The Impact of Measurement Error in Health in Health-Related Counterfactuals

SMU Brown Bag

Martin Garcia–Vazquez (garciavazquez@wustl.edu)
Luis Pérez (**luisperez@smu.edu**)
Yuta Maeda (maeda025@umn.edu)

March 26, 2025

This Paper

- **Motivation.** Increasing body of research recognizes the importance of health in shaping economic decisions and outcomes
  - Most studies assume that health is perfectly measured
  - Yet evidence suggests that "health" is complex and hard to measure

This Paper

- **Motivation.** Increasing body of research recognizes the importance of health in shaping economic decisions and outcomes
  - Most studies assume that health is perfectly measured
  - Yet evidence suggests that “health” is complex and hard to measure
- **Research question:**

  How important is imperfect observability of health to evaluate costs of bad health?

This Paper

- **Motivation.** Increasing body of research recognizes the importance of health in shaping economic decisions and outcomes
  - Most studies assume that health is perfectly measured
  - Yet evidence suggests that “health” is complex and hard to measure
- **Research question**:

  How important is imperfect observability of health to evaluate costs of bad health?

- **Why do we care?**
  1. Better understanding of how health shapes decisions and outcomes
  2. Gives a sense of how biased previous studies may be
  3. Informative for future research

This Paper

What We Do.

- **Estimate canonical structural life-cycle model** of savings and labor supply **with health risk under two assumptions**

This Paper

What We Do.

- **Estimate canonical structural life-cycle model** of savings and labor supply **with health risk under two assumptions**:

  A1. Health is perfectly observable (standard assumption)

This Paper

What We Do.

- **Estimate canonical structural life-cycle model** of savings and labor supply **with health risk under two assumptions**:
  - A1. Health is perfectly observable (standard assumption)
  - A2. Health is not observable, but a battery of noisy measures is

This Paper

What We Do.

- **Estimate canonical structural life-cycle model** of savings and labor supply **with health risk under two assumptions**:

  A1. Health is perfectly observable (standard assumption)

  A2. Health is not observable, but a battery of noisy measures is

- **Canonical structural model** (similar to French 2005):
  - Designed to fit the institutional set-up of the UK (ELSA data)
  - Exogenous: health, labor-productivity risk
  - Endogenous: consumption, savings, labor supply
  - Health affects pecuniary resources, time endowment, future health

This Paper

What We Do.

- **Counterfactuals in the two estimated models** (w/ and w/o ME in health)

This Paper

What We Do.

- **Counterfactuals in the two estimated models** (w/ and w/o ME in health)
  - We focus on the **costs of bad health**, as measured by foregone labor earnings, hours worked, consumption, and assets

This Paper

What We Do.

- **Counterfactuals in the two estimated models** (w/ and w/o ME in health)
  - We focus on the **costs of bad health**, as measured by foregone labor earnings, hours worked, consumption, and assets
  - **Costs of bad health in outcome X**: change the distribution of health shocks and calculate $\text{X}(\text{counterfactual}) - \text{X}(\text{benchmark})$

This Paper

What We Do.

- **Counterfactuals in the two estimated models** (w/ and w/o ME in health)
  - We focus on the **costs of bad health**, as measured by foregone labor earnings, hours worked, consumption, and assets
  - **Costs of bad health in outcome X**: change the distribution of health shocks and calculate $X(\text{counterfactual}) - X(\text{benchmark})$
- **Quantify the bias introduced by ME in health** by finding difference in counterfactuals between the two models

This Paper

What We Do.

- **Counterfactuals in the two estimated models** (w/ and w/o ME in health)
  - We focus on the **costs of bad health**, as measured by foregone labor earnings, hours worked, consumption, and assets
  - **Costs of bad health in outcome X**: change the distribution of health shocks and calculate $X(\text{counterfactual}) - X(\text{benchmark})$
- **Quantify the bias introduced by ME in health** by finding difference in counterfactuals between the two models

Findings.

This Paper

What We Do.

- **Counterfactuals in the two estimated models** (w/ and w/o ME in health)
  - We focus on the **costs of bad health**, as measured by foregone labor earnings, hours worked, consumption, and assets
  - **Costs of bad health in outcome X**: change the distribution of health shocks and calculate $\mathrm{X}(\text{counterfactual}) - \mathrm{X}(\text{benchmark})$
- **Quantify the bias introduced by ME in health** by finding difference in counterfactuals between the two models

Findings.

1. **Ignoring ME in health leads to underestimating the persistence of health and overestimating the fixed costs of participation**

This Paper

What We Do.

- **Counterfactuals in the two estimated models** (w/ and w/o ME in health)
  - We focus on the **costs of bad health**, as measured by foregone labor earnings, hours worked, consumption, and assets
  - **Costs of bad health in outcome X**: change the distribution of health shocks and calculate $X(\text{counterfactual}) - X(\text{benchmark})$
- **Quantify the bias introduced by ME in health** by finding difference in counterfactuals between the two models

Findings.

