Healthy Habits and Inequality

Dante Amengual[☆] Jesús Bueren[◇] Josep Pijoan-Mas[□]

[☆]CEMFI

[◇]EUI

[□]CEMFI, CEPR

CLIC, January 2024

Introduction

- In recent decades, there has been an increase in in both income and health inequalities

- Two important facts:
  a) Strong association between economic and health inequality
     Kitagawa and Hauser (1973); Pijoan-Mas and Rios-Rull (2014); Chetty *et al* (2016)
  b) Growing educational gradients of health inequality
     Preston and Elo (1995); Meara et al. (2009); Montez *et al* (2011); Case and Deaton (2015)

  → Reasons not well-understood

- We aim to study to which extent differences in lifestyles across education groups and time can account for these facts
  - Differences in lifestyles are an important driver of health outcomes
    Li *et al* (2018); Zaninotto, Head, Steptoe (2020)
  - More educated individuals tend to adopt healthier habits
    Lantz et al. (1998); Martikainen et al. (2013)

Objectives

1. Understand the joint determination of education and lifesyle
   → Why is there an education gradient of lifestyles?

2. Quantify the effect of increased economic inequality on the increase in the educational gradient of health outcomes

What we do: Model

- We build an heterogeneous agents model featuring two different stages

  1. Early life heterogeneous individuals in their cost of studying and engaging in protective health behaviors (exercising, smoking, etc.) invest:
     - Education: affects their future income
     - Lifestyle: drives health transitions and survival

  2. Working/retirement age: individuals solve a standard life-cycle model with idiosyncratic labor income and health risks

     (conditional on given education and lifestyle)

What we do: Model

- Complementarities between education and lifestyle investments
  - An extra year of life is more valuable with higher consumption possibilities
  - The benefit in health transitions of investing in protective health behavior differs across education groups
- Complementarities of health and education investment are going to affect the selection of individuals in different education categories:
  - Individuals facing lower health behavior cost are more likely to choose higher education investments

What we do: Model

Numbers

- We take health dynamics conditional on education and lifestyle as given
- Calibrate the model to match:

  1. The joint distribution of education and lifestyles choices for different cohorts: born in the 30's, 50's and 70's.
  2. Savings decisions by education and lifestyles of individuals born in 1930's.

- Main results:

  → Education gradient of lifestyles. Both the income and health advantages of college education are equally important

  → Increase in LE gradient .
    - The model is able to explain 50% pf the increase in LE gradient between the cohorts born in 1930s and 1970s.
    - 80% of the declines in LE of high-school dropouts explained by deteriorating economic conditions
    - 20% of the declines in LE of high-school dropouts explained by selection

Literature on health and economic inequality

- Models with *exogenous health*
  Hosseini et al. (2021); De Nardi et al. (2023)

  → We model endogenous health

- Models with *endogenous monetary health investments*
  Fonseca et al. (2023); Hong et al.(2023)

  → We focus on health-related behaviour

- Models with *endogenous health behaviour investments*
  Cole et al. (2019); Mahler and Yum (2023); Margaris and Wallenius (2023)

  → We focus on once-and-for-all choices of lifestyles and education

  → Identification of lifestyles on health based on differences in health dynamics and mortality.

→ We study how long-run changes in economic inequality shape changes in health inequality

The Model

The Model

Two different stages

1. Early life
   - Choice of education and lifestyle
2. Life cycle
   a) Working age: standard life-cycle incomplete-markets model of consumption with health and labor market risks
   b) Retirement: as before, but without labor market risks

Stage 1: Early life

Set up

- Teenager/parents make once-and-for-all simultaneous choices of
  - education $\mathsf{e} \in \{\mathrm{HSD}, \mathrm{HSG}, \mathrm{CG}\}$
  - lifestyle $\mathsf{y} \in \{\mathrm{DET}, \mathrm{PRO}\}$
- They solve $\max\limits_{\mathsf{e},\mathsf{y}} \left\{ V_0^{\mathsf{eyc}} - \tau_{\mathsf{e}} - \tau_{\mathsf{y}} \right\}$
  - Value V_0^{eyc} of starting stage 2 with type $(\mathsf{e}, \mathsf{y}, \mathsf{c})$
  - Cost τ_{e} of education e:

$$\tau_{\mathrm{HSD}} = 0 \quad \Big| \quad \tau_{\mathrm{HSG}} = \epsilon_e \quad \Big| \quad \tau_{\mathrm{CG}} = \mu_{\mathrm{CG}} \epsilon_e, \; \epsilon_{\mathrm{PRO}} \sim N(\mu_e, \sigma_e)$$

