $\Pr(\text{CG}|\tau_{\text{PRO}}) \equiv \Pr(\tau_{\text{CG}} \leq \tau^*_{\text{CG}}(\tau_{\text{PRO}}))$ is weakly decreasing with τ_{PRO}, and so is $\Pr(\text{CG}|\tau \leq \tau_{\text{PRO}}) = \int^{\tau_{\text{PRO}}} \Pr(\text{CG}|\tau) \frac{f(\tau)}{F(\tau_{\text{PRO}})} d\tau$.[27] Using Bayes theorem, we have that

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

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

$$\frac{\Pr(\text{CG}|\tau \leq \tau_{\text{PRO}})}{\Pr(\text{CG})} \geq 1$$

For high school dropouts we have equivalently:

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

thus:

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

with strict inequality for some τ_{PRO}

Finally, note that whenever we are in the case of independent choices, Lemma 1 states that $\tau^*_{\text{CG}}(\tau_{\text{PRO}})$ is just a constant independent from τ_{PRO}. Hence, $\Pr(\text{CG}|\tau_{\text{PRO}})$ is also independent from τ_{PRO} and we have $F(\tau_{\text{PRO}}|\text{HSD}) = F(\tau_{\text{PRO}}) = F(\tau_{\text{PRO}}|\text{CG}) \;\; \forall \tau_{\text{PRO}}$. $\square$

[27] To see this last part, note that $\frac{\partial \Pr(\text{CG}|\tau \leq \tau_{\text{PRO}})}{\partial \tau_{\text{PRO}}} = \Pr(\text{CG}|\tau_{\text{PRO}}) - \Pr(\text{CG}|\tau \leq \tau_{\text{PRO}}) < 0$, 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}.

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. (2025) 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 5.2.1 we can obtain the total differential:

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

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

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

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 $0.01 \times dx/ds_t^{\mathrm{e,y}}(h)$ units of income. Putting 100 identical agents together, we would have one death on average in exchange for $dx/ds_t^{\mathrm{e,y}}(h)$ units of income.

Appendix E: Robustness exercises

In this appendix, we consider two different model robustness exercises.

Correlated heterogeneity. In the benchmark model, the unobserved costs of investing in education and in good lifestyles are uncorrelated. The model endogenously delivers a correlation of these costs conditional on choices. As a robustness exercise, we allow for the initial distribution of ϵ_{PRO} and ϵ_e to be correlated. In particular, we impose $\text{corr}(\epsilon_{\text{PRO}}, \epsilon_e) = 0.25$ and reestimate the model. The model fit changes very little, with the most noticeable difference being in the model's slightly lower ability to reproduce the fall in the share of HSD between the 1930s and 1970s cohorts, see Figure E.1. In terms of the recovered first-stage parameters, the model with correlated heterogeneity requires a larger dispersion of unobserved heterogeneity (larger σ_{HSG}, σ_{CG}, and σ_{PRO}) and lower average cost τ_{PRO} of *protective* lifestyle (see Panel B in Table E.1). The calibration recovers more dispersion in unobserved heterogeneity because the correlation between ϵ_{PRO} and ϵ_e is an additional force correlating education and lifestyle choices. Hence, in order to match the same data, the estimation requires de-emphasizing the role of income differences across education groups, which is achieved precisely through the increase in dispersion in unobserved heterogeneity. In particular, in the main model, the income gap between college-educated workers and high-school dropouts leads to a 1.3-year gap in life expectancy through different lifestyle choices, while in the model with correlated heterogeneity, the income gap across education groups only explains 0.4 years, see Table E.2. In contrast, and by construction, selection becomes more important as individuals with lower costs of going to college are assumed to also have lower costs of investing in *protective* lifestyle. In particular, the role of selection increases from 1 year in the benchmark model to 2.3 years in the model with correlated heterogeneity. Regarding changes across cohorts, we observe the same pattern of lower importance for the direct effect of the increase in wage inequality, see Table E.3.