1. Ignoring ME in health leads to underestimating the persistence of health and overestimating the fixed costs of participation
2. Lower persistence of health and higher costs of participation lead to underestimating the costs of bad health by more than a factor of two

Our Contributions

1. Identification result for non-stationary Hidden Markov Models + novel estimator

Our Contributions

1. Identification result for non-stationary Hidden Markov Models $+$ novel estimator
2. Estimate structural life-cycle model with health risk and ME in health, taking into account ME in each stage of estimation

Our Contributions

1. Identification result for non-stationary Hidden Markov Models $+$ novel estimator
2. Estimate structural life-cycle model with health risk and ME in health, taking into account ME in each stage of estimation
3. Produce new estimates of the costs of bad health

Our Contributions

1. Identification result for non-stationary Hidden Markov Models + novel estimator
2. Estimate structural life-cycle model with health risk and ME in health, taking into account ME in each stage of estimation
3. Produce new estimates of the costs of bad health
4. Speak to structural literature
    - Previous studies likely to highly underestimate lifetime costs of bad health (eg, Capatina 2015, De Nardi et al 2024)
    - Future research needs to worry about ME in health

Rest of the Talk

1. Toy Model
2. Structural Model
3. Data and Estimation
   - Ignoring vs. Acknowledging Measurement Error in Health
4. The Costs of Bad Health
5. Conclusion

Toy Model

Simple Example

- Static model where workers can be in good or bad health, $H \in \{G, B\}$

Simple Example

- Static model where workers can be in good or bad health, $H \in \{G, B\}$
- Disutility of work varies with health, $\phi_H \in \{\phi_G, \phi_B\}$, where $\phi_B > \phi_G > 0$

Simple Example

- Static model where workers can be in good or bad health, $H \in \{G, B\}$
- Disutility of work varies with health, $\phi_H \in \{\phi_G, \phi_B\}$, where $\phi_B > \phi_G > 0$
- Individuals are heterogeneous in their Frisch elasticity ϵ_i

Simple Example

- Static model where workers can be in good or bad health, $H \in \{G, B\}$
- Disutility of work varies with health, $\phi_H \in \{\phi_G, \phi_B\}$, where $\phi_B > \phi_G > 0$
- Individuals are heterogeneous in their Frisch elasticity ϵ_i
- Workers solve the following problem:

$$\max_{c,h \geq 0} \quad c - \phi_H \frac{h^{1+\frac{1}{\epsilon_i}}}{1+\frac{1}{\epsilon_i}}$$
$$\text{s.t.} \quad c \leq wh$$

Simple Example

- Static model where workers can be in good or bad health, $H \in \{G, B\}$
- Disutility of work varies with health, $\phi_H \in \{\phi_G, \phi_B\}$, where $\phi_B > \phi_G > 0$
- Individuals are heterogeneous in their Frisch elasticity ϵ_i
- Workers solve the following problem:

$$\max_{c,h \geq 0} \quad c - \phi_H \frac{h^{1+\frac{1}{\epsilon_i}}}{1+\frac{1}{\epsilon_i}}$$
$$\text{s.t.} \quad c \leq wh$$

- For this model, we can identify the **labor supply effect of bad health**:

$$\mathbb{E}[\log h | H = G] - \mathbb{E}[\log h | H = B] = -\mathbb{E}[\epsilon_i](\log \phi_G - \log \phi_B)$$

Identification

Importance of Measurement Error

- Suppose we observe a noisy measure of the true binary health index, call it H^*, whose distribution conditional on true health H is:

$$H^* = \begin{cases} H & \text{w.p. } 1-p \\ e & \text{w.p. } p \in (0,1) \end{cases}, \qquad \text{where} \quad e = \begin{cases} G & \text{w.p. } 1/2 \\ B & \text{w.p. } 1/2 \end{cases}$$

Importance of Measurement Error

- Suppose we observe a noisy measure of the true binary health index, call it H^*, whose distribution conditional on true health H is:

$$H^* = \begin{cases} H & \text{w.p. } 1-p \\ e & \text{w.p. } p \in (0,1) \end{cases}, \qquad \text{where} \quad e = \begin{cases} G & \text{w.p. } 1/2 \\ B & \text{w.p. } 1/2 \end{cases}$$

- **Ignoring measurement error**:

$$\underbrace{\mathbb{E}[\log h | H^* = G] - \mathbb{E}[\log h | H^* = B]}_{\text{ESTIMATED labor-supply effect of bad health}} = \underbrace{(1-p)}_{\text{attenuation bias}} \times \underbrace{\big(\mathbb{E}[\log h | H = G] - \mathbb{E}[\log h | H = B]\big)}_{\text{TRUE labor-supply effect of bad health}}$$

(Similar to attenuation bias in classical regression analysis)