  - Cost τ_{y} of lifestyle y:

$$\tau_{\mathrm{DET}} = 0 \quad \Big| \quad \tau_{\mathrm{PRO}} = \epsilon_{\mathrm{PRO}}, \; \epsilon_{\mathrm{PRO}} \sim LN(\mu_{\mathrm{PRO}}, \sigma_{\mathrm{PRO}})$$

Stage 2: Life cycle

State variables

- Working agents are heterogeneous with respect to:

1. Types
   - Education e $\in \{\text{HSD}, \text{HSG}, \text{CG}\}$
   - Lifestyle y $\in \{\text{DET}, \text{PRO}\}$
   - Cohort c $\in \{1930, 1970\}$
2. Exogeneous and deterministic state
   - Age $t \in \{25, 27, 29, ...\}$
3. Exogeneous and stochastic states
   - Health status $h_t \in \{h_g, h_b\}$
   - Employment status $l_t \in \{0, 1\}$
   - Shock to earnings $\zeta_t \in \mathbb{R}$
4. Endogenous state
   - Cash-on-hand $x_t \in [\underline{x}, \infty)$

Ingredients

- Health dynamics $s_t^{\text{ey}}(h)$ and $\Gamma_t^{\text{ey}}(h'|h)$ as estimated before
  - Survival and health transition are *not* cohort-specific
- Medical expenses $m_t^{\text{e}}(\xi, h)$
- Employment status $l_t^{\text{e}}(l_{-1}, \varepsilon, h)$ depends on
  - Education e and age t
  - Previous period employment l_{-1} and shock ε
  - Health h
- Labor earnings $w_t^{\text{ec}}(\zeta, \epsilon, h)$ depend on
  - Education e, cohort c, and age t
  - Persistent and transitory stochastic component ζ and ϵ
  - Health h
- Education costs $\bar{\tau}_t^{\text{ec}}$ paid over 8 first years of working life
- Progressive tax system $T()$ and minimum income floor $\underline{x}$

Worker's problem

- Worker's problem can be written as:

$$V_t^{\text{eyc}}(h,l,\zeta,x) = \max_{c,k'} \Big\{ u(c) + \beta s_t^{\text{ey}}(h) \sum_{h'} \Gamma_t^{\text{ey}}(h'|h)\, \mathbb{E}\left[V_{t+1}^{\text{eyc}}(h',l',\xi',\zeta',x')\right] + \beta^{T-t}\left(1 - s_t^{\text{ey}}(h)\right) v(k') \Big\}$$

s.t.

$$\begin{aligned}
k' &= x - c \\
\tilde{x}' &= (1+r)\,k' + T\left[l_{t+1}^{\text{e}}(l,\varepsilon',h') w_{t+1}^{\text{ec}}(\zeta',\epsilon',h')\right] - m_{t+1}^{\text{e}}(\xi',h') - \bar{\tau}_{t+1}^{\text{ec}} \\
x' &= \min\{\tilde{x}', \underline{x}\} \\
l' &= l_{t+1}^{\text{e}}(h',l,\varepsilon')
\end{aligned}$$

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

$$\text{Bequest motive:} \quad v(k) = \theta_1 \frac{(k+\theta_2)^{1-\sigma}}{1-\sigma}$$

Calibration

1. Define lifestyles and associated health dynamics

   – Exploit patterns in health behavior, health status, and survival.

2. Working retirement phase:

2.a/ External: parameters related to demographics, taxes, social security, and the stochastic processes for earnings and for health dynamics

   - This includes cohort-specific wages $w_t^{\text{ec}}(\zeta, h, \epsilon)$ and education costs $\bar{\tau}_t^{\text{ec}}$

2.b/ Internal: SMM to calibrate remaining parameters

   - Median wealth across age (by education and lifestyles) for the 1930s cohort
   - Value of statistical life

3. Early life: calibrate cost of education and lifestyles and variance/covariance of taste shocks

   – Joint distribution of education and lifestyles choices across cohorts

Calibration: Lifestyles

Lifestyles

Data

- The HRS and PSID provide an unbalanced panel of individuals $i = 1, ..., N$ followed for $t = 1, ..., T$ 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. Exercise

Lifestyles

Latent variables

- We want to incorporate heterogeneity in lifestyles into a structural model.
- In principle, we could have define a lifestyle for each possible combination of the observed health behavior variables : 2^6 lifestyles.
- Instead, we are going to 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 $y \in \{y_1, y_2, ...\}$.
- We interpret the latent factor (y) as the lifestyle
  - Allocate individuals to lifestyles
  - Measure the importance of lifestyles on health dynamics

