Healthy Habits and Inequality

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

[☆] CEMFI

[◇] EUI

[□] CEMFI, CEPR

European Economic Association, August 2024

Introduction

- In recent decades, there has been an increase in inequalities
  (*wealth, consumption, income, health outcomes*)

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

  → Reasons not well-understood

- We study to which extent differences in lifestyles across education groups can account for these facts
  - Lifestyles are an important driver of health outcomes
    McGinnis, Foege (1993); Li et al. (2018); Zaninotto et al. (2020)
  - More educated individuals tend to adopt healthier lifestyles
    Lantz et al. (1998); Cutler, Lleras-Muney (2010); Polvinen et al. (2013)

Objectives

1. Measure the impact of lifestyles on health dynamics and economic outcomes
2. Understand the joint determination of education and lifesyles
   → Why is there an education gradient of lifestyles?
3. Understand the increase in the education gradient of life expectancy
   → Quantify the role played by the increase in the education wage premium
      a) Income effect on lifestyle choices
      b) Selection: different composition of education groups

Literature on health and economic inequality

- Models with *exogenous health* dynamics

  De Nardi et al. (2010); Ameriks et al. (2020); Bueren (2023); Nakajima, Telyukova (2023)

  Capatina (2015); Braun et al (2019); Hosseini et al (2021); De Nardi et al. (2023)

  → We model endogenous health dynamics

- Models with *endogenous monetary health investments*

  Fonseca et al. (2023); Ozkan (2023); Hong, Pijoan-Mas, Ríos-Rull (2024)

  → Scarce causal evidence of effects of money on health

  → We focus on health-related behaviour

- Models with *endogenous health behaviour investments*

  Cole et al. (2019); Mahler, Yum (2023); Margaris, Wallenius (2023)

  → Joint decision of education and lifestyle: deal with selection

  → Effect of lifestyle identified by long-run health dynamics

  Hai, Heckman (2024)

  → We model different chanels

→ We study long-run changes of economic and health inequality

Health Dynamics and Health Behavior

The Data

- HRS and PSID
  - Unbalanced panels of individuals $i = 1, ..., N$ followed $t = 1, ..., T$ periods

- Demographic information: birth cohort (c), sex (s), 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 $z_{mt} \in \{0, 1\}$:
    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

Health behavior and health outcomes

- Across demographic groups: health behaviors associated w/ health outcomes
- But, at individual level:
  - Different health behaviors imperfectly correlated across individuals and over time [See data]
  - Hard to identify their long-run effect on health outcomes
  - Curse of dimensionality if you want to put them in a model
- Objective: identify latent types in lifestyle behavior based on:
  - Cross-sectional and panel variation of health behavior
  - Health dynamics

→ *Types are permanent, but health behaviors change with age and health status*

Latent types

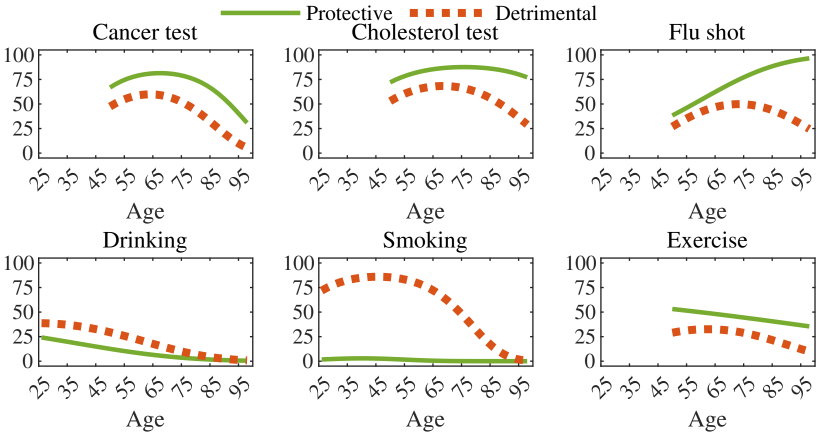
- We assume that observed health behavior is the (noisy) result of a latent time-invariant factor $y \in Y \equiv \{y_1, y_2, ...\}$
  - We interpret y as the individual lifestyle: propensity to engage in healthy behaviors
- We propose a novel econometric model to
  - Allocate individuals to lifestyles y
  - Measure the importance of lifestyles on health dynamics
- We *jointly estimate* health dynamics and lifestyles using a mixture model:

$$
\begin{aligned}
p(\boldsymbol{z}, \boldsymbol{h} | \mathsf{c}, \mathsf{s}, \mathsf{e}, a_0) &= \sum_{y \in Y} p(\boldsymbol{z}, \boldsymbol{h} | \mathsf{c}, \mathsf{s}, \mathsf{e}, a_0, y)\, p(y | \mathsf{c}, \mathsf{s}, \mathsf{e}, a_0) \\
&= \sum_{y \in Y} p(\boldsymbol{z} | \boldsymbol{h}, \mathsf{c}, \mathsf{s}, \mathsf{e}, a_0, y)\, p(\boldsymbol{h} | \mathsf{c}, \mathsf{s}, \mathsf{e}, a_0, y)\, p(y | \mathsf{c}, \mathsf{s}, \mathsf{e}, a_0) \\
&\simeq \sum_{y \in Y} \underbrace{p(\boldsymbol{z} | \boldsymbol{h}, a_0, y)}_{\text{health behavior}}\ \underbrace{p(\boldsymbol{h} | \mathsf{s}, \mathsf{e}, a_0, y)}_{\text{health dynamics}}\ \underbrace{p(y | \mathsf{c}, \mathsf{s}, \mathsf{e}, a_0)}_{\text{lifestyle distribution}}
\end{aligned}
$$

Results: Lifestyles and health behavior

1. Lifestyles "well" approximated by 2 types: protective and detrimental

Probability of reporting health behaviors by lifestyle



Results: Lifestyles and health dynamics

2. LE at age 50 is 8.6 years larger for protective than for detrimental

LE at age 50 across education and lifestyles: males born in 1970s

	All		HSD		HSG		CG		ΔLE (CG-HSD)		
	%	LE	%	LE	%	LE	%	LE	Data	(a)	(b)
All	100	**29.3**	100	24.9	100	28.0	100	32.8	7.9	4.7	3.2
PRO	74.4	**31.5**	44.3	28.5	69.0	30.1	93.6	33.4	4.9		
DET	25.6	**22.9**	55.7	22.0	31.0	23.3	6.4	23.8	1.9		
Δ	48.8	**8.6**	-11.4	6.6	37.9	6.8	87.2	9.6	3.0		

(a) Gradient explained by difference in health dynamics across education groups for given lifestyle, $\Delta_e p(\boldsymbol{h}|e, y)$
(b) Gradient explained by difference in lifestyles across education groups for given health dynamics, $\Delta_e p(y|e)$

Results: Lifestyles, education, and health dynamics

3. Effect of lifestyle on LE larger for the more educated (3 years)

LE at age 50 across education and lifestyles: males born in 1970s

	All		HSD		HSG		CG		ΔLE (CG-HSD)		
	%	LE	%	LE	%	LE	%	LE	Data	(a)	(b)
All	100	29.3	100	24.9	100	28.0	100	32.8	7.9	4.7	3.2
PRO	74.4	31.5	44.3	28.5	69.0	30.1	93.6	33.4	4.9		
DET	25.6	22.9	55.7	22.0	31.0	23.3	6.4	23.8	1.9		
Δ	48.8	**8.6**	-11.4	**6.6**	37.9	**6.8**	87.2	**9.6**	3.0		

(a) Gradient explained by difference in health dynamics across education groups for given lifestyle, $\Delta_{\text{e}} p(\boldsymbol{h}|\text{e}, y)$
(b) Gradient explained by difference in lifestyles across education groups for given health dynamics, $\Delta_{\text{e}} p(y|\text{e})$

Results: Lifestyles, education, and health dynamics

4. Lifestyles explain around 40% of the education gradient of LE

LE at age 50 across education and lifestyles: males born in 1970s

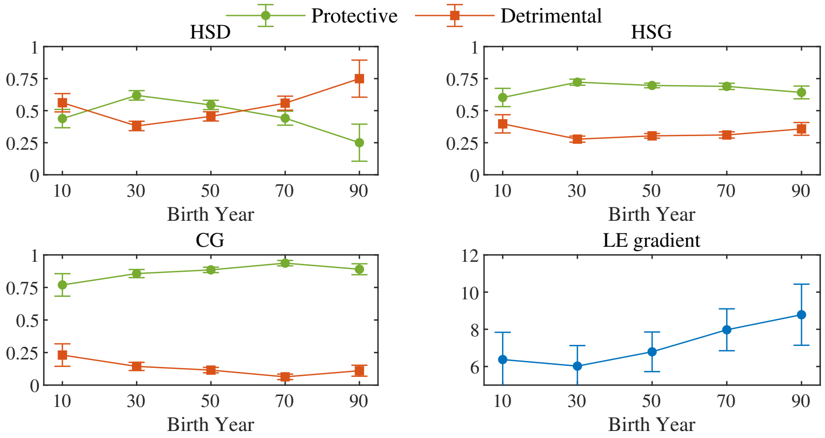
	All		HSD		HSG		CG		ΔLE (CG-HSD)		
	%	LE	%	LE	%	LE	%	LE	Data	(a)	(b)
All	100	29.3	100	24.9	100	28.0	100	32.8	**7.9**	**4.7**	**3.2**
PRO	74.4	31.5	**44.3**	28.5	**69.0**	30.1	**93.6**	33.4	4.9		
DET	25.6	22.9	**55.7**	22.0	**31.0**	23.3	**6.4**	23.8	1.9		
Δ	48.8	8.6	**-11.4**	6.6	**37.9**	6.8	**87.2**	9.6	3.0		