Importance of Measurement Error

- Suppose we observe a noisy measure of the true binary health index, call it H^*, whose distribution conditional on true health H is:

$$H^* = \begin{cases} H & \text{w.p. } 1-p \\ e & \text{w.p. } p \in (0,1) \end{cases}, \qquad \text{where} \quad e = \begin{cases} G & \text{w.p. } 1/2 \\ B & \text{w.p. } 1/2 \end{cases}$$

- **Ignoring measurement error**:

$$\underbrace{\mathbb{E}[\log h | H^* = G] - \mathbb{E}[\log h | H^* = B]}_{\text{ESTIMATED labor-supply effect of bad health}} = \underbrace{(1-p)}_{\text{attenuation bias}} \times \underbrace{\big(\mathbb{E}[\log h | H = G] - \mathbb{E}[\log h | H = B]\big)}_{\text{TRUE labor-supply effect of bad health}}$$

(Similar to attenuation bias in classical regression analysis)

- If we think health is complex and hard to measure, ME likely an important issue

Structural Model

Structural Model

- Individuals aged 50+ (ELSA core household members)
- Biannual life-cycle model: $t \in \{50\text{–}51, 52\text{–}53, \ldots, 86\text{–}87\}$
- Individuals decide how much to work, consume and save
  - Partial equilibrium
- Health affects pecuniary resources, time endowment, and mortality
- Government:
  - Taxes income (progressively)
  - Provides social security (in the form of pensions)
  - Gives mean-tested transfers

Structural Model

- The household's state vector:

$$\boldsymbol{X}_t = \Big(\underbrace{H_t}_{\text{health}}, \underbrace{a_t}_{\text{assets}}, \underbrace{ae_t}_{\substack{\text{average}\\ \text{earnings}}}, \underbrace{u_t}_{\substack{\text{persistent wage}\\ \text{component}}} \Big)$$

- Log-wages:

$$\log W_t(H_t, t) = a_0 + a_1 t + a_2 t^2 + a_H \mathbf{1}_{\{H_t = \text{Good}\}} + u_t$$
$$u_t = \rho u_{t-1} + \xi_t, \qquad \xi_t \sim \text{iid}$$

- **Average earnings** evolve according to:

$$ae_{t+1} = \frac{W_t N_t + (12 + t - 1) ae_t}{12 + t}$$

(We assume workers start working at age 26, hence the 12)

Structural Model

- **Spousal income** is a deterministic function of health and age:

$$ys_t = \begin{cases} ys(t, H_t), & \text{if } t \leq R_a + 1 \\ 0, & \text{if } t > R_a + 1 \end{cases}$$

- **Public pension benefits** are a function of average earnings at 64:

$$\text{pbb}_t = \begin{cases} g(ae_{64}), & t \geq R_a \\ 0, & \text{otherwise} \end{cases}$$

- **Private benefits** also a function of average earnings at age 64:

$$\text{privben}_t = \begin{cases} f(ae_{64}), & t \geq R_a \\ 0, & \text{otherwise} \end{cases}$$

(All choices motivated by data)

- Gvt gives transfers to household heads to ensure min level of consumption
  - Min level of consumption allowed to depend on age, $C_{\min,t}$
    (Retirees face different mean-tested programs than non-retirees in the UK)
  - $C_{\min,t}$ changes at $t = R_a$:

$$C_{\min,t} = \begin{cases} C^y_{\min}, & t < R_a \\ C^o_{\min}, & t \geq R_a \end{cases}$$

- Government transfers:

$$\text{tr}_t = \begin{cases} \max\left\{0, C_{\min,t} - \left[W_t N_t + (1+r)a_t + ys(t, H_t)\right]\right\}, & t < R_a \\ \max\left\{0, C_{\min,t} - \left[W_t N_t + (1+r_t)a_t + ys(t, H_t) + \text{pbb}_t + \text{privben}_t\right]\right\}, & t \geq R_a \end{cases}$$

Structural Model

- Households face **no-borrowing constraint**: $a_{t+1} \geq 0$
- Save at **constant interest rate** r
- **Tax system** captured by *after-tax* function $y_t(\cdot)$
- **Household's budget constraint**:

$$c_t + a_{t+1} = y_t(ra_t + W_t N_t) + ys(t, H) + a_t + \text{tr}_t \qquad (\text{working age: } t < R_a)$$

$$c_t + a_{t+1} = y_t(ra_t + W_t N_t + \text{pbb}_t + \text{privben}_t) + ys(t, H) + a_t + \text{tr}_t \qquad (\text{retirement age: } t \geq R_a)$$

