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 across education groups

- 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
   $\rightarrow$ 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 individuals heterogeneous in their willingness to study and preferences for engaging in detrimental health behaviors (smoking, exercising, 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
- The model allows us to identify to which extent an increase in the education premiums could lead to deterioration of health outcomes for lower-educated individuals because:
  1. Direct effect: a decline in their economic prospects.
  2. Selection: as a more negatively selected pool of individuals chooses to remain lower-educated.

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:

$$\tau_{\mathrm{HSD}} = 0 \quad \Big| \quad \tau_{\mathrm{HSG}} = \mu_{\mathrm{HSG}} + \epsilon_{\mathrm{HSG}} \quad \Big| \quad \tau_{\mathrm{CG}} = \mu_{\mathrm{CG}} + \epsilon_{\mathrm{CG}}$$

  - Cost τ_{y} of lifestyle y:

$$\tau_{\mathrm{DET}} = 0 \quad \Big| \quad \tau_{\mathrm{PRO}} = \mu_{\mathrm{PRO}} + \epsilon_{\mathrm{PRO}}$$

  - Where μ_{e}, μ_{y} are average costs of actions e and y
  - Where ϵ_{e}, ϵ_{y} are i.i.d normally distributed with variance σ_{e} and $\sigma_{\mathsf{y}}(e)$

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 $s_t^{\mathsf{ey}}(h)$ and $\Gamma_t^{\mathsf{ey}}(h'|h)$ as estimated before
  - Survival and health transition are *not* cohort-specific
- Medical expenses $m_t^{\mathsf{e}}(\xi, h)$
- Employment status $l_t^{\mathsf{e}}(l_{-1}, \varepsilon, h)$ depends on
  - Education e and age t
  - Previous period employment l_{-1} and shock ε
  - Health h
- Labor earnings $w_t^{\mathsf{ec}}(\zeta, \epsilon, h)$ depend on
  - Education e, cohort c, and age t
  - Persistent and transitory stochastic component ζ and ϵ
  - Health h
- Education costs $\bar{\tau}_t^{\mathsf{ec}}$ 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:

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

s.t.

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

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

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

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:

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

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:

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

- 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

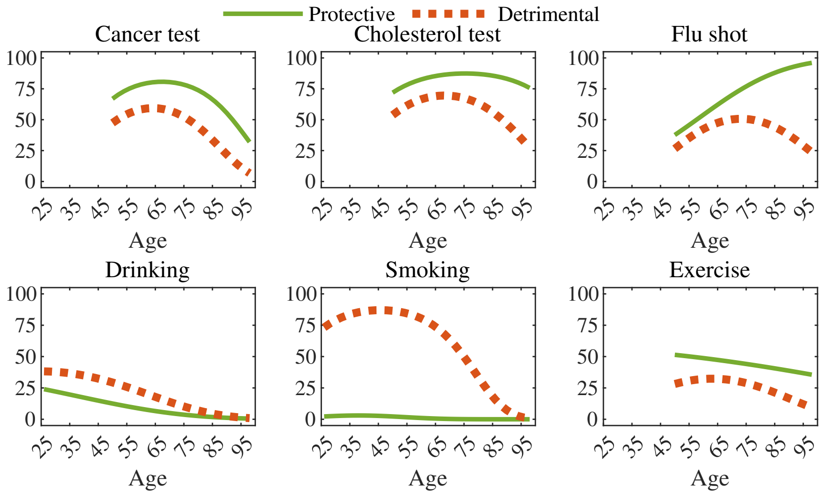


Figure 1: Probability of reporting health behaviors by lifestyle

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

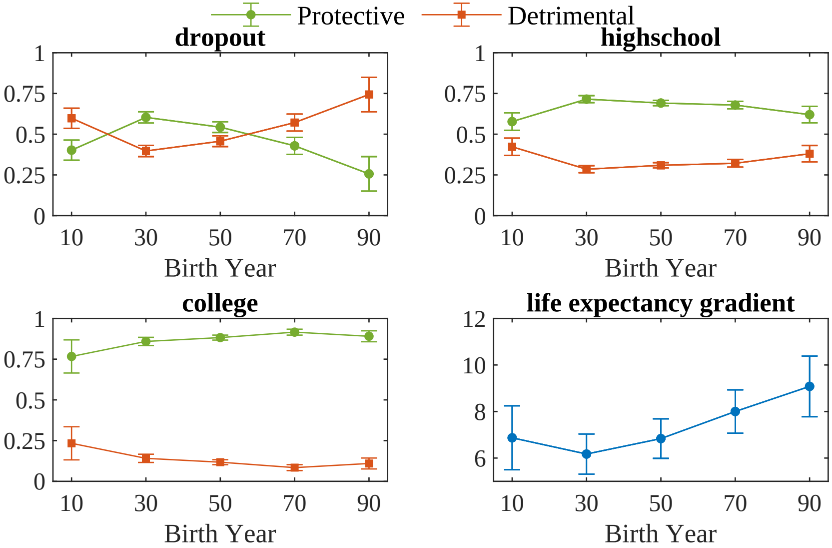


Figure 2: Probability of lifestyle at age 50 across cohorts. Males.

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:

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

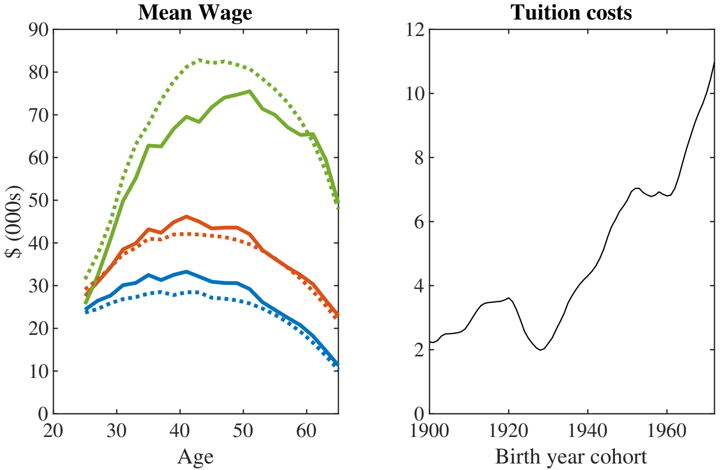
- Medical expenses (HRS):

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

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	16.05
θ_1	bequest motive: marginal utility	9.57
θ_2	bequest motive: non-homoteticity	130.36
b	value of life	0.66

Internal parameters

Model Fit

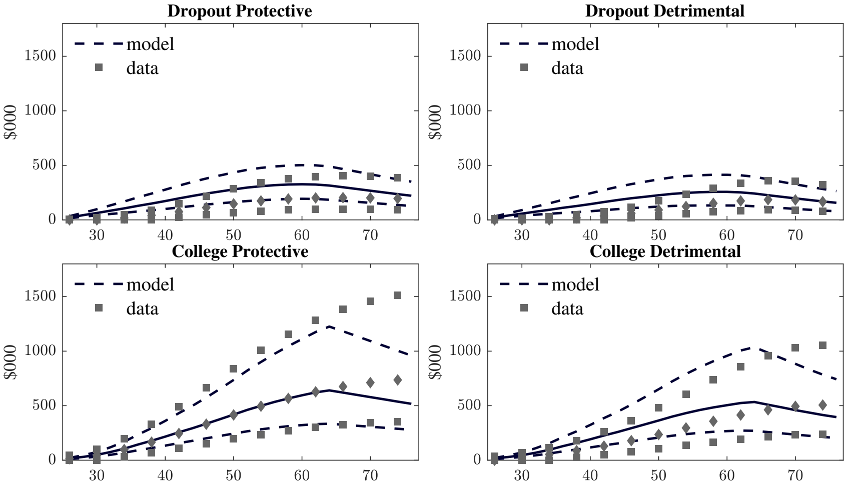


Figure 4: Wealth distribution: model vs data

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_{y} drive the <u>average</u> share of e and y over time
  - σ^2_{e}, σ^2_{y} drive <u>changes</u> 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_{PRO}	11.7	σ_{HSG}	1.9
μ_{HSG}	9.7	σ_{CG}	82.3
μ_{CG}	30.3		
$\sigma_{\mathrm{PRO}}(\mathrm{HSD})$	1.7		
$\sigma_{\mathrm{PRO}}(\mathrm{HSG})$	3.9		
$\sigma_{\mathrm{PRO}}(\mathrm{CG})$	5.2		