Higher VSL. The range of estimates for the VSL is relatively large. In the benchmark calibration, we have selected a VSL of 1 million dollars, which is at the low end of the estimates. As an alternative, in this Section, we re-estimate our model with a VSL of 4 million dollars. The model fit, shown in Figure E.2, remains similar to the benchmark calibration but overestimates the decline in high school dropout rates between the 1930s and 1970s cohorts. With a higher VSL, each year of life is more valuable, so the estimated average cost of adopting a protective lifestyle must increase to match the observed share of individuals choosing this lifestyle (see Table **??**). Moreover, since the effect of a protective lifestyle on survival is stronger for highly educated individuals, the contribution of the health effect to explaining the life expectancy gradient becomes more pronounced. Specifically, the health effect increases from 1.0 years with a VSL of 1 million to 1.7 years with a VSL of 4 million, while the income effect diminishes from 1.4 to 0.7 years (see Table E.4). Additionally, a higher VSL strengthens the complementarity between health and education, leading to a more pronounced selection effect. With a VSL of 4 million, selection explains 1.7 years of the LE

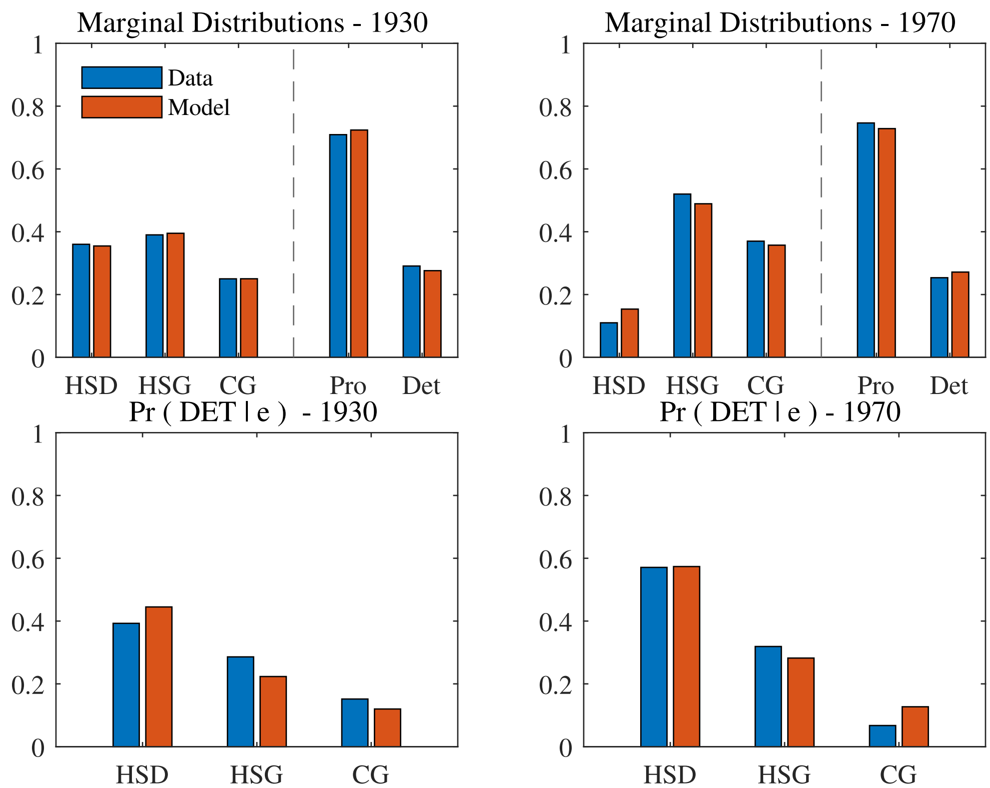
Table E.1: Calibrated parameters: robustness checks

Parameter	Description	Value	
		Corr. hetero.	Larger VSL
Panel A: Adult Life			
β	Discount factor	0.952	0.952
$\underline{x}$	Income floor	17.97	17.97
b_0	Bequest motive: marginal utility	15.11	15.11
b_1	Bequest motive: non-homoteticity	296.36	296.36
$\bar{b}$	Value of life	3.44	10.55
Panel B: Early Life			
μ_{HSG}	Average cost of HSG education	4.68	7.00
μ_{CG}	Average cost of CG education	23.07	26.05
μ_{PRO}	Average cost of PRO lifestyle	-1.36	9.87
σ_{HSG}	Sd. cost of HSG education	2.42	0.23
σ_{CG}	Sd. cost of CG education	21.11	5.49
σ_{PRO}	Sd. cost of PRO lifestyle	7.84	3.98

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. The first column refers to the model parameters for the model with correlated unobserved heterogeneity. The second column refers to the model parameters for the calibration with a higher VSL.

gradient, compared to 1.0 year in the benchmark case. When comparing across cohorts, the model is able to explain a slightly larger fraction of the observed increase in the life expectancy gradient (1.1 instead of 1.0 years). As the value of life increases, the direct effect of income on lifestyle and life expectancy differences diminishes (0.1 instead of 0.4 years).