Lifestyles

Mixture model

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

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

  - 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)

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,

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

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

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

Econometric Model

2. Health Dynamics

- We model the probability of reporting some health $h' \in \{Good, Bad, Dead\}$ next period as a multinomial probit model
  - There are latent variables ($h^*_{h,h'}$) that depend on gender (g), education (e), type (y), health (h), age (a), and an idiosyncratic shock ($\epsilon_{h'}$)

$$h^*_{h,h'} \quad = \quad f(a, s, e, y; \boldsymbol{\beta}_{h,h'}) + \epsilon_{h'}$$

with,

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

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:

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

Lifestyles

Results

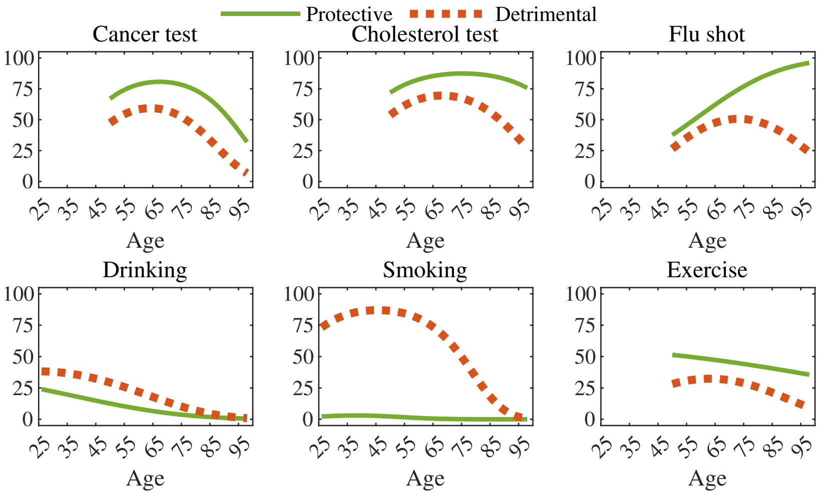


Figure 1: Probability of reporting health behaviors by lifestyle

Lifestyles

Results: Males

Table 1: LE at age 50 across education and lifestyles: males born in 1950s

	Dropouts		High-school		College		College-Dropout	
	Share	LE	Share	LE	Share	LE	ΔLE	ΔLE
All	100.0	25.5	100.0	28.0	100.0	32.3	6.8	4.3
Protective	54.3	29.0	69.1	30.3	88.3	33.4	4.4	
Detrimental	45.7	21.4	30.9	23.0	11.8	24.5	3.1	
Δ	8.6	7.6	38.2	7.2	76.5	8.9	1.4	

Lifestyles

Results: Females

Table 2: LE at age 50 across education and lifestyles: females born in 1950s

	Dropouts		High-school		College		College-Dropout	
	Share	LE	Share	LE	Share	LE	ΔLE	ΔLE
All	100.0	28.1	100.0	31.5	100.0	34.3	6.2	4.6
Protective	68.5	30.3	75.9	33.2	90.0	34.9	4.6	
Detrimental	31.5	23.2	24.2	26.1	10.0	28.2	5.0	
Δ	37.0	7.1	51.7	7.2	80.1	6.7	-0.4	

Lifestyles

Results

- More educated individuals tend to adopt healthier lifestyles.
  - The fraction of males with harmful lifestyle is 4 times bigger among high-school dropouts than among college graduates.
- If dropout males had the same lifestyles as college males, their life expectancy would increase by 2.5 extra years.
  - This corresponds to 37% of the observed difference in life-expectancy.

Lifestyles

Results

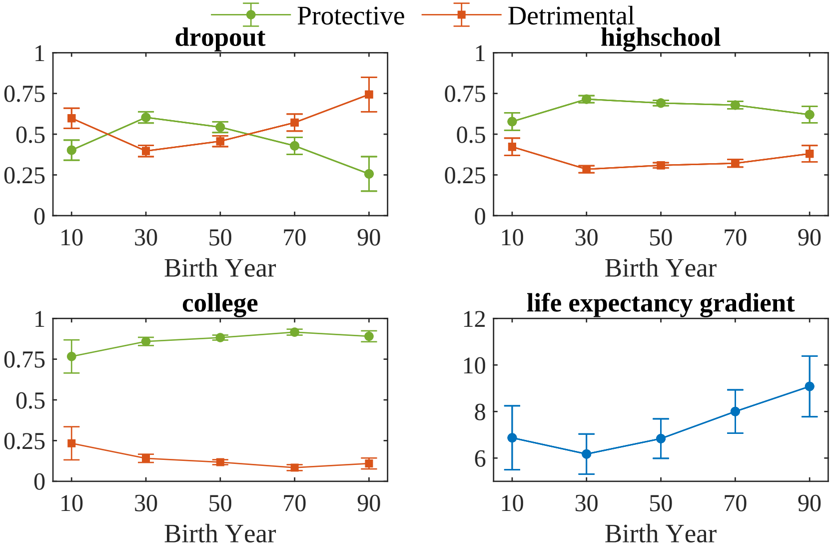


Figure 2: Probability of lifestyle at age 50 across cohorts. Males.

