Healthy Habits and Inequality

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

[☆]CEMFI

[◇]EUI

[□]CEMFI, CEPR

Very preliminary

Introduction

- Large body of research documenting economic inequality
  (*wealth, consumption, income, wages*)
- Growing interest in inequality in health outcomes
- Three important features of inequality in health outcomes
  - Inequality in health outcomes is large
  - It has an important socio-economic gradient
  - Much remains beyond the observed socio-economic gradient
- Economics research on health inequality typically focuses on its socio-economic gradient

  → We will look at inequality in health outcomes more broadly

Inequality in lifespans

Table 1: Life expectancy and lifespan SD at age 25, year 2000

	White		Black	
	Females	Males	Females	Males
Life Expectancy	55.8	51.1	51.6	45.3
Lifespan SD	13.0	13.9	15.3	15.6

Notes: Statistics reported from Sasson (Demography 2016).
Source: Death counts from U.S. Multiple Cause of Death and population from IPUMS

Inequality in lifespans

Table 1: Life expectancy and lifespan SD at age 25, year 2000

	White		Black	
	Females	Males	Females	Males
Life Expectancy	55.8	51.1	51.6	45.3
0-11	51.5	45.2	49.5	42.2
12	55.6	50.0	50.4	43.9
13-15	56.0	52.3	51.6	48.6
16+	58.7	54.9	54.5	50.4
Lifespan SD	13.0	13.9	15.3	15.6

Notes: Statistics reported from Sasson (Demography 2016).
Source: Death counts from U.S. Multiple Cause of Death and population from IPUMS

Inequality in lifespans

Table 1: Life expectancy and lifespan SD at age 25, year 2000

	White		Black	
	Females	Males	Females	Males
Life Expectancy	55.8	51.1	51.6	45.3
0-11	51.5	45.2	49.5	42.2
12	55.6	50.0	50.4	43.9
13-15	56.0	52.3	51.6	48.6
16+	58.7	54.9	54.5	50.4
Lifespan SD	13.0	13.9	15.3	15.6
0-11	16.1	16.5	17.6	17.2
12	13.3	14.6	14.9	15.2
13-15	12.4	13.4	14.1	15.0
16+	11.9	12.2	14.0	14.3

Notes: Statistics reported from Sasson (Demography 2016).
Source: Death counts from U.S. Multiple Cause of Death and population from IPUMS

Objectives

- Two main questions
  - *What are the reasons behind health inequality?*
  - *How does health inequality connect with economic inequality?*

- We conjecture that heterogeneity in healthy habits / lifestyle is an important driver of health inequality
  - Smoking, regular exercise, or healthy diet linked to better health outcomes
    Zaninotto, Head, Steptoe (2020); Li *et al* (2018)
  - Free access to medical care does not seem to matter much
    Aron-Dine, Einav, Finkelestein (JEP 2013); Finkelstein *et al* (NEJM 2016)

Objectives

- We seek to
  - Measure lifestyles and quantify its effects on health dynamics
  - Understand the determinants of different lifestyles
  - Make a connection between health and economic inequality
  - Quantify the economic costs of unhealthy lifestyles

What we do

1. Data

- We identify patterns in health behavior (preventive tests, substance abuse, obesity) driving health dynamics in both HRS and PSID
- We find that
  - Health behavior is well represented by three (unobserved) healthy habits types: protective, detrimental, and harmful
  - Large LE gradient: 12 years difference 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
    - Healthy habits explain 1/3 of the education gradient in life expectancy
  - Individuals holding harmful health behaviors tend to accumulate less wealth
    - And this is true within education groups

What we do

2. Model

- We build a life-cycle heterogeneous agents model with idiosyncratic labor market and health risks
  - Individuals differ in health habits and education as well as wealth and health
  - Health dynamics driven by previous estimates
  - Education and healthy habit choices taken together early in life
- Estimate the model
  - → Understand sources of heterogeneity across individuals
- Counterfactuals
  - → Understand role of healthy habits on economic inequality