Figure E.1: Model Fit. First-Stage (correlated unobserved heterogeneity)



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.

Table E.2: Decomposition: 1970s cohort (correlated unobserved heterogeneity)

	Pr(e)		Pr(PRO)	Pr(PRO|e)			LE(e)		
	CG	HSD		CG	HSD	CG-HSD	CG	HSD	CG-HSD
Data	0.37	0.11	0.75	0.93	0.43	0.50	32.8	24.8	8.0
Model	0.36	0.15	0.73	0.87	0.43	0.44	32.2	24.8	7.4
Same lifestyle				0.87	0.87	0.00	32.2	27.7	4.5
Income				0.87	0.51	0.37	32.3	25.3	7.0
Health				0.87	0.49	0.39	32.2	25.2	7.1
Income + Health				0.87	0.58	0.29	32.3	25.8	6.5
Selection				0.87	0.79	0.09	32.2	27.2	5.1

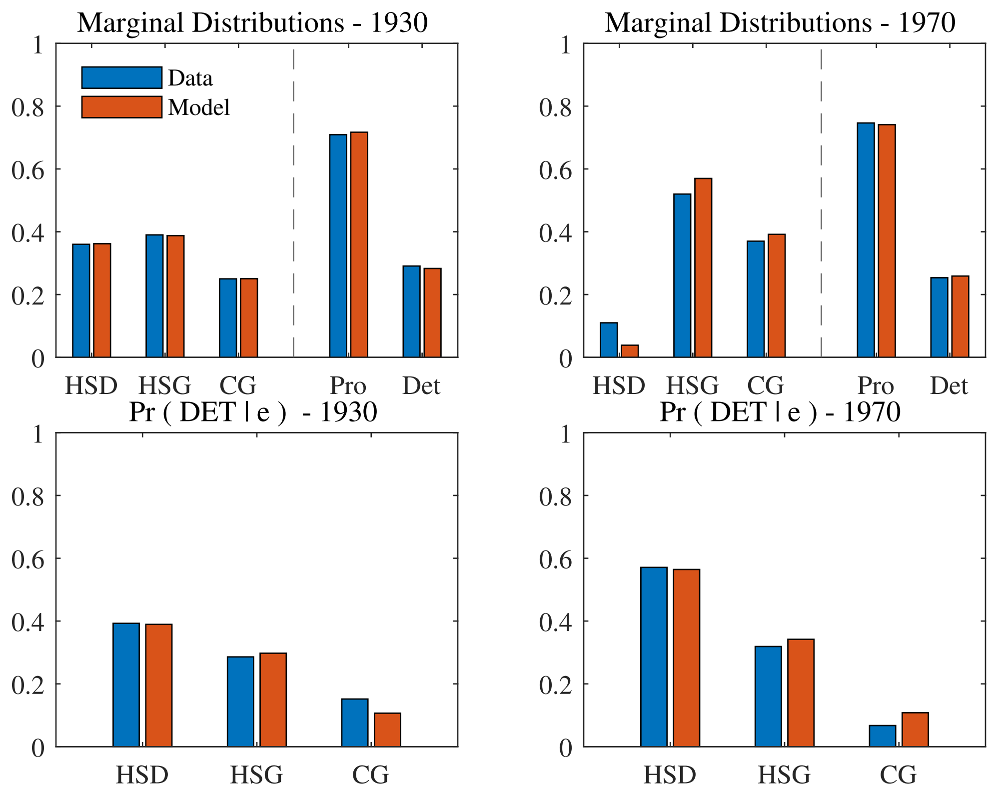
Notes: See footnote in Table 5.

Table E.3: Decomposition: changes across cohorts (correlated unobserved heterogeneity)

	Δ_cPr(e)		Δ_cPr(PRO)	Δ_cPr(PRO|e)			Δ_cLE(e)		
	CG	HSD		CG	HSD	CG-HSD	CG	HSD	CG-HSD
Data	0.12	-0.25	0.04	0.08	-0.18	0.26	0.8	-1.2	2.0
Model	0.11	-0.20	0.00	-0.01	-0.13	0.12	-0.1	-0.8	0.8
Income				0.00	-0.02	0.02	0.0	-0.1	0.1
Selection				-0.01	-0.11	0.10	-0.1	-0.7	0.6

Notes: See footnote in Table 6.

Figure E.2: Model Fit. First-Stage (larger VSL)



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.

