

Healthy Habits and Inequality

Dante Amengual^{*}

Jesús Bueren[◇]

Josep Pijoan-Mas[□]

^{*}CEMFI

[◇]EUI

[□]CEMFI, CEPR

Preliminary

Introduction

- Health is an important determinant of earnings, survival and overall welfare.

Introduction

- Health is an important determinant of earnings, survival and overall welfare.
- Body of research documenting large inequalities in health outcomes across SES groups.

Introduction

- Health is an important determinant of earnings, survival and overall welfare.
- Body of research documenting large inequalities in health outcomes across SES groups.
- Identifying drivers of health might therefore be important to understand both health and economic inequality.

Introduction

- Health is an important determinant of earnings, survival and overall welfare.
- Body of research documenting large inequalities in health outcomes across SES groups.
- Identifying drivers of health might therefore be important to understand both health and economic inequality.
- By examining the role played by lifestyles, we aim to “*kill two birds with one stone*” and shed light on an important driver of both health and economic inequality.

Main ideas

- We conjecture that heterogeneity in lifestyles is an important driver of health inequality and thus also on economic inequality
 - Smoking, regular exercise, or healthy diet linked to better health outcomes
Zaninotto, Head, Steptoe (2020); Li et al (2018)
 - High SES tend to adopt healthier habits
Hiscock et al (2012); McLaren (2007)
- We seek to
 - Measure lifestyles and quantify their effects on health dynamics and economic outcomes
 - Understand the determinants of different lifestyle choices
 - Run counterfactuals to quantify how lifestyles affect the economic and welfare inequality in comparison with other policies in the literature.

What we do

1. Data

- We identify patterns in lifestyle (preventive tests, substance abuse, and obesity) driving health dynamics in both HRS and PSID
- We find that
 - Health behavior is well represented by three different lifestyles: **protective**, **detrimental**, and **harmful**
 - Large LE_{50} gradient: 9 years between **protective** and **harmful**
 - Healthy habits are correlated w/ education but carry independent information
 - Healthy habit gradient of similar size within education groups
 - **Harmful** types much more frequent among the less educated
 - Lifestyles explain 40% of the education gradient in LE_{50}

What we do

2. Model

- We build a life-cycle heterogeneous agents model with idiosyncratic labor market and health risks
 - Individuals differ in lifestyles and education as well as wealth and health
 - Health dynamics driven by previous estimates
 - Education and lifestyle choices taken together early in life
- Estimate the model
 - Understand sources of heterogeneity across individuals
- Counterfactuals
 - Understand role of lifestyles on economic inequality
 - Compare to changes in the progressivity in the tax system

The Data

- The HRS and PSID provide an unbalanced panel of individuals $i = 1, \dots, N$ followed for $t = 1, \dots, T$ periods
- Standard demographic information: gender (g), education (e), age (a_t)
- Wide array of information on **health status** and **health behavior**
 - Health state (h_t): self-reported health (good/bad) + death
 - Health behavior (z_{mt}):
 - 1 Preventive cancer tests (mammography / prostate check)
 - 2 Cholesterol test
 - 3 Flu shot
 - 4 Heavy drinking (2+ drinks on the day they drink)
 - 5 Smoking
 - 6 Obesity

Latent types

- We assume that observed health behavior (z_{mt}) is the result of some **latent time-invariant** factor (y)
 - The latent factor is represented by a few discrete groups $y \in \{y_1, y_2, \dots\}$.
- We interpret the latent factor (y) as the lifestyle
- We propose an econometric model exploiting both the cross-sectional and the time-series dimension of our data to
 - Allocate individuals to lifestyles
 - Measure the importance of lifestyles on health dynamics

Econometric Model

Overview

- We *jointly estimate* health dynamics and lifestyles using a mixture model:

$$\begin{aligned} p(\mathbf{z}, \mathbf{h} | c, s, e, a, h_0) &= \sum_{y \in Y} p(\mathbf{z}, \mathbf{h} | c, s, e, a, h_0, y) p(y | c, s, e, a, h_0) \\ &= \sum_{y \in Y} p(\mathbf{z} | \mathbf{h}, a, h_0, y) p(\mathbf{h} | s, e, a, h_0, y) p(y | c, s, e, a, h_0) \end{aligned}$$

- By estimating types and transition jointly, we find the types that better represent both the observed behaviour and the health transitions
(vs. k-means clustering on habits and then transitions)
- Conditional on the lifestyle (y)
 - Health behaviours (z_{mt}) are iid, modelled through a probit
 - Health outcomes (h_t) are first-order markovian, modelled through a nested probit

Econometric Model

1. Healthy Habits

- We model the probability of individual i of reporting the m 'th behaviour ($z_{mt} = 1$) at time t as a probit model.