Calibration

Fit: marginal distributions

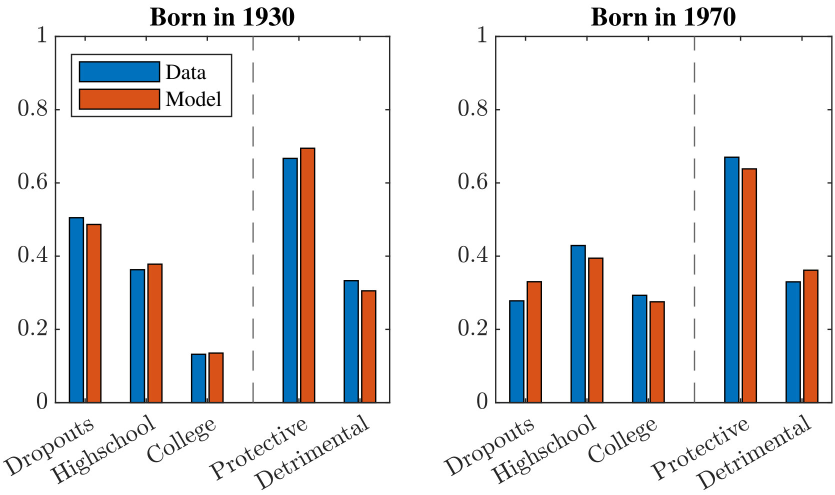


Figure 5: Marginal distributions: Education and Health Behavior

Calibration

Fit: conditional distributions

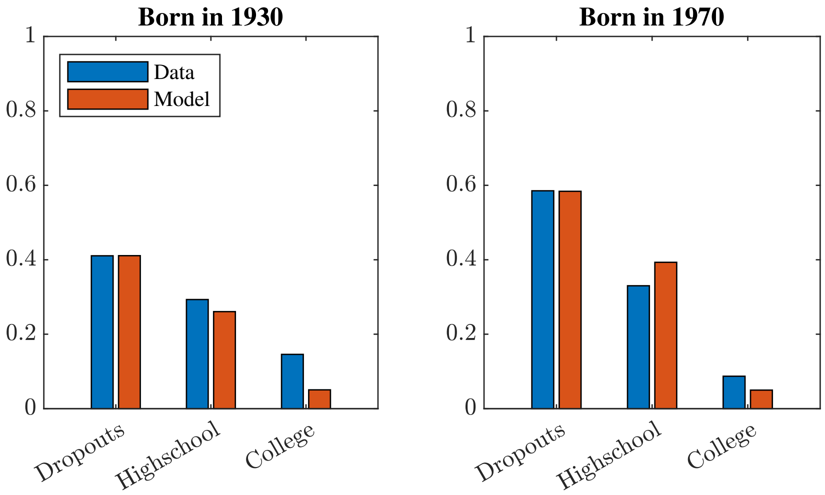


Figure 6: Conditional distribution of Detrimental Behavior by Education

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 two mechanisms through which the incentive to adopt a given health behavior varies across educational choices:

  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 health behavior are more favorable for those with college education

- Complementarities between education and health investment trigger selection: preferences for protective behavior vary over education levels.

- What is the effect of income on health behavior?