Health Dynamics and Health Behavior

The Data

- The HRS provides an unbalanced panel of individuals $i = 1, \ldots, N$ followed for $t = 1, \ldots, T$ 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)
  - Health behavior (z_{mt}):
    1. Preventive cancer tests (mammography / prostate check)
    2. Cholesterol test
    3. Flu shot
    4. Heavy drinking (3+ 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 unobserved time-invariant latent 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 / healthy habit type
- We propose an econometric model exploiting both the cross-sectional and the time-series dimension of our data to
  - Allocate individuals to healthy habit types
  - Measure the importance of healthy habit types on health dynamics

Econometric Model

Overview

- We *jointly estimate* health dynamics and healthy habits types
  - 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 healthy habit type (y)
  - Health behaviours (z_{mt}) are iid, modelled through a probit
  - Health outcomes (h_t) are markovian, modelled through a nested probit
- Estimation by maximum likelihood
  - Gibbs sampling algorithm to reduce the complexity of the full likelihood into smaller simpler blocks.

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), and an idiosyncratic shock (ϵ_t)

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

  - Then,

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

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

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

Econometric Model

2. Health Dynamics

- We model the probability of reporting some health $h' \in \{Good, Bad, Dead\}$ next period as a 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 ($\epsilon_{h'}$)

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

with,

$$f(a, g, e, y; \boldsymbol{\beta}_{h'}) = \beta_{0,y,e,g,h,h'} + \beta_{1,y,e,g,h,h'} a + \beta_{2,y,e,g,h,h'} a^2$$

  – Then,

$$\text{Prob}\left(h' = Dead\right) = \text{Prob}\left(h^*_{h,h'=Dead} > h^*_{h,h'=Alive}\right)$$
$$\text{Prob}\left(h' = Good\right) = \text{Prob}\left(h^*_{h,h'=Alive} > h^*_{h,h'=Dead},\ h^*_{h,h'=Good} > h^*_{h,h'=Bad}\right)$$

Results: Healthy Habits

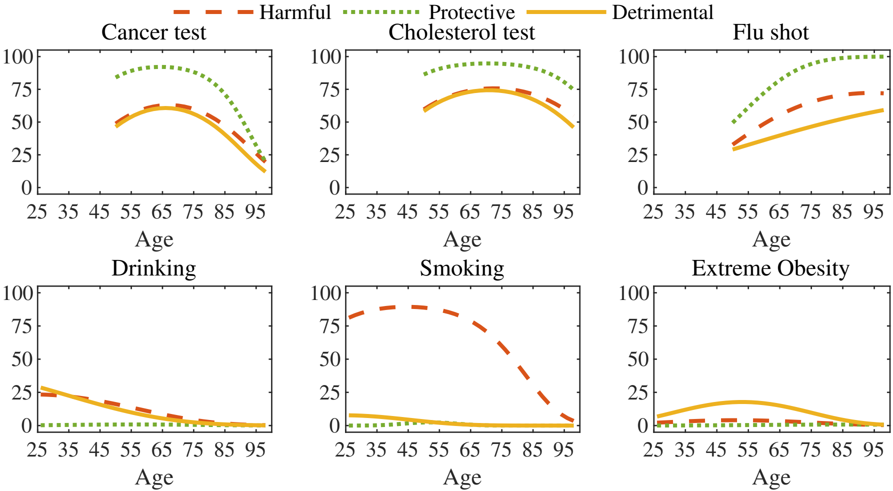

Figure 1: Probability of having a health habit by health behavior type as individuals age

Results: Health Dynamics

Males

Table 2: Expected duration of each health state at age 25 across behavior types

Health Behavior	Fraction	Life Expectancy	=	Good Health	+	Bad Health
			Men: Dropouts			
Protective	31.2	49.4		34.9		14.4
Detrimental	35.2	48.0		33.7		14.2
Harmful	33.6	38.9		25.7		13.2
Average	-	45.2		31.3		13.9
			Men: High-school			
Protective	40.3	54.8		51.2		3.6
Detrimental	33.8	52.1		48.4		3.8
Harmful	25.9	43.5		38.6		4.9
Average	-	50.3		46.2		4.1
			Men: College			
Protective	58.9	58.2		56.5		1.7
Detrimental	30.6	54.2		52.4		1.8
Harmful	10.5	45.2		41.6		3.5
Average	-	53.8		51.5		2.3