(a) Gradient explained by difference in health dynamics across education groups for given lifestyle, $\Delta_e p(\boldsymbol{h}|e, y)$
(b) Gradient explained by difference in lifestyles across education groups for given health dynamics, $\Delta_e p(y|e)$

Results: Changes across cohorts

5. Education gradient of lifestyles widens over time

Probability of lifestyle at age 50 across cohorts. Males



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 in cohort c make once-and-for-all simultaneous choices of
  - education $e \in \{HSD, HSG, CG\}$
  - lifestyle $y \in \{DET, PRO\}$
- They solve $\max_{e,y} \left\{ V_0^{eyc} - \tau_e - \tau_y \right\}$
  - Value V_0^{eyc} of starting stage 2 with type (e, y, c)
  - Cost τ_e of education e:

$$\tau_{HSD} = 0 \quad | \quad \tau_{HSG} \sim N(\mu_{HSG}, \sigma_{HSG}) \quad | \quad \tau_{CG} \sim N(\mu_{CG}, \sigma_{CG})$$

  - Cost τ_y of lifestyle y:

$$\tau_{DET} = 0 \quad | \quad \tau_{PRO} \sim N(\mu_{PRO}, \sigma_{PRO})$$

- The choices are independent if $V_0^{CG,PRO} - V_0^{CG,DET} = V_0^{HSD,PRO} - V_0^{HSD,DET}$
- There are complementarities between education and lifestyle if:

$$V_0^{CG,PRO} - V_0^{CG,DET} > V_0^{HSD,PRO} - V_0^{HSD,DET}$$

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 $x_t \in [\underline{x}, \infty)$

Worker's problem

- Worker's problem can be written as:

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

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

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

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

Calibration

Calibration

- Life-cycle:
  - External:
    - Parameters related to: demographics, taxes, social security
    - Cohort-specific wages $w_t^{ec}(\zeta, \epsilon, h)$ and labor force participation $l_t^{ec}(\varepsilon, h)$
    - Cohort-independent health dynamics $s_t^{ey}(h)$ and $\Gamma_t^{ey}(h'|h)$
    - Cohort-independent medical spending $m_t^{e}(\xi, h)$
  - Internal: SMM to calibrate remaining 4 parameters ($\underline{x}$, θ_1, θ_2, b)
    - Median wealth across age (by education and lifestyles) for the 1930s cohort
    - Value of statistical life
- Early life:
  - Estimate cost parameters by matching the marginal distributions of education and lifestyles for two different cohorts: 1930 and 1970.

Counterfactuals

Two questions

1. Why is there an education gradient of lifestyles?
2. What has been the effect of the rise in the education wage premium on the increase in education gradient of LE?

Question 1: Education gradient of lifestyle

Education gradient of lifestyle

Mechanisms

- Why higher educated individuals are more likely to be protective?

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

   Higher expected income for the more educated motivates healthier behavior because the value of life increases w/ consumption possibilities

$$V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} > V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$$

2. Complementarities in health dynamics: $s_t^{\text{ey}}(h)$ and $\Gamma_t^{\text{ey}}(h'|h)$

   Gains in life expectancy due to protective health behavior are larger for those with a college education

$$V_0^{\text{CG,PRO}} - V_0^{\text{CG,DET}} > V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}}$$