Table E.4: Decomposition: 1970s cohort (larger VSL)

	Pr(e)		Pr(PRO)	Pr(PRO|e)			LE(e)		
	CG	HSD		CG	HSD	CG-HSD	CG	HSD	CG-HSD
Data	0.37	0.11	0.75	0.93	0.43	0.50	32.8	24.8	8.0
Model	0.39	0.04	0.74	0.89	0.44	0.45	32.4	24.9	7.6
Same lifestyle				0.89	0.89	0.00	32.4	27.8	4.6
Income				0.89	0.53	0.36	32.4	25.5	6.9
Health				0.89	0.69	0.20	32.4	26.5	5.9
Income + Health				0.89	0.76	0.13	32.4	27.0	5.4
Selection				0.89	0.70	0.20	32.4	26.6	5.9

Notes: See footnote in Table 5.

Table E.5: Decomposition: changes across cohorts (larger VSL)

	Δ_cPr(e)		Δ_cPr(PRO)	Δ_cPr(PRO|e)			Δ_cLE(e)		
	CG	HSD		CG	HSD	CG-HSD	CG	HSD	CG-HSD
Data	0.12	-0.25	0.04	0.08	-0.18	0.26	0.8	-1.2	2.0
Model	0.14	-0.32	0.02	0.00	-0.17	0.17	0.0	-1.1	1.1
Income				0.00	-0.01	0.02	0.0	-0.1	0.1
Selection				0.00	-0.15	0.15	0.0	-1.0	1.0

Notes: See footnote in Table 6.

Appendix F: Extra figures and tables

Table F.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 F.1: Health transitions across education and lifestyles