Lifestyles

Results

- Differences in lifestyles across education groups have increased.
  - The share of dropouts holding a detrimental lifestyle has increased from 40% for those born in 1930 to 57% for those born in 1970.
  - The share of college graduates holding a detrimental lifestyle has decreased from 14% for those born in 1930 to 8.4% for those born in 1990.
- This divergence in lifestyles across education groups has led to an increase in the life expectancy gradient.
  - From 6.2 years in 1950 to 8.0 years in 1970.
  - The importance of the differences in lifestyles to explain the life expectancy gradient has increased from 31% to 46% for those born in 1930 and 1970, respectively.

Calibration: Life cycle

Calibration

External parameters

- Labor income (PSID):
  - Extensive margin:

$$l_t^*(\mathsf{e}, l_{t-1}, \varepsilon_t, h_t) = f_t(\mathsf{e}, l_{t-1}, h_t) + \varepsilon_t, \quad \varepsilon_t \sim N(0,1)$$
$$l_t^{\mathsf{e}}(l_{t-1}, \varepsilon_t, h_t) = 1 \text{ if } l_t^*() > 0$$

  - Intensive margin:

$$w_t^{\mathsf{ec}}(\zeta_t, \epsilon_t, h_t) = \omega_t^{\mathsf{ec}}(h_t) + \zeta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2)$$
$$\zeta_t = \rho_\zeta \zeta_{t-1} + \nu_t, \quad \nu_t \sim N(0, \sigma_\nu^2)$$

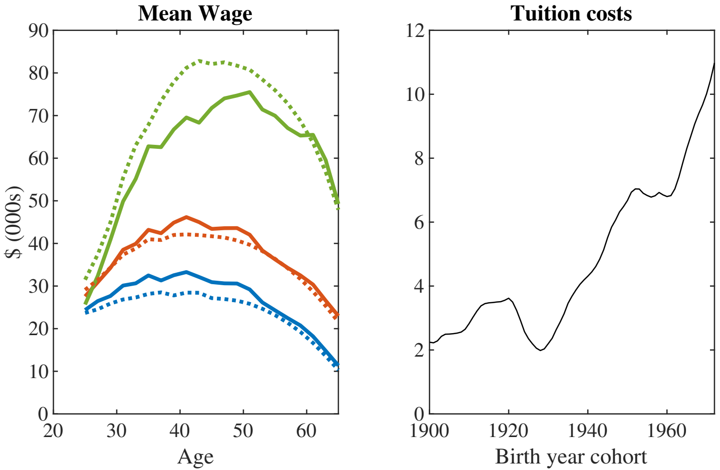
- Medical expenses (HRS):

$$m_t^{\mathsf{e}}(\xi_t, h_t) = \lambda_t^e(h_t) + \xi_t, \quad \xi_t \sim N(0, \sigma_\xi^2)$$

Estimation

Externally estimated parameters

Figure 3: College premium and tuition fees across cohorts



Source: PSID; Autor (2014); Donovan and and Herrington (2019)

Internal Parameters

- The model is able to replicate
  - Higher wealth accumulation for the more educated
  - Higher wealth accumulation for the protective (conditional on education)

- Parameter values

Parameter	Description	Value
$\underline{x}$	income floor	17.60
θ_1	bequest motive: marginal utility	3.90
θ_2	bequest motive: non-homoteticity	103.71
b	value of life	-0.63