Structural Model

Household head's decision problem:

$$\max_{c_t, N_t, a_{t+1} \geq 0} \quad \mathbb{E}_0 \sum_{t=0}^{T} \beta^t \left[\underbrace{\frac{s_t(H_t)}{1-\gamma} \left\{ c_t{}^{\nu} \left[\overline{L} - \phi_P \mathbf{1}_{\{N_t > 0\}} - N_t - \phi_H \mathbf{1}_{\{H_t = \text{Bad}\}} \right]^{1-\nu} \right\}^{1-\gamma}}_{\text{utility conditional on survival}} + \underbrace{\left[1 - s_t(H_t)\right] b(a_{t+1})}_{\text{utility at death}} \right]$$

subject to

$$b(a_{t+1}) = \theta_B \frac{(\kappa_B + a_{t+1})^{(1-\gamma)\nu}}{1-\gamma}$$

Budget constraint

Transition functions

Initial conditions

Numerical Procedure

Data and Estimation

Data

- We use data from ELSA (English Longitudinal Study of Ageing):
  - Survey data
  - Representative of the old English population
  - Individuals aged +50 (and their partners)
  - Bi-annual interviews since 2002

Data

- We use data from ELSA (English Longitudinal Study of Ageing):
  - Survey data
  - Representative of the old English population
  - Individuals aged +50 (and their partners)
  - Bi-annual interviews since 2002
- We focus on the UK to avoid unnecessary complications

Data

- We use data from ELSA (English Longitudinal Study of Ageing):
  - Survey data
  - Representative of the old English population
  - Individuals aged $+50$ (and their partners)
  - Bi-annual interviews since 2002
- We focus on the UK to avoid unnecessary complications
  - No need to model employer-provided health insurance and OOP medical expenditures
  - NHS provides universal health care
  - Private health care used by $\approx$ 10% as top-up to NHS

Estimation

Two-step estimation procedure

1. Estimate some parameters outside the model and set some others to values taken from the literature

Estimation

Two-step estimation procedure

1. Estimate some parameters outside the model and set some others to values taken from the literature
   - **Tax parameters and interest rate**: adapted from O'Dea 2018 See
   - **Wage shock parameters**: FE regression + Minimum Distance See
   - **Pension parameters**: OLS regressions on average earnings See
   - **Spousal income, average earnings, initial distribution of states**
   - **Health process**: next slides

Estimation

Two-step estimation procedure

1. Estimate some parameters outside the model and set some others to values taken from the literature
    - **Tax parameters and interest rate**: adapted from O'Dea 2018 See
    - **Wage shock parameters**: FE regression + Minimum Distance See
    - **Pension parameters**: OLS regressions on average earnings See
    - **Spousal income, average earnings, initial distribution of states**
    - **Health process**: next slides
2. Estimate remaining parameters inside model using Indirect Inference

Parameters Estimated Outside the Model

- Two options for health process:

- Two options for health process:
    1. Ignore ME in health

Parameters Estimated Outside the Model

- Two options for health process:
    1. Ignore ME in health
        - Consider SRHS as “true health”

Poor	Fair	Good	Very Good	Excellent

- Two options for health process:
  1. Ignore ME in health
     - Consider SRHS as "true health"

Poor	Fair	Good	Very Good	Excellent

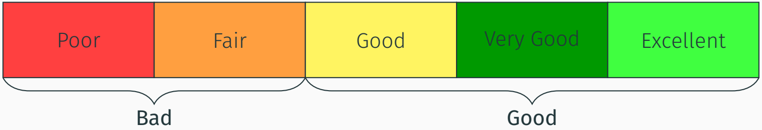
     - Classify individuals into good and bad health

$$\underbrace{\text{Poor} \quad \text{Fair}}_{\text{Bad}} \quad \underbrace{\text{Good} \quad \text{Very Good} \quad \text{Excellent}}_{\text{Good}}$$

Parameters Estimated Outside the Model

- Two options for health process:
  1. Ignore ME in health
     - Consider SRHS as “true health”

     - Classify individuals into good and bad health



     - Obtain health dynamics using empirical transition probabilities for each age group

Parameters Estimated Outside the Model

- Two options for health process:

    2. Acknowledge ME in health

Parameters Estimated Outside the Model

- Two options for health process:
  2. Acknowledge ME in health
     - Consider “true health” as unobservable

Parameters Estimated Outside the Model

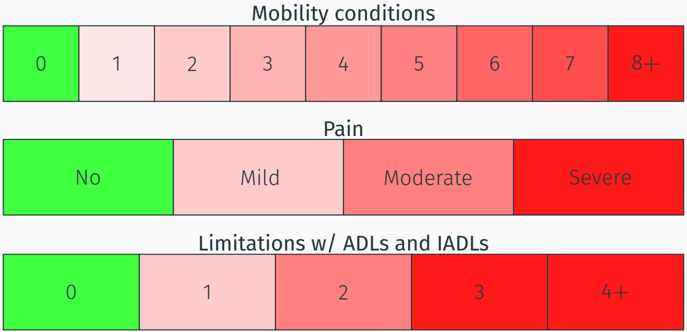
- Two options for health process:

  2. Acknowledge ME in health

     - Consider "true health" as unobservable
     - Find (at least) three noisy measures of health