Results: Health Dynamics

Females

Table 3: Expected duration of each health state at age 25 across behavior types

Health Behavior	Fraction	Life Expectancy	=	Good Health	+	Bad Health
			Women: Dropouts			
Protective	31.9	53.2		38.8		14.4
Detrimental	42.1	50.4		28.2		22.2
Harmful	26.0	41.7		23.7		18.1
Average	-	47.7		29.7		18.0
			Women: High-school			
Protective	45.7	58.3		54.5		3.8
Detrimental	33.1	54.6		48.9		5.8
Harmful	21.2	47.0		39.6		7.4
Average	-	53.8		48.4		5.4
			Women: College			
Protective	59.1	61.4		59.8		1.7
Detrimental	31.1	57.0		54.2		2.8
Harmful	9.8	49.2		43.8		5.4
Average	-	57.2		54.2		2.9

Results: Health Dynamics

- More educated individuals tend to adopt healthier habits.
  - The probability that a college male has a harmful health behavior is 3 times smaller that a dropout.
- If dropout males had the same proportion of health behavior types than college males, their life expectancy would increase by 2.6 extra years.
  - This corresponds to 31% of the observed difference in life-expectancy at age 25 between college graduates males and high-school dropouts males.
- If dropout females had the same proportion of health behavior types than college females, their life expectancy would increase by 3.6 extra years.
  - This corresponds to 38% of the observed difference in life-expectancy at age 25 between college graduates females and high-school dropouts females.
- To which extent can the observed correlation be explained by differences in preferences for education and/or health behaviors?

Types and economic outcomes

Asset Accumulation (males)

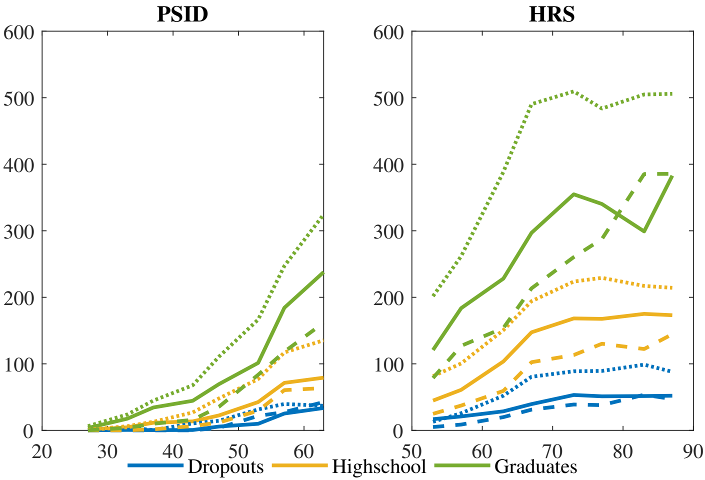

Figure 2: Median Assets across Health Behavior types in Thousand of Dollars (dotted line: protective; solid line: detrimental; dash line: harmful)

Results: Health Dynamics

- Conditional on education, individuals with healthier habits tend to accumulate more wealth.
- Differences in wealth holdings across individuals of different health behavior types but same education category can be driven by:
  - Higher survival rates.
  - Less frequent unhealthy states: higher wages & lower medical expenses.
  - Unobserved heterogeneity shaping both saving behavior and healthy habits

The Model

The Model

Timing

- We follow closely De Nardi et al. (WP)
- A model period is one year.
- Each individuals lives at most T periods.
- During the first $R-1$ periods of life an individual chooses whether to work or not.
- At age R all individuals retire.