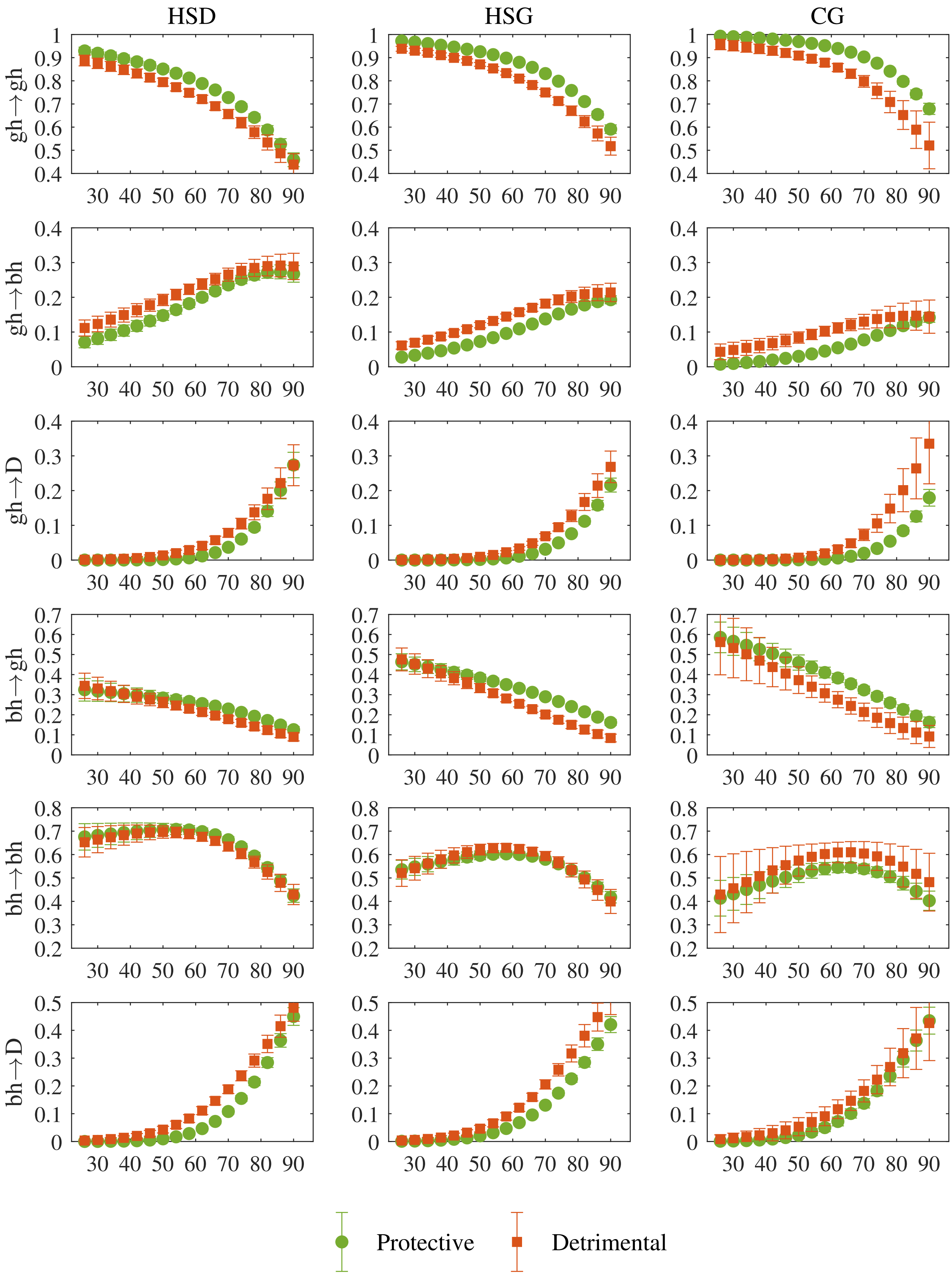
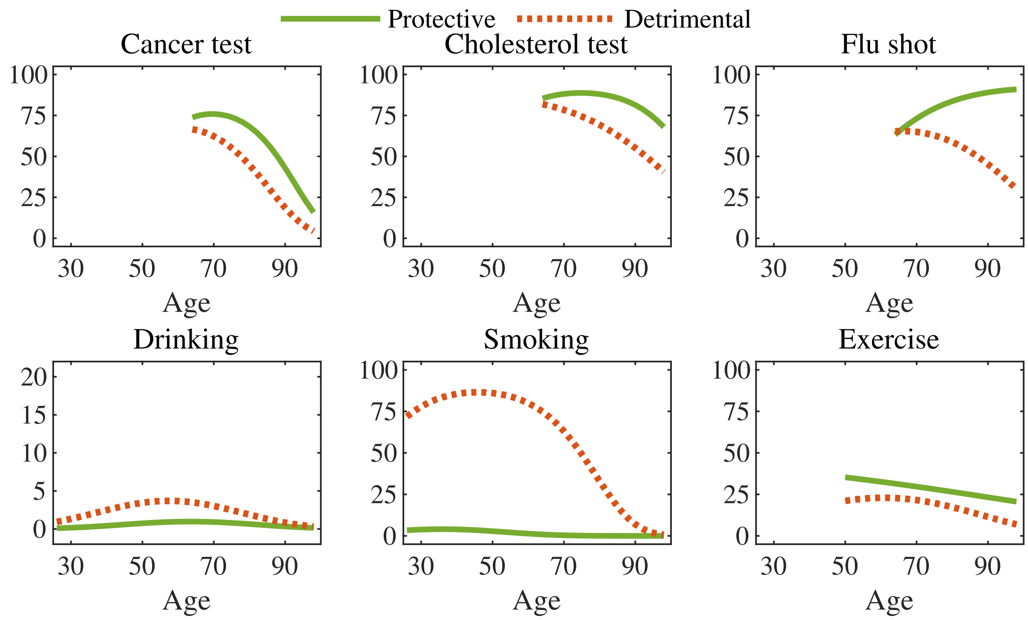
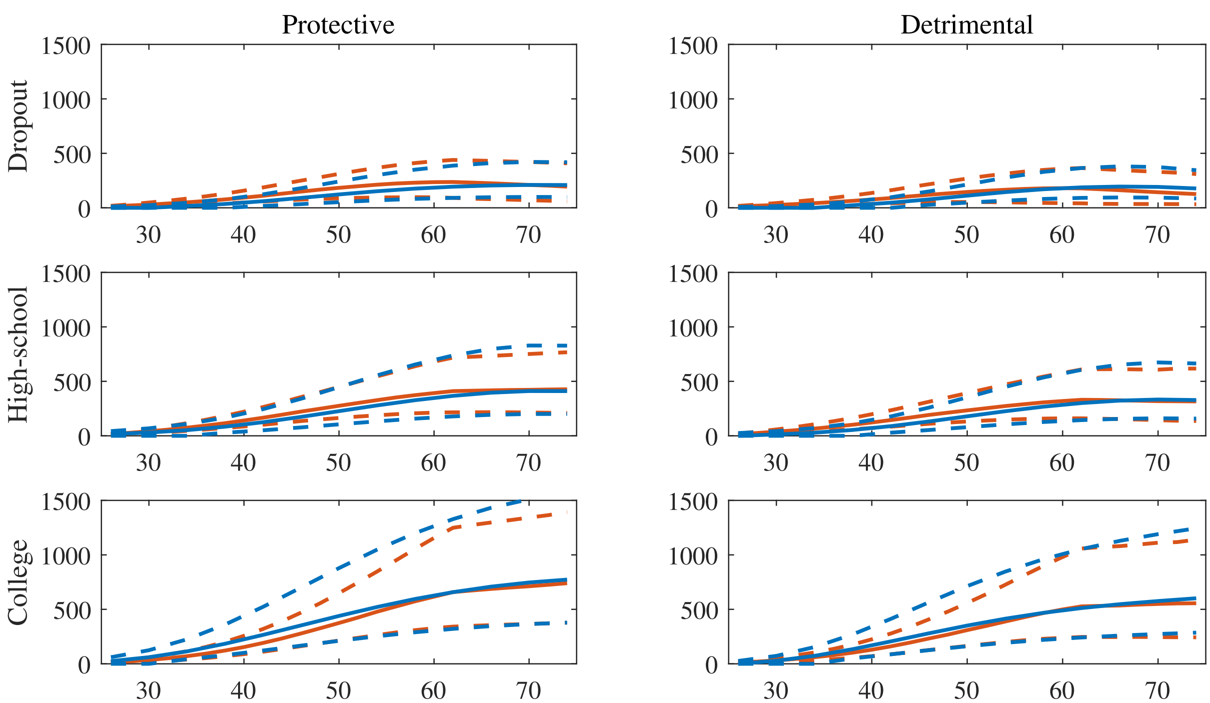