Parameters Estimated Outside the Model

- **Two options for health process:**
  2. Acknowledge ME in health
     - Consider “true health” as unobservable
     - Find (at least) three noisy measures of health



Parameters Estimated Outside the Model

- **Two options for health process:**

   2. Acknowledge ME in health

      - Consider “true health” as unobservable
      - Find (at least) three noisy measures of health

Mobility conditions

0	1	2	3	4	5	6	7	8+

Pain

No	Mild	Moderate	Severe

Limitations w/ ADLs and IADLs

0	1	2	3	4+

      - Build + estimate ME model for health (non-stationary hidden Markov model)

Parameters Estimated Outside the Model

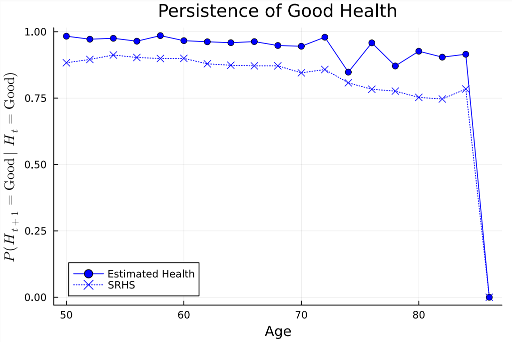
Health Process

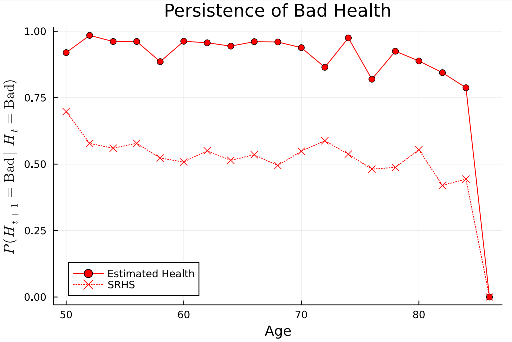
1. Ignoring ME: health and its dynamics identified and estimated using SRHS and empirical transition probabilities between health states
2. Acknowledging ME: true health unobservable, but noisy measures available. Use non-stationary hidden Markov model for health ME Model Identification Estimation Algorithm More

Parameters Estimated Outside the Model

Health Process

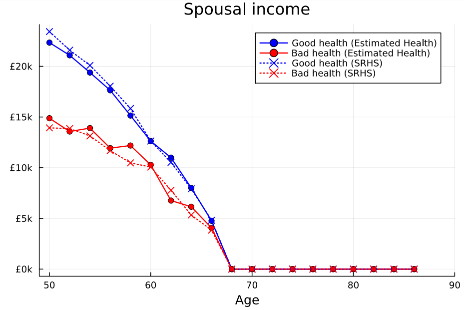
1. **Ignoring ME**: health and its dynamics identified and estimated using SRHS and empirical transition probabilities between health states
2. **Acknowledging ME**: true health unobservable, but noisy measures available. Use non-stationary hidden Markov model for health ME Model Identification Estimation Algorithm More





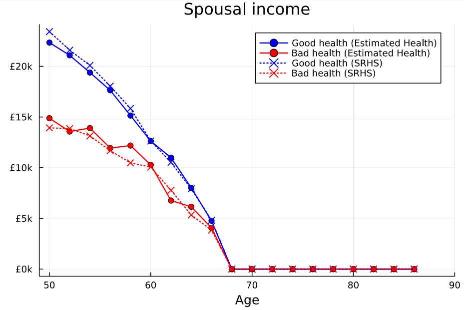
Parameters Estimated Outside the Model

- **Spousal income.** Empirical counterpart of the spousal income function is given by $\mathrm{ys}(t, H) = \mathbb{E}[\mathrm{ys}_{it} \mid H_{it}]$ Identification & Estimation with ME



Parameters Estimated Outside the Model

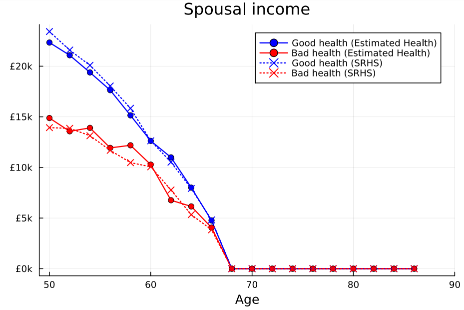
- **Spousal income.** Empirical counterpart of the spousal income function is given by $\text{ys}(t, H) = \mathbb{E}\left[\text{ys}_{it} \mid H_{it}\right]$ Identification & Estimation with ME



- Need a measure of **average earnings** (state variable), but we only observe individuals starting at age 50 (at the earliest) Average Earnings

Parameters Estimated Outside the Model

- **Spousal income.** Empirical counterpart of the spousal income function is given by $\text{ys}(t, H) = \mathbb{E}\left[\text{ys}_{it} \mid H_{it}\right]$ Identification & Estimation with ME