The Model

Health

- At age t, agent's i health $h_{i,t}$ can be either good ($h_{i,t} = 1$) or bad ($h_{i,t} = 0$)
- Individual's health behavior type y_i can be protective ($y_i = 1$) or detrimental ($y_i = 1$)
- Health evolves according to the process defined in the empirical section.

Working Age

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

Working Age

- Worker's problem can be written as:

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

subject to

$$\begin{array}{rcl} k' & = & \mathbb{1}_{x > \underline{c}}(x - c) \\ x' & = & (1+r)\,k' + w_t^e(\zeta',h') - m_t(\xi',h') + T(\underline{\mathsf{c}}) \\ T(\underline{\mathsf{c}}) & = & \max\{0, \underline{\mathsf{c}} - (1+r)\,k' - l w_t^e(\zeta',h') + m_t(\xi',h')\} \\ \xi' & = & \rho_\xi \xi + \epsilon_\xi, \quad \epsilon_\xi \sim N(0,\sigma_\xi) \\ \zeta' & = & \rho_\zeta \zeta + \epsilon_\zeta, \quad \epsilon_\zeta \sim N(0,\sigma_\zeta) \end{array}$$

Retirement

- Retired agents are heterogeneous with respect to:
  - discount factor (β)
  - education (e)
  - health behavior (y)
  - pension (ss)
  - wealth net of medical expenses (x_t)
  - health status (h_t)
  - persistent shock to medical expenses (ξ_t)
- Pension is determined from last persistent shock, education, and health behavior type.

Retirement

- Retiree's problem can be written as:

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

subject to

$$\begin{aligned} k' &= \mathbb{1}_{x>\underline{c}}(x-c) \\ x' &= (1+r)\,k' + ss - m_t(\xi',h') + T(\underline{c}) \\ T(\underline{c}) &= \max\{0, \underline{c} - (1+r)\,k' - ss + m_t(\xi',h')\} \\ \xi' &= \rho_\xi \xi + \epsilon_\xi, \quad \epsilon_\xi \sim N(0,\sigma_\xi) \end{aligned}$$

Initial age

Formulation 1

- When young, individuals make once-and-for-all choices of:
  - Education $e \in \{e_1, e_2, e_3\}$
  - Lifestyle $y \in \{y_g, y_b\}$

- Individuals differ in
  - Discount factor $\beta \in \{\beta_i, \beta_2, ..., \beta_N\}$
  - Utility cost of studying $\theta \in \mathbb{R}_{++}$
  - Utility cost of good lifestyle $\eta \in \mathbb{R}_{++}$
  - They are potentially correlated:

$$(\log\theta, \log\eta) \sim N\left[\mu(\beta), \Sigma(\beta)\right]$$

  $\rightarrow$ This gives 5 free parameters per β type

- Higher education and better lifestyle provide
  - Higher lifetime utility $V^{\beta,e,y}()$
  - Higher flow utility cost $C^{\theta,\eta}(e,y)$

Initial age

Formulation 1

- Individuals chose the option with highest payoff

$$max_{e,y}\Big\{ - C^{\theta,\eta}\left(e,y\right) + \beta\left[V_0^{\beta,e,y}\left(x_0,h_0,\zeta_0,\xi_0\right)\right]\Big\}$$

- With the cost function,

$$C^{\theta,\eta}\left(e,y\right) = \mathbb{I}_{e=e_2}\theta + \mathbb{I}_{e=e_3}\bar{\theta}\left(\beta\right)\theta + \mathbb{I}_{y=y_g}\eta$$

- With the extra parameter $\bar{\theta}\left(\beta\right)$ we have 6 parameters (per β type)
- And we have 5 statistics from the data (per β type): the frequency of people in each of the 6 states, which needs to add up to one
- Something should be normalized, like the mean of θ or η?
- Then, the model-implied probability of each state can be obtained through the properties of the normal distribution and a NLES can be used to recover the parameters

Initial age

Formulation 2

- When young, individuals make once-and-for-all choices of:
  - Education $e \in \{e_1, e_2, e_3\}$
  - Lifestyle $y \in \{y_g, y_b\}$