- There is a latent variable (z_{mt}^*) that depends on type (y), age (a_t), health (h_t), and an idiosyncratic shock (ϵ_t)

$$z_{mt}^* = \gamma_{0,m,y} + \gamma_{1,m,y}a_t + \gamma_{2,m,y}a_t^2 + \gamma_{3,m,y}h_t + \epsilon_t, \quad \epsilon_t \sim N(0, 1)$$

- Then,

$$\text{Prob}(z_{mt} = 1) = \text{Prob}(z_{mt}^* > 0) = \underbrace{\alpha_m(y, a_t, h_t)}_{\alpha_{mt}}$$

- Considering independence of health behaviour given type, the probability of observing a sequence of health behaviours z for an individual across time, is assumed to be given by:

$$p(z|h, y) = \prod_{t=1}^T \prod_{m=1}^M \alpha_{mt}^{z_{mt}} (1 - \alpha_{mt})^{1-z_{mt}}$$

Econometric Model

2. Health Dynamics

- We model the probability of reporting some health $h' \in \{Good, Bad, Dead\}$ next period as a nested probit model
 - ① First nest: Alive/Dead
 - ② Second nest: Good/Bad cond on survival
- There are latent variables ($h_{h,h'}^*$) that depend on gender (g), education (e), type (y), health (h), age (a), and an idiosyncratic shock ($\epsilon_{h'}$)

$$h_{h,h'}^* = f(a, s, e, y; \beta_{h,h'}) + \epsilon_{h'}$$

with,

$$f(a, g, e, y; \beta_{h,h'}) = \beta_{0,y,e,g,h,h'} + \beta_{1,y,e,g,h,h'} a$$

- Then,

$$\begin{aligned} \text{Prob}(h' = Dead) &= \text{Prob}(h_{h,h'=Dead}^* > 0) \\ \text{Prob}(h' = Good | Survival) &= \text{Prob}(h_{h,h'=Good}^* > 0 | Survival) \end{aligned}$$

Econometric Model

3. Weights

- The mixture weights at the initial age (age 25) are modeled as a multinomial probit model:

$$\begin{aligned} y_1^* &= \lambda_{0,s,e,c}^1 + \lambda_{1,s,e}^1 bh + \epsilon_1 \\ &\vdots \\ y_Y^* &= \lambda_{0,s,e,c}^Y + \lambda_{1,s,e}^Y bh + \epsilon_Y, \end{aligned}$$

- We compute weights for future ages using the health transition model:

$$p(y, h_t | s, e, c) = \sum_{h_{t-1}} p(h_t | h_{t-1}, y, s, e, c) p(y, h_{t-1} | s, e, c)$$