Figure F.2: Health behaviors and lifestyles: bad health



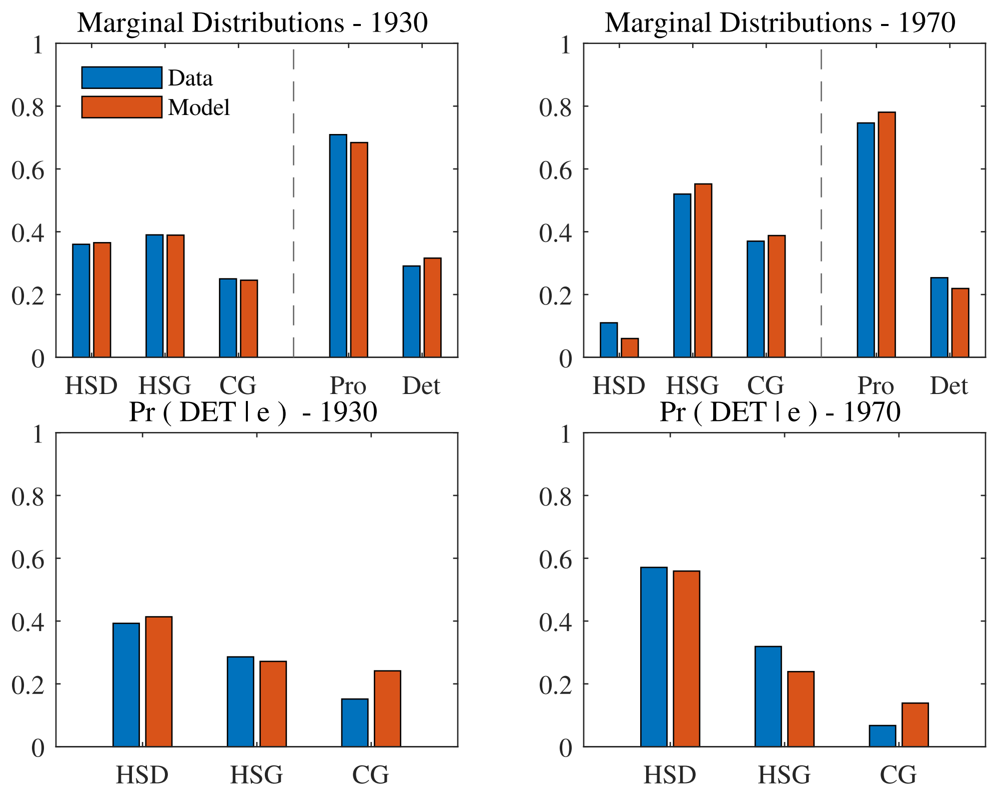
Notes: Estimation results. Probability of engaging in each health behavior by age and type, for male individuals in bad health.

Figure F.3: 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 F.4: 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.