- Individuals differ in
  - Discount factor $\beta \in \{\beta_i, \beta_2, ..., \beta_N\}$

- In addition, every individual draws a specific utility cost $\epsilon_{e,y}$ of pursuing each option
  - They are potentially correlated: $\epsilon_{e,y} \sim N\left[\mu\left(\beta\right), \Sigma\left(\beta\right)\right]$

  → This gives 6+21 free parameters per β type

- Higher education and better lifestyle provide
  - Higher lifetime utility $V^{\beta,e,y}\left(\right)$
  - Higher flow utility cost $\epsilon_{e,y}$

Initial age

Formulation 2

- Individuals chose the option with highest payoff

$$max_{e,y}\Big\{-\epsilon_{e,y}+\beta\Big[V_0^{\beta,e,y}\left(x_0,h_0,\zeta_0,\xi_0\right)\Big]\Big\}$$

- Again, the model-implied probability of each state can be obtained through the properties of the normal distribution
- However, we have 6+21 free parameters for the same 5 statistics. It seems that all the variance-covariance matrix is redundant and we identify only 5 means (with a normalization of the sixth one)
- How this formulation captures "correlation" between choices? Well, if $\mu_{e_1,y_g}-\mu_{e_1,y_b}$ is larger than $\mu_{e_3,y_g}-\mu_{e_3,y_b}$ it means that good lifestyle is more costly for e_1-educated than for e_3-educated individuals

Initial age

Formulation 3

- When young, individuals make once-and-for-all choices of:
  - Education $e \in \{e_1, e_2, e_3\}$
  - Lifestyle $y \in \{y_g, y_b\}$
- Individuals differ in
  - Discount factor $\beta \in \{\beta_i, \beta_2, ..., \beta_N\}$
- In addition, every individual draws a specific utility value $\epsilon_{e,y}$ of pursuing each option
  - These shocks are Type-I Extreme Value
- Higher education and better lifestyle provide
  - Higher lifetime utility $V^{\beta,e,y}()$
  - Higher flow utility cost $C(e,y)$
  - Higher flow utility $\epsilon_{e,y}$

Initial age

Formulation 3

- Individuals chose the option with highest payoff

$$max_{e,y}\Big\{\underbrace{-C\,(e,y) + \beta\Big[V_0^{\beta,e,y}\,(x_0,h_0,\zeta_0,\xi_0)\Big]}_{W^{\beta}(e,y)} + \epsilon_{ey}\Big\}$$

- With the cost function,

$$C\,(e,y) = c_{e,y}$$

This gives 6 parameters

- Then, the fraction of individuals of type β taking each choice is,

$$q\,(e_i,y_j|\beta) = \frac{exp\big[W^{\beta}\,(e_i,y_j)\big]}{\sum_{e,y} exp\,[W^{\beta}\,(e,y)]}$$

- These are 5 statistics to identify 5 parameters c_{ey} (with one nromalized to zero)
- How does the correalation between choices appear? As in formulation 2: if $c_{e_1,y_g} - c_{e_1,y_b}$ is larger than $c_{e_3,y_g} - c_{e_3,y_b}$ it means that good lifestyle is more costly for e_1-educated than for e_3-educated individuals

Comments 1

- It seems that the 3 formulations are equivalent, but I am not 100% sure
- Formulation 3 is very convenient analytically, the c_{ey} can be recovered w/o non-linear equation solver

Comments 2

- Heterogeneity in β alone will generate perfect correlation between of β, e, y:
  - → more patient households find it optimal to invest more in their future

- How do we create mixing of lifestyle, education, and wealth?
  1) Heterogeneity in either θ or η separates β from e and y, but generates perfect correlation between e and y:
     - → investments in education and lifestyle are strong complements (higher lifetime income means investing in health is more profitable and longer life expectancy means that investing in education is more profitable)
  2) Heterogeneity in both θ or η may help mix e and y

- If we allow for a flexible initial distribution of β, θ, η (with a free variance-covariance matrix), we may not need the extreme-value shocks
  - – How tractable is a joint log-normal distribution?