Internal parameters

Model Fit

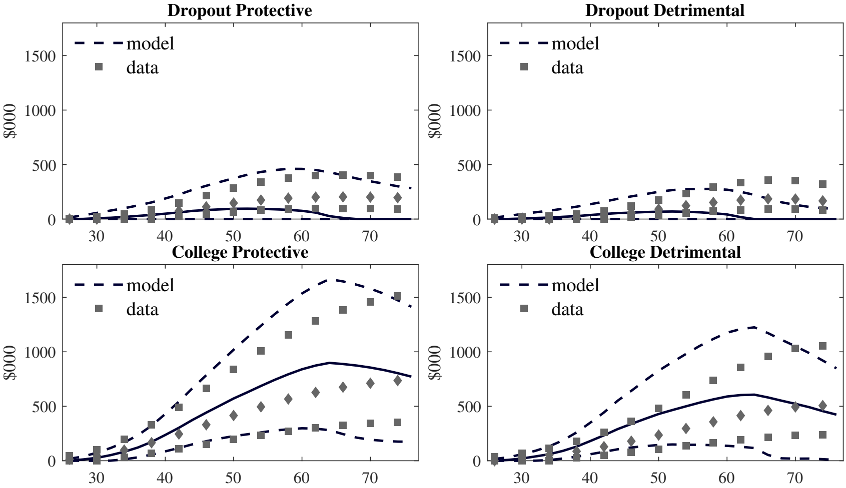


Figure 4: Wealth distribution: model vs data

Calibration: Early life

Calibration

- Match the joint distribution of e and y in each cohort c
  - → A total of 10 parameters and 10 statistics
- Identification
  - μ_{e}, μ_{PRO}, μ_{CG} drive the average share of e and y over time
  - σ^2_{e}, σ^2_{PRO} drive changes in e and y over cohorts c as V_0^{eyc} changes (due to changes in wages and tuition fees)

Calibration

Parameters

Parameter	Value	Parameter	Value
μ_e	6.58	σ_e	4.02
μ_{PRO}	8.52	σ_{PRO}	1.45
μ_{CG}	5.26		

Calibration

Fit: marginal distributions

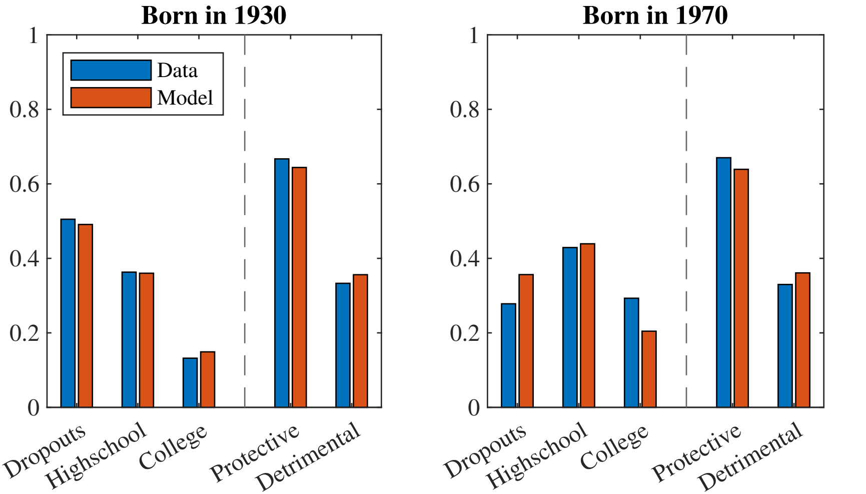


Figure 5: Marginal distributions: Education and Health Behavior

Calibration

Fit: conditional distributions

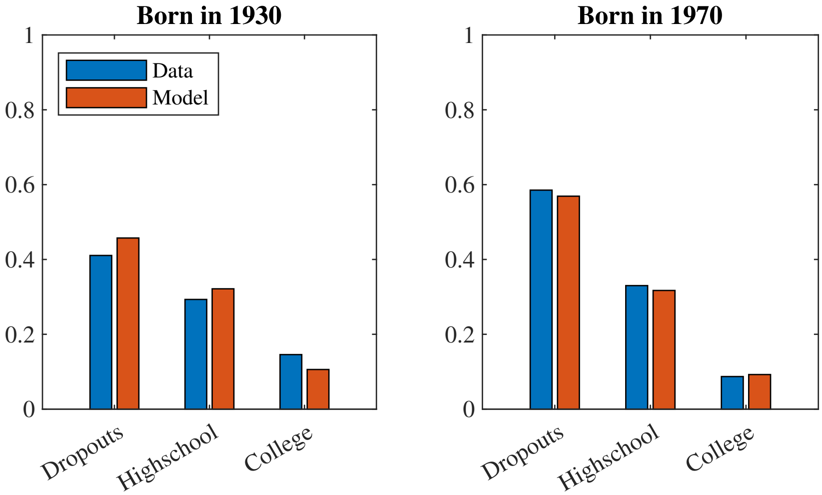


Figure 6: Conditional distribution of Detrimental Behavior by Education

Counterfactuals

Two questions

1. Why do individuals with higher education opt for more protective health behaviors?

2. To what extent the rise in earnings inequality has led to increased health inequalities?

Question 1: Education gradient of lifestyle and LE

Mechanisms

- Our model incorporates three mechanisms to explain why higher educated individuals are more likely to be protective:

1. Income gradient: $w_t^{\text{ec}}(\zeta, \epsilon, h)$

   Higher expected income for the more educated motivates healthier behavior as the value of life increases

2. Complementarity of health investments: $\Gamma_t^{\text{ey}}(h'|h)$

   Gains in life expectancy due to protective health behavior are more favorable for those with college education

3. Selection:

   Given the complementarities between health and education investment, early in life, individuals facing lower cost of protective behavior (τ_{PRO}) are more likely to choose higher education.