- Need a measure of **average earnings** (state variable), but we only observe individuals starting at age 50 (at the earliest) Average Earnings
- **Initial distribution of states** needed to simulate model Details

Parameters Estimated Inside the Model

We target profiles for assets, hours worked, participation (all by measured health), and the coefficients of the FE regression for wages

Parameters Estimated Inside the Model

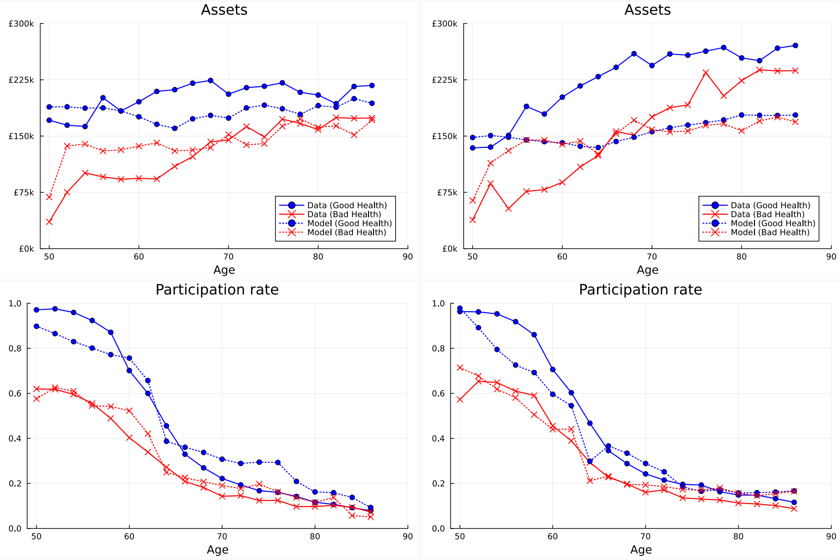
We target profiles for assets, hours worked, participation (all by measured health), **and the coefficients of the FE regression for wages**

Parameter	Acknowledging ME	Ignoring ME
β: bi-annual discount factor	0.99	1.06
γ: CRRA coefficient	4.32	4.88
ν: consumption weight in utility function	0.55	0.65
$\boldsymbol{\phi_P}$: **fixed cost of participation**	1,230 hours	2,437 hours
θ_B: weight on bequest	0.013	0.37
ϕ_H: time cost of bad health	130 hours	123 hours
$C^y_{\min}$: consumption floor when young	£20,337	£25,112
$C^o_{\min}$: consumption floor when old	£47,332	£60,420
a_0: constant term of wage profile	1.73	1.93
a_1: linear age-term of wage profile	0.03	0.01
a_2: quadratic age-term of wage profile	−0.002	−0.001
a_H: health coefficient of wage profile	0.279	0.085

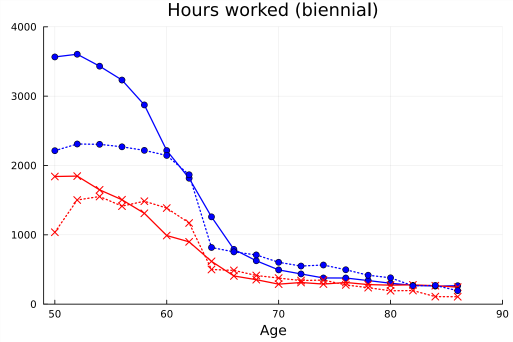
Table Notes. Parameters ϕ and $C_{\min}$ should be interpreted in terms of bi-annual hours and bi-annual GBP, respectively.

Model Fit: accounting for ME (left), ignoring it (right)

Data Profiles



Model Fit: accounting for ME (left), ignoring it (right)



- Similar fit when taking into account and ignoring ME
- Sometimes missing the levels, roughly capturing the trends
- Some trouble in fitting data (work in progress)

The Costs of Bad Health

The Costs of Bad Health

- We follow De Nardi, Pashchenko, and Porapakkarm 2024 and calculate the costs of bad health as measured by many outcomes

The Costs of Bad Health

- We follow De Nardi, Pashchenko, and Porapakkarm 2024 and calculate the costs of bad health as measured by many outcomes
- Exercise:
    1. Simulate model imposing everyone is always in good health
        → Individuals' histories of earnings, hours worked, consumption, assets

The Costs of Bad Health

- We follow De Nardi, Pashchenko, and Porapakkarm 2024 and calculate the costs of bad health as measured by many outcomes
- Exercise:
  1. Simulate model imposing everyone is always in good health
     - → Individuals' histories of earnings, hours worked, consumption, assets
  2. Simulate model letting health evolve according to estimated transition matrix
     - → Individuals' histories of earnings, hours worked, consumption, assets