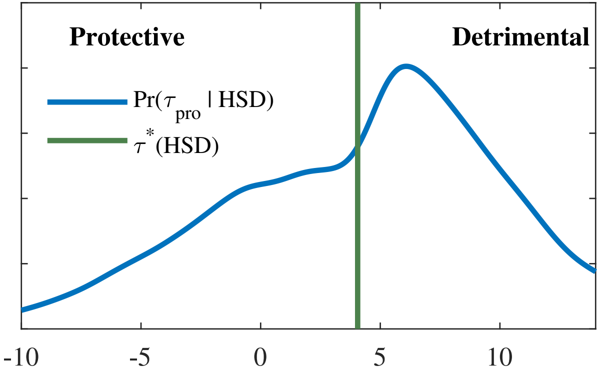
3. Selection:

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

Education gradient of lifestyle

Lifestyle Choice for HSD

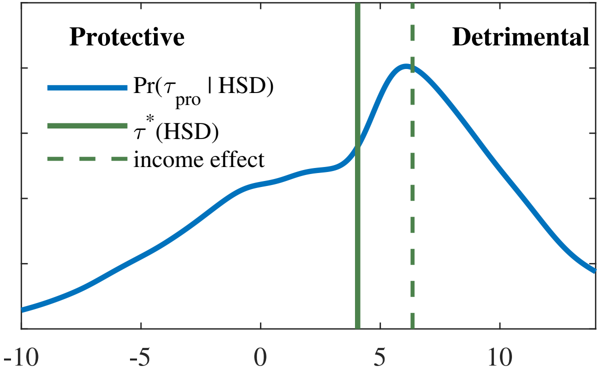
- HSD choose y = PRO iff $\tau_{\text{PRO}} < V_0^{\text{HSD,PRO}} - V_0^{\text{HSD,DET}} \equiv \tau^*(\text{HSD})$



Education gradient of lifestyle

1. Income Effect

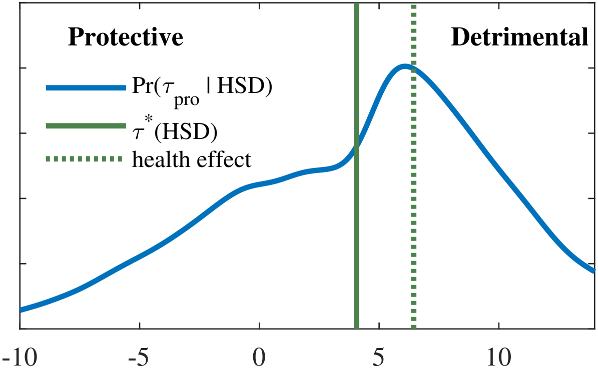
→ If HSD had same income as CG: 17pp more of PRO (out of 45pp gap)



Education gradient of lifestyle

2. Health Effect

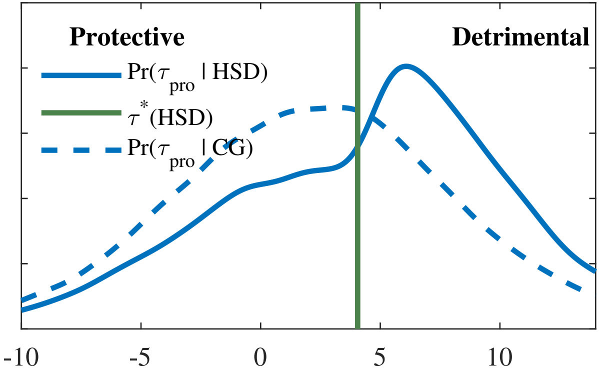
→ If HSD had same health gain of PRO as CG: 17pp more of PRO (out of 45pp gap)



Education gradient of lifestyle

3. Selection

→ If HSD had same distribution of τ_{PRO} as CG: 19pp more of PRO (out of 45pp gap)



Question 2: Changes over time

Changes over Time

Mechanisms

- Income effect: Increases in the education wage premium strengthen the complementarity between education and lifestyles,

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

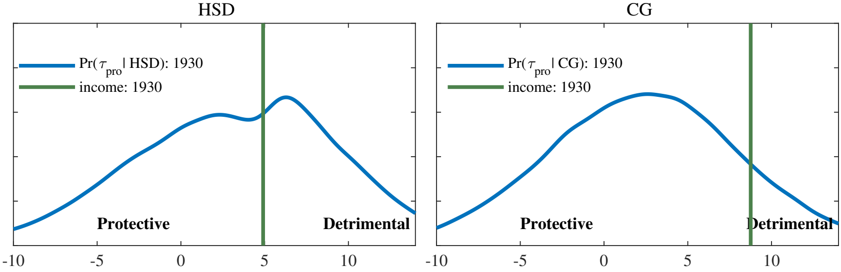
  → This increases the education gradient in $\Pr\left[y = \text{PRO} \,|\, e\right]$

- Selection effect: worse pool of individuals in terms τ_{PRO} for both HSD and CG.