Results cohort 1930s

- Compute how differently lifestyles choices of the high-school dropout would be if they faced the same:

  1. income prospects
  2. health transitions
  3. cost distribution for health behavior

  as the college graduates

Results cohort 1930s

Benchmark

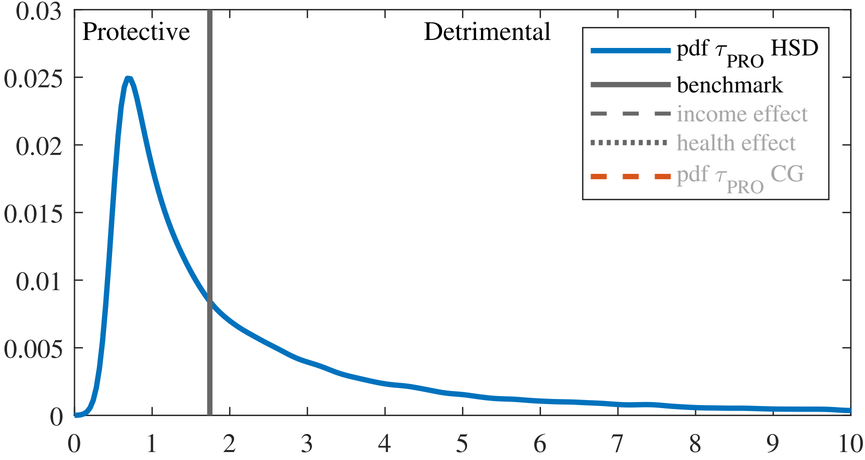


Figure 7: Cost of protective lifestyle

Results cohort 1930s

Income Effect

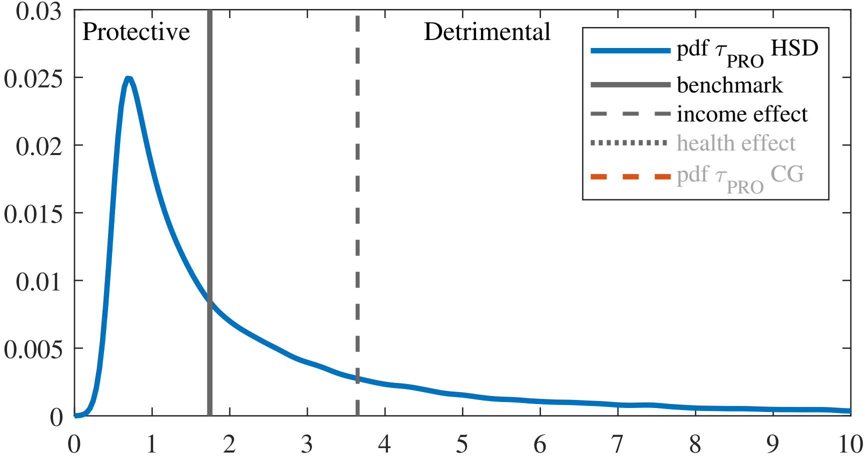


Figure 8: Cost of protective lifestyle

Results cohort 1930s

Health Effect

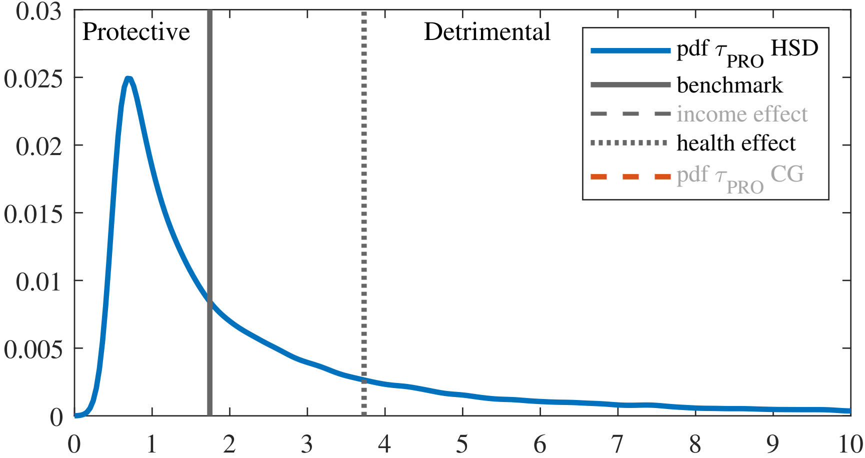


Figure 9: Cost of protective lifestyle

Results cohort 1930s

Selection Effect

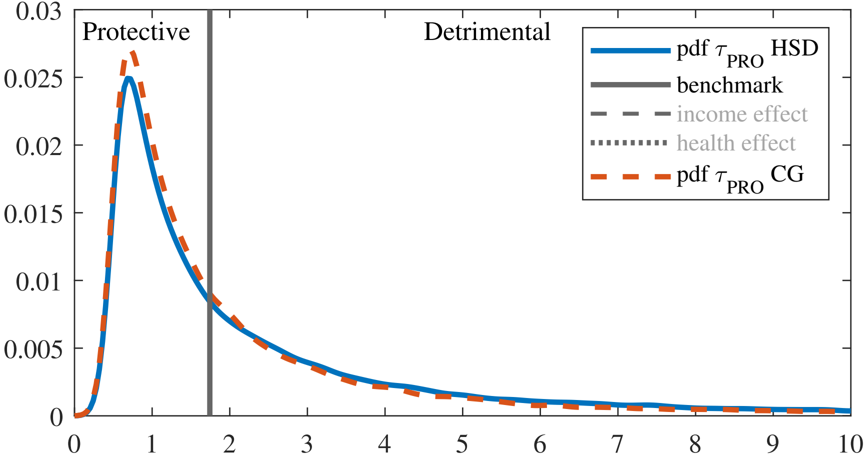


Figure 10: Cost of protective lifestyle

Results cohort 1930s

- The fraction of high-school dropout with detrimental health behavior would decrease from 45.7% to

1. Income effect: 22.9%
2. Health effect: 22.3%
3. Selection: 40.7%

Question 2: Changes over time

Mechanisms

- In the data, we observe an increase in the LE gradient between CG and HSD of 1.9 years.
- The model is able to explain 50% of the increase: 1.0 years
- Increase in the education wage premium
  a) Increases (decreases) the fraction of e = CG (e = HSD) individuals
  b) Effect on Pr [y = PRO | e] and LE(e)
     - Direct: return on health investments increases more for the more educated,

$$\Delta_{\mathsf{c}}\left(V_0^{\mathrm{CG},\mathrm{PRO},\mathsf{c}} - V_0^{\mathrm{CG},\mathrm{DET},\mathsf{c}}\right) > \Delta_{\mathsf{c}}\left(V_0^{\mathrm{HSD},\mathrm{PRO},\mathsf{c}} - V_0^{\mathrm{HSD},\mathrm{DET},\mathsf{c}}\right)$$

       → This increases the education gradient in Pr [y = PRO | e] and LE(e)
     - Selection: worse pool of individuals (in terms of τ_{PRO}) within HSD
       → *Ex ante* ambiguous effect on education gradient in Pr [y = PRO | e] and LE(e) as the pool of individuals within HSG will change

Results: summary

- Direct effect:
  - Accounts for 78% of the fall in LE among the HSD
  - Accounts for all the increase in LE among the CG
- Selection effect:
  - Accounts for 22% of the fall in LE among the HSD
  - Does not quantitatively drive changes in LE among the CG