The Costs of Bad Health

- We follow De Nardi, Pashchenko, and Porapakkarm 2024 and calculate the costs of bad health as measured by many outcomes
- Exercise:
  1. Simulate model imposing everyone is always in good health
     $\rightarrow$ Individuals' histories of earnings, hours worked, consumption, assets
  2. Simulate model letting health evolve according to estimated transition matrix
     $\rightarrow$ Individuals' histories of earnings, hours worked, consumption, assets
  3. Find differences in annual means between histories in (1) and (2)

The Costs of Bad Health

- We follow De Nardi, Pashchenko, and Porapakkarm 2024 and calculate the costs of bad health as measured by many outcomes
- Exercise:
  1. Simulate model imposing everyone is always in good health
     → Individuals' histories of earnings, hours worked, consumption, assets
  2. Simulate model letting health evolve according to estimated transition matrix
     → Individuals' histories of earnings, hours worked, consumption, assets
  3. Find differences in annual means between histories in (1) and (2)
- **Do this exercise twice** (with and without taking into account ME)

The Costs of Bad Health

- We follow De Nardi, Pashchenko, and Porapakkarm 2024 and calculate the costs of bad health as measured by many outcomes
- Exercise:
  1. Simulate model imposing everyone is always in good health
     → Individuals' histories of earnings, hours worked, consumption, assets
  2. Simulate model letting health evolve according to estimated transition matrix
     → Individuals' histories of earnings, hours worked, consumption, assets
  3. Find differences in annual means between histories in (1) and (2)
- **Do this exercise twice** (with and without taking into account ME)
- **Find difference in counterfactuals between models**

The Costs of Bad Health (all individuals)

Ignoring ME in health leads to substantially underestimating costs of bad health

Outcome	Taking into account ME	Ignoring ME	Difference (%)
Earnings	£1,477	£784	88%
Hours worked	144	78	85%
Consumption	£2,695	£2,259	19%
Assets	£21,445	£16,951	27%

Notes. All variables are means in annual terms. Mean earnings and hours worked computed until age 65.

The Costs of Bad Health (all individuals)

Ignoring ME in health leads to substantially underestimating costs of bad health

Outcome	Taking into account ME	Ignoring ME	Difference (%)
Earnings	£1,477	£784	88%
Hours worked	144	78	85%
Consumption	£2,695	£2,259	19%
Assets	£21,445	£16,951	27%

Notes. All variables are means in annual terms. Mean earnings and hours worked computed until age 65.

Mainly **two forces driving the results**:

1. Higher **persistence of health** when taking into account ME
2. Lower **fixed costs of participation** when taking into account ME

The Costs of Bad Health (initially unhealthy)

Costs of bad health higher than for the overall population (since health is persistent)

Outcome	Acknowledging ME	Ignoring ME	Difference (%)
Earnings	£7,027	£3,218	118%
Hours worked	718	362	98%
Consumption	£5,345	£3,405	57%
Assets	£30,685	£18,160	69%

Notes. All variables are means in annual terms. Mean earnings and hours worked are computed until age 65

Conclusion

Taking Stock

- **Question**: How important is the imperfect observability of health to evaluate the costs of bad of health?
- **Method**:
  1. **Estimate dynamic structural life-cycle model under two assumptions**:
     - A1. Health is perfectly observable (standard)
     - A2. Health is not observable, but noise measures are
  2. **Calculate costs of bad health** (foregone earnings, hours worked, consumption, assets)
  3. **Calculate bias introduced by ME in health by finding difference b/w models**
- **Answer**: Ignoring ME leads to underestimating costs of bad health by 20–120%
- **Implications**: Literature likely to highly underestimate lifetime costs of bad health

Questions?

Thank You!

(Email: luisperez@smu.edu)

(Website: `https://luisperezecon.com`)

Appendix

Identification: Toy Model

- Suppose that the cross-sectional distribution of Frisch elasticities is known and independent of health

- Then **we can identify** ϕ_B **and** ϕ_G **from** $\mathbb{E}[\log h|H = B]$ **and** $\mathbb{E}[\log h|H = G]$**, respect.** To see why note that:

$$\mathbb{E}[\log h|H] = \mathbb{E}[\epsilon_i] \log w - \mathbb{E}[\epsilon_i] \log \phi_H$$

- Given this, **we can identify the labor supply effect of bad health**:

$$\mathbb{E}[\log h|H = G] - \mathbb{E}[\log h|H = B] = -\mathbb{E}[\epsilon_i](\log \phi_G - \log \phi_B)$$

(In this very simple case this quantity is directly observable)

Back

Model Solution: Numerical Procedure

- There are four states (apart from age): Assets, health, average earnings, and stochastic component of wages
  - Health already discrete. Rest are discretized and placed on a grid
- There two continuous choices: Assets tomorrow and hours worked
  - Also discretized and placed on a grid.
- Wage shock discretized using extension to life-cycle models of the Rouwenhorst method by Fella Gallipolli Fan 2019
  - This produces a transition matrix and a grid for each age
- Value function at each age found by backward induction
  - Given value function at $t+1$ problem at t solved by grid search
- Expectations of the value function are taken using the transition function for health and for the discretized wage shocks
- Average earnings tomorrow can be outside the grid $\Longrightarrow$ use linear interpolation