Results: Lifestyles

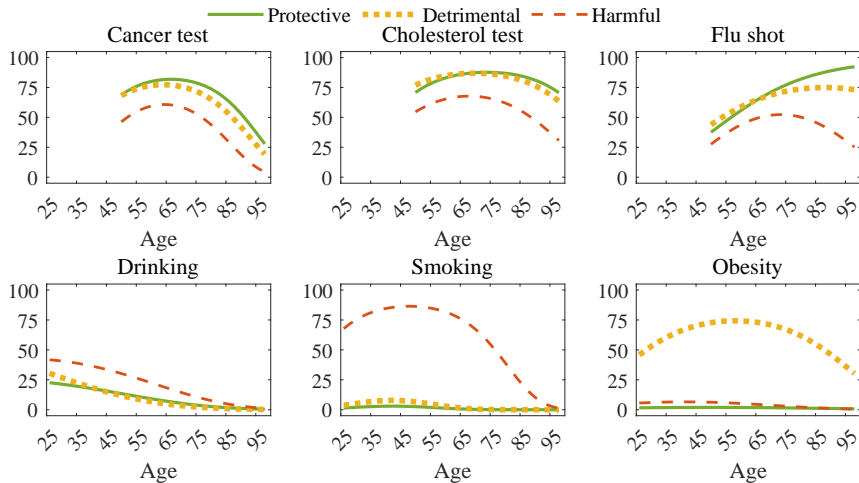


Figure: Probability of reporting health behaviors by lifestyle

Results: Mixture Weights

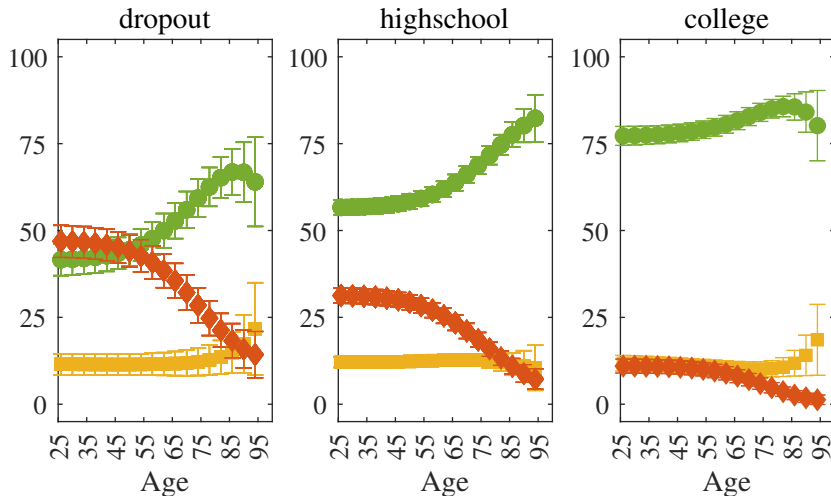


Figure: Probability of lifestyle. Males.

Results: Health Dynamics

Males

Table: Expected duration of each health state at age 50 across lifestyles

Lifestyle	Fraction	Life Expectancy	=	Good Health	+	Bad Health
Dropouts						
Protective	43.7	27.9		17.3		10.6
Detrimental	11.2	24.8		12.3		12.5
Harmful	45.1	19.3		10.7		8.6
Average	-	23.7		14.3		11.2
High-school						
Protective	57.9	29.5		23.9		5.6
Detrimental	12.4	27.3		18.4		8.9
Harmful	29.7	20.9		14.8		6.1
Average	-	26.7		21.2		6.6
College						
Protective	78.3	32.9		29.7		3.2
Detrimental	11.2	29.8		22.4		7.4
Harmful	10.5	22.5		17.9		4.6
Average	-	31.5		27.6		3.8

Results: Health Dynamics

Females

Table: Expected duration of each health state at age 50 across lifestyles

Lifestyle	Fraction	Life Expectancy	=	Good Health	+	Bad Health
Women: Dropouts						
Protective	43.0	29.4		15.9		13.6
Detrimental	27.1	27.1		9.9		17.2
Harmful	29.9	20.5		8.8		11.7
Average	-	26.1		12.2		14.1
Women: High-school						
Protective	56.7	32.9		26.9		6.1
Detrimental	21.6	29.3		18.7		10.6
Harmful	21.8	24.2		16.5		7.7
Average	-	30.2		22.9		7.4
Women: College						
Protective	73.9	34.8		31.0		3.8
Detrimental	17.0	30.1		23.7		6.4
Harmful	9.2	26.7		20.0		6.7
Average	-	33.3		28.4		4.5

Results: Health Dynamics

- More educated individuals tend to adopt healthier lifestyles.
 - The probability that a college male has a harmful lifestyle is 4.3 times smaller than a dropout.
- If dropout males had the same lifestyles than college males, their life expectancy would increase by 3 extra years.
 - This corresponds to 40% of the observed difference in life-expectancy at age 50 between college graduates males and high-school dropouts males.
- If dropout females had the same lifestyles than college females, their life expectancy would increase by 2.1 extra years.
 - This corresponds to 30% of the observed difference in life-expectancy at age 50 between college graduates females and high-school dropouts females.