Results cohort 1930s

- Compute how differently lifestyles of the the higher educated would be if they faced the income prospects of the high-school dropouts.
  - The share of college graduate that would be of the detrimental type would triple from 5% to 15%.
- Compute how differently lifestyles of the the higher educated would be if they faced the health transitions of the high-school dropouts.
  - The share of college graduate that would be of the detrimental type would double from 5% to 12%.
- The combination of the two is still not going to close the gap because college educated are better selected in terms of their preferences towards health. (to be done)

Results cohort 1930s

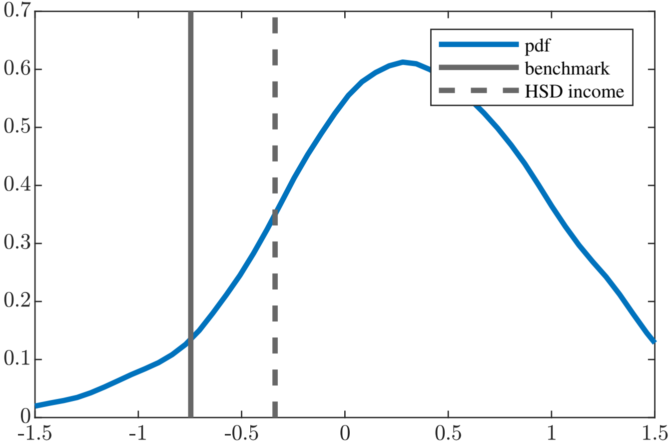


Figure 7: Taste shocks for protective lifestyle: College graduates

Question 2: Changes over time

Mechanisms

1. Increase in the education wage premium
   a) Increases (decreases) the fraction of $e = \text{CG}$ ($e = \text{HSD}$) individuals
   b) Effect on $\Pr[y = \text{PRO} \mid e]$ and $LE(e)$
      - Direct: return on health investments increases more for the more educated,

$$\Delta_c \left(V_0^{\text{CG},\text{PRO},c} - V_0^{\text{CG},\text{DET},c} \right) > \Delta_c \left(V_0^{\text{HSD},\text{PRO},c} - V_0^{\text{HSD},\text{DET},c} \right)$$

        → This increases the education gradient in $\Pr[y = \text{PRO} \mid e]$ and $LE(e)$
      - Selection: worse pool of individuals (in terms of ϵ_{PRO}) within CG and HSD
        → *Ex ante* ambiguous effect on education gradient in $\Pr[y = \text{PRO} \mid e]$ and $LE(e)$

2. Increase in college enrollment fees
   – Reversed patterns

Mechanism

- The fraction of detrimental types among dropouts increased from 41% to 59%.
- What part of it is driven back worsening of economic conditions (direct effect) versus selection?
- Couterfactual: how differently an individual who decided to be a high-school dropout in 1930's would have behaved in terms of their lifestyle if faced the income prospects of the 1970 cohort.

Cahnges over time

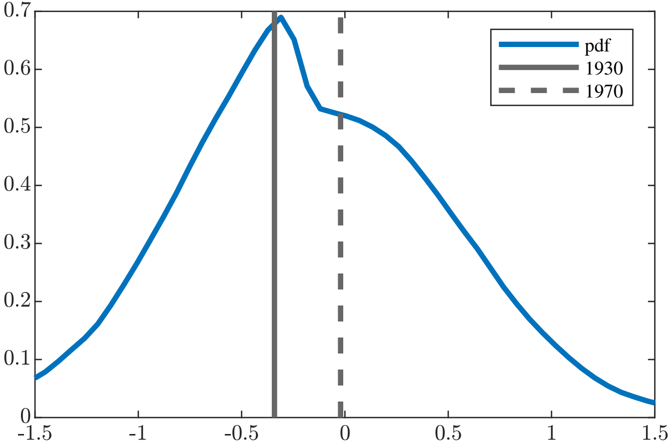


Figure 8: Taste shocks for protective lifestyle: High-school dropouts

Conclusions

Conclusions

- We develop a econometric model that allows us to incorporate lifestyles into an HA model to analyze complementarities in education and health investments
- Differences in lifestyles across education groups can account for around $1/3$ of the LE gradient
- Both income differences and larger health advantages for college individuals drive differences in lifestyles across education groups.
- Our model implies that increases in wage inequality are tightly linked to increases in health inequality.
- Worsening in the economic conditions for the high-school dropouts account for $3/4$ of their decline in lifestyle choices.