  → Unclear effect on the education gradient

Changes over Time: 1930 vs 1970 cohorts

1930 cohort



Changes over Time: 1930 vs 1970 cohorts

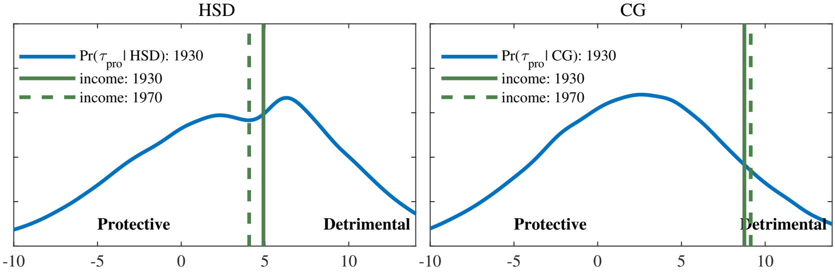
Income effect

Pr(τ_{PRO} |e) as in 1930, education premium widens

→ Pr(PRO|HSD) declines 5pp

→ Pr(PRO|CG) increases 1pp

→ LE gradient widens 5 months



Changes over Time: 1930 vs 1970 cohorts

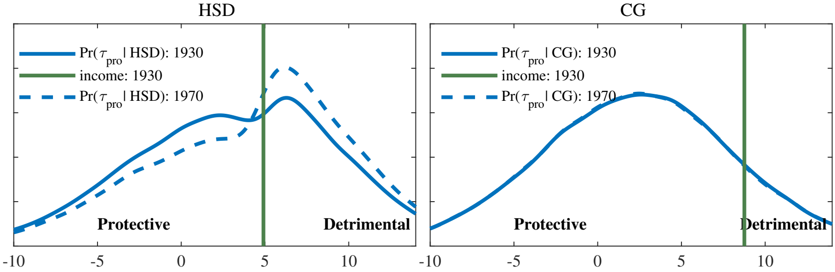
Selection effect

$\Pr(\tau_{\text{PRO}} | e)$ changes and education premium as in 1930

→ Pr(PRO|HSD) declines 9pp

→ Pr(PRO|CG) increases 0pp

→ LE gradient widens 8 months



Conclusions

Conclusions

- We develop an econometric model to identify latent types in lifestyles

⇒ Differences in lifestyles across education groups account for 40% of the LE gradient (3.1 out of 7.8 years)

- HA model w/ complementarities in edu and health investments

1. Education gradient of lifestyles
   - Income and Health advantage explains 15% (⇒ 1.1 years of LE gradient)
   - Selection explains 17% (⇒ 1.5 years)
2. 1.9 years increase in education gradient of LE between 1930s and 1970s cohorts
   - Changes in wages explains 55% of the increase that we see in the data
   - 42% explained by changes in wages
   - 60% explained by changes in selection

Econometric Model

1. Health Behavior

- 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:

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

Back to Mixture Model

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,

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

Back to Mixture Model

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:

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

Back to Mixture Model

Health Behaviour Data

Mean health behavior and 4-year auto-correlation

	Mean					AC	
	HSD	HSG	CG	50-60	70-80	50-60	70-80
Drinking	0.09	0.10	0.07	0.13	0.05	0.53	0.48
Smoking	0.23	0.16	0.07	0.21	0.08	0.81	0.78
Cancer test	0.66	0.76	0.85	0.71	0.77	0.42	0.41
Cholesterol	0.77	0.84	0.89	0.79	0.89	0.37	0.30
Flu shot	0.58	0.62	0.69	0.49	0.77	0.55	0.62
Exercise	0.26	0.39	0.55	0.42	0.38	0.40	0.39

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

Back to Data

Health Behaviour Data

Cross correlation health behaviors

	Drinking	Smoking	Cancer	Cholesterol	Flu shot	Exercise
Drinking	1.00	0.08	-0.02	-0.03	-0.03	0.02
Smoking	0.18	1.00	-0.10	-0.07	-0.06	-0.08
Cancer test	-0.04	-0.13	1.00	0.26	0.19	0.11
Cholesterol	-0.04	-0.11	0.39	1.00	0.24	0.07
Flu shot	-0.05	-0.05	0.23	0.24	1.00	0.02
Exercise	-0.01	-0.14	0.08	0.04	0.02	1.00

Notes: HRS. Upper diagonal: individuals aged 70 to 80. Lower diagonal: individuals aged 50 to 60.

Back to Data