Results: Cohorts

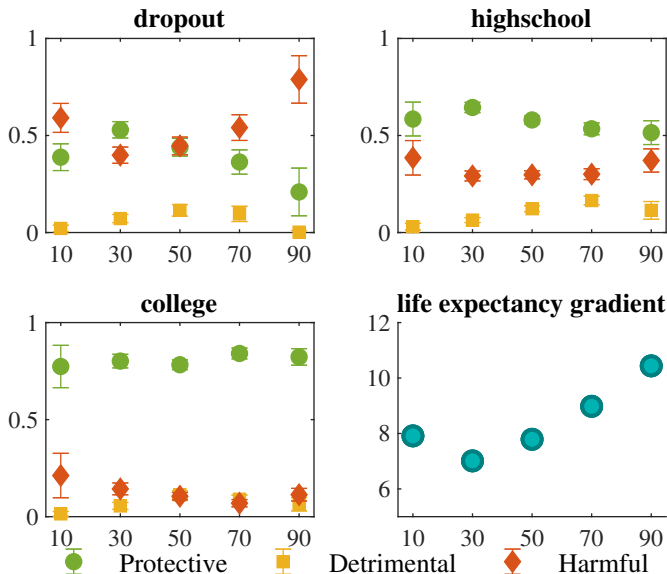


Figure: Probability of lifestyle at age 50 across cohorts. Males.

The Model

Three different stages

① Early life

- Choice of education and lifestyle

② Working Age

- Standard life-cycle incomplete-markets model of consumption with health and labor market risks

De Nardi, Pashchenko, Porapakkarm (2022)

- A model period is two years, individuals retire at age R

③ Retirement

- As above, but without labor market risks

Working Age (stage 2)

State variables

- Working agents are heterogeneous with respect to:
 - education (e)
 - lifestyle (y)
 - health status (h_t)
 - persistent shock to earnings (ξ_t)
 - persistent shock to medical expenses (ζ_t)
 - cash-on-hand (x_t)

Working Age (stage 2)

Worker's problem

- Worker's problem can be written as:

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

s.t.

$$k' = x - c$$

$$x' = \min\{(1+r)k' + w_t^{e,y}(\zeta', h', \epsilon') - m_t(\xi', h') - Tax, \underline{c}\}$$

$$Tax = T(w_t^{e,y}(\zeta', h', \epsilon')) + \tau_{MCR} w_t^{e,y}(\zeta', h', \epsilon') + \tau_{ss} \min\{w_t^{e,y}(\zeta', h', \epsilon'), w_{ss}\}$$

$$\text{Flow utility: } u(c, h) = (1 - \delta_{bh}) \frac{c^{1-\sigma}}{1-\sigma} + b$$

$$\text{Bequest motive: } v(a) = \theta \frac{(a + \underline{k})^{1-\sigma}}{1-\sigma}$$

Model Estimation (stage 2 & 3)

A two-step estimation strategy

- ① We set parameters related to demographics, taxes, social security benefits, and estimate the shock processes directly from the data.
- ② Method of Simulated Moments to estimate our remaining model parameters
 - We match p25, p50, p75 wealth moments across age, education and lifestyles.
 - We minimize the sum square of the difference between the targeted and simulated wealth moments.

Model Estimation (stage 2 & 3)

Internally estimated parameters

Parameter	Description	Value
β	discount factor	0.932
\underline{c}	consumption floor	3.93
δ_{bh}	disutility from bad health	0.278
\underline{k}	bequest motive: non-homoteticity	45.5
θ	bequest motive: marginal utility	0.107
b	value of life	1.33

Model Estimation (stage 2 & 3)