Changes over time

High-School Dropouts

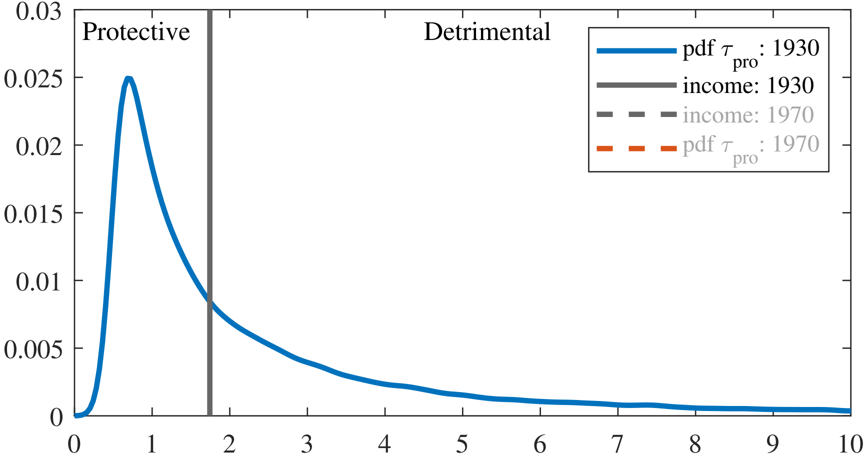


Figure 11: Cost of protective lifestyle

Changes over time

High-School Dropouts

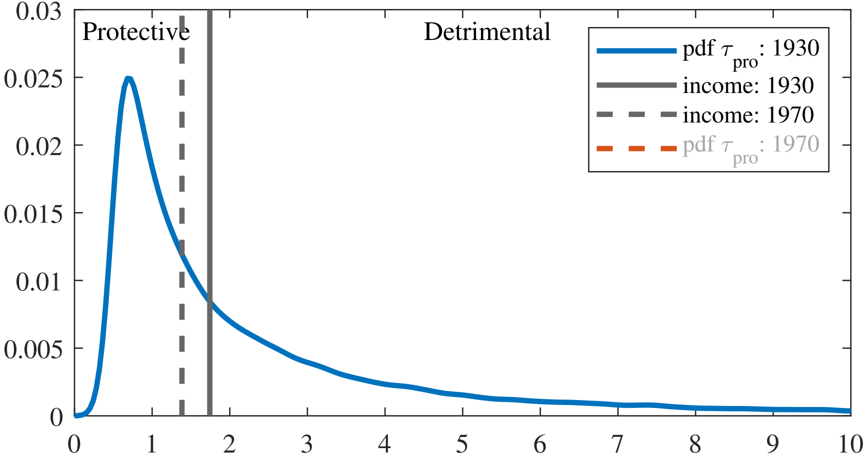


Figure 12: Cost of protective lifestyle

Changes over time

High-School Dropouts

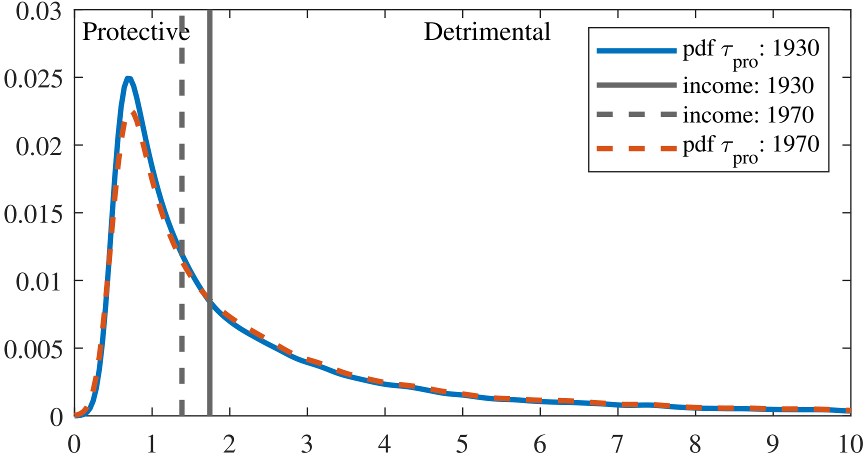


Figure 13: Cost of protective lifestyle

Changes over time

College Graduates

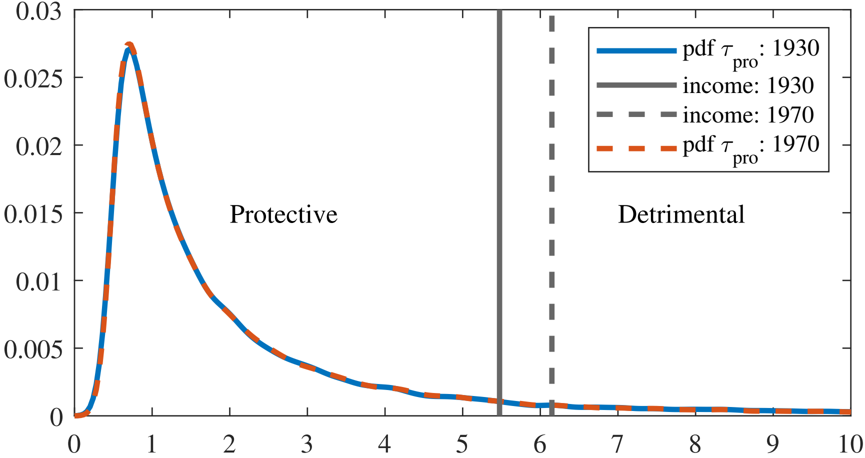


Figure 14: Cost of protective lifestyle

Conclusions

Conclusions