Model Fit

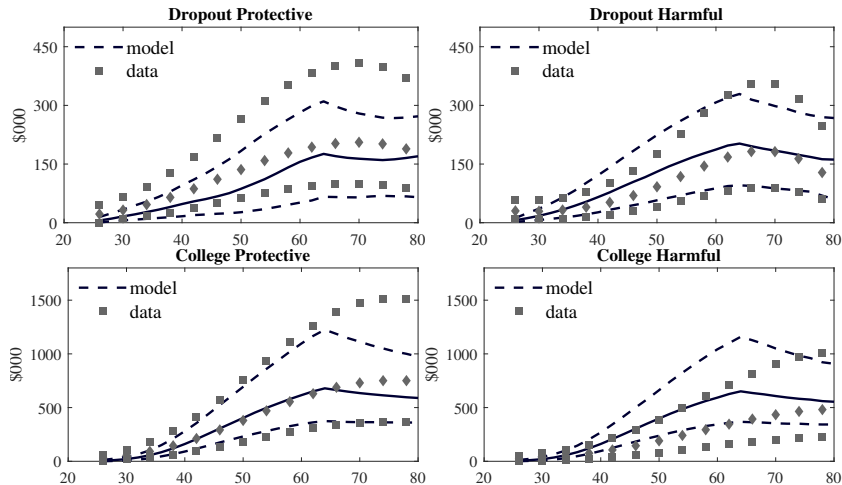


Figure: Wealth distribution: model vs data

Model Fit

Stage 2 & 3

- Good fit for the differences in wealth accumulation profiles across education categories
- Not so good fit for the differences in wealth accumulation across lifestyles
 - Differences in β s across lifestyles?

Early life (stage 1)

- Let $V_0^{e,y}$ be the value of starting working life with type (e, y)
(coming from Stage 2)
 - Let $c_{e,y}$ be the average cost of undertaking choice (e, y)
(we normalize $c_{\tilde{e},\tilde{y}} = 0$)
 - Let $\epsilon_{e,y}$ be the idiosyncratic cost of undertaking choice (e, y)
- Teenagers/parents make once-and-for-all choices of (e, y) :

$$\max_{e,y} \left\{ V_0^{e,y} - c_{e,y} - \sigma \epsilon_{e,y} \right\}$$

Early life (stage 1)

- With extreme value distributed shocks, the fraction of individuals taking each choice is,

$$q_{e,y} = \frac{\exp[(V_0^{e,y} - c_{e,y})/\sigma]}{\sum_{e_i,y_j} \exp[(V_0^{e_i,y_j} - c_{e_i,y_j})/\sigma]}$$

- One can write

$$\log q_{e,y} - \log q_{\tilde{e},\tilde{y}} = \frac{1}{\sigma} (V_0^{e,y} - V_0^{\tilde{e},\tilde{y}}) - \frac{1}{\sigma} c_{e,y}$$

- Use this equation to:
 - Estimate σ by OLS (covariance between differences in type frequencies)
 - Recover investment costs from regression residuals

Early life (stage 1)

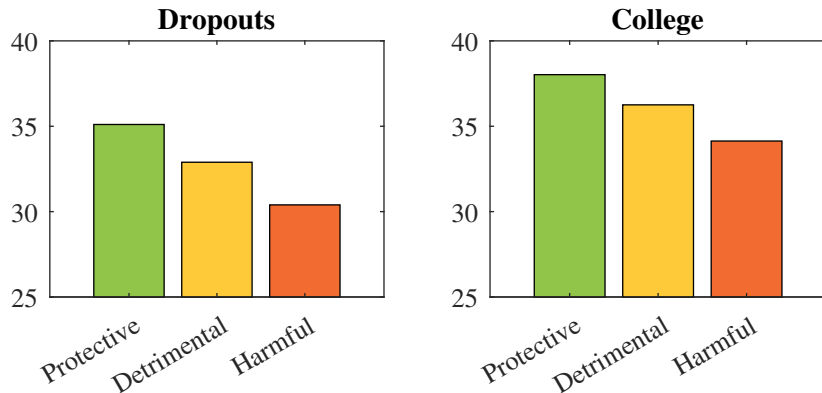
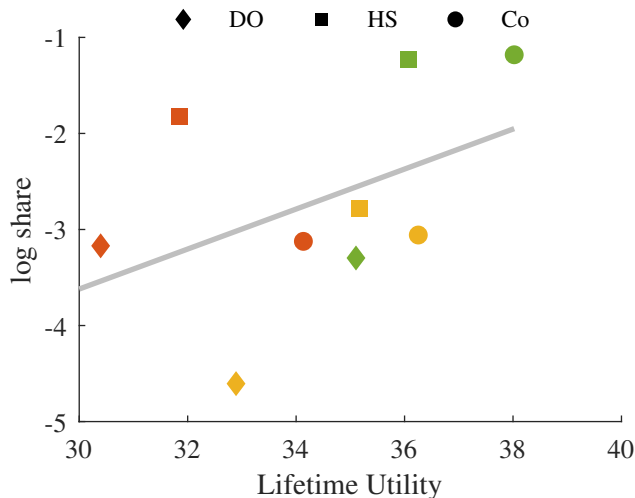


Figure: Lifetime utility at age 23: $V_0^{e,y}$

- Dropouts: Harmful: 47%; Protective: 42%
- College: Harmful: 10%; Protective: 77%

Early life (stage 1)



Early life (stage 1)

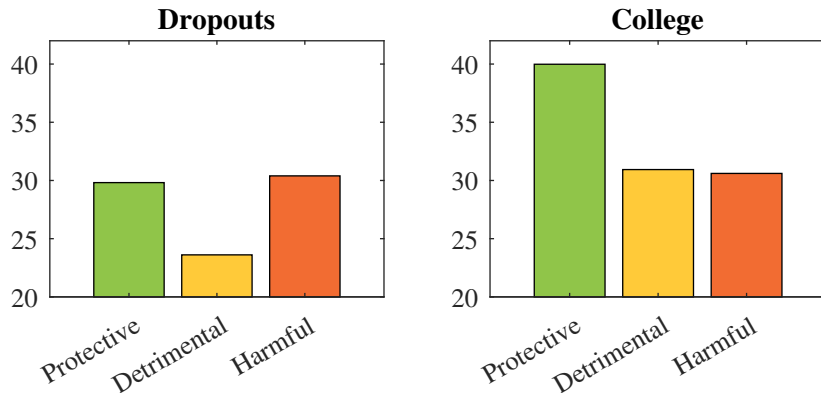


Figure: Lifetime utility at age 23: $V_0^{e,y}$ upper panel; $V_0^{e,y} - c_{e,y}$ lower panel

Results

- Differences in lifetime utility at age 23 are unable to explain why dropouts don't adopt healthier lifestyles as dropouts would benefit more than college graduates from doing so.
- In order to match the large share of harmful-dropout and protective-college types negative costs are needed.
 - For a dropout it is more costly to adopt a protective than a harmful lifestyle
 - For a college, the reverse is true.
- This suggests that there are complementarities in education and health behavior choices beyond the ones incorporated in the model that drive the large health inequalities observed in the data

Counterfactuals

- We ran two counterfactuals to quantify the effect of costs in the decision of types at young ages:
 - ① Choices taken based exclusively on $V^{e,y}$ at the initial age: $c_{e,y} = \bar{c} \forall e, y$
 - ② Revenue-neutral tax reform of labor earning taxes: the US becomes Denmark (top decile marginal tax rate 38% \rightarrow 55%)

	College (%)	Protective (%)	LE ₅₀ (yr)	LE ₅₀ gradient (yr)	\bar{V}	$var(V)$
Benchmark	40	63	26.3	7.9	35.6	5.5
No costs	45	47	26.0	5.0	35.4	4.1
<i>US \rightarrow DNK</i>	38	62	26.2	8.0	35.8	5.1

Counterfactuals

- Unobserved initial factors (parental investment, peers, genes) are key drivers of the observed health and welfare inequality.
 - Without costs, the gradient in LE would decrease in 3.5 years or 45%
 - Dropouts behave better: fraction of harmful 47.5% \rightarrow 22%
 - College behave worse: fraction of harmful 11% \rightarrow 24%
- Increasing the progressivity of the tax system decreases the variance in welfare inequality but deteriorates health inequalities.
 - The economic incentive of well behaving for the dropouts decreases.

Conclusions

- We developed a econometric model to summarize life styles exploiting health behavior and health outcomes.
- We solve and estimate an HA model with incomplete markets where education and lifestyle choices are made early in life.
- We find that economic and health shocks which happen after age 23 explain little of the observed differences in lifestyles across SES.
- Policies related to unobserved factors that happen early in life have a quantitatively larger impact at reducing health and welfare inequalities.