Parameters from Literature Back

Parameter	Value	Source
κ_B: curvature of bequests	650,000	O'Dea 2018
$\overline{L}$: total endowment of bi-annual hours	8,760	12 daily hours
r: interest rate non-housing wealth	0.0323	O'Dea 2018

Table 1: Income Tax Thresholds from O'Dea 2018

Parameter	Age		
	< 64	64–73	≥ 74
κ_1	16,210	21,000	21,200
κ_2	84,940	89,740	89,940

$$\text{Income taxes}(ti, t) = \begin{cases} 0, & \text{if } ti \leq \kappa_1^t \\ 0.2(ti - \kappa_1^t), & \text{if } \kappa_1^t < ti \leq \kappa_2^t \\ 0.2(\kappa_2^t - \kappa_1^t) + 0.4(ti - \kappa_2^t), & \text{if } \kappa_2^t < ti \end{cases},$$

Parameters Estimated Outside the Model Back

- **Wage parameters.** FE regression:

$$\log(W_{it}^{\text{data}}) = a_0 + a_1 t + a_2 t^2 + a_H \mathbf{1}_{\{H_{it}=\text{Good}\}} + \underbrace{\eta_i + u_{it} + m_{it}}_{=\varepsilon_{it}}$$

$$u_{it} = \rho u_{it-1} + \xi_t, \quad \xi_1 \sim \mathcal{N}(0, \sigma^2_{\xi,1}), \quad \xi_t \sim \mathcal{N}(0, \sigma^2_{\xi,t}) \ \forall t > 1,$$

$$m_{it} \sim \mathcal{N}(0, \sigma^2_m), \qquad \eta_i \sim \mathcal{N}(0, \sigma^2_\eta), \qquad \eta_i \perp \xi_t, \ \forall i, t.$$

Wage parameters biased due to selection of workers in labor market

Identification of wage-shock parameters

Minimum Distance Estimates **Wage-shock parameter**	**Health (with DLs)**	**Health (with SRHS)**
ρ: autocorrelation	0.8764	0.8790
σ^2_η: variance fixed effect	0.0806	0.0775
$\sigma^2_{\xi,1}$: variance innovations at $t = 1$	0.1949	0.1991
$\sigma^2_{\xi,t}$: variance innovations at $t > 1$	0.0581	0.0577
σ^2_m: variance measurement error	0.1645	0.1645

Parameters Estimated Outside the Model Back

- **Pension parameters**. Estimate parameters that relate public and private pension benefits with avg. earnnigs with OLS:

$$pbb_{i,65} = ss_1 ae_{i,64} + ss_2 ae^2_{i,64} + \varepsilon_i, \qquad \text{for} \;\; ae_{i,64} \leq \widehat{ae}_{ss},$$
$$priv ben_{i,65} = pp_0 + pp_1 ae_{i,64} + pp_2 ae^2_{i,64} + \xi_i,$$

($\widehat{ae}_{ss} = 75k$: threshold at which quadratic relationship between pbb and ae starts to decrease)

Parameter	Value	S.E.
ss_1	0.65	0.00
ss_2	−3.56E-06	1.31E-08
pp_0	5,981	591
pp_1	0.34	0.03
pp_2	5.23E-07	2.45E-07

Measurement Error Model for Health

1. At each time t, **individual can be in one of $r-1$ different unobserved health states** $H_t \in \{1, 2, \ldots, r\}$, where $r =$ dead
   - Eg, $H_t \in \{\text{Good health } (=1), \text{Bad health } (=2), \text{Dead } (=3)\}$
2. **Health evolves according to non-stationary Markov model** with transition matrices $\{K_t\}$, where
$$K_t(j,k) := \mathbb{P}_t(H_{t+1} = k | H_t = j)$$
3. **Econometrician cannot observe true health status** (except for mortality), **but can observe at least 3 discrete noisy measures**:
$$Y_t^m \in \{1, \ldots, \kappa_m, \kappa_{m+1}\}$$
   - Eg, $\boldsymbol{Y}_t = \{\text{Mobility conditions, Pain severity, \#ADL + \#IADL limitations}\}$
4. Conditional distribution of Y_t^m is given by matrix P_t^m, where
$$P_t^m(c,j) := \mathbb{P}(Y_t^m = c \mid H_t = j)$$

Back

Identification of ME Model for Health

Assumptions:

- A1. Access to three conditionally-independent noisy measures Y_t^m of unobserved state H_t
- A2. (i) $\mathbb{P}(H_{t+1} \mid H_t, Y_t^1) = \mathbb{P}(H_{t+1} \mid H_t)$ for $t = 0, \ldots, T-1$
  (ii) $\mathbb{P}(Y_t^1 \mid H_{t+1}, Y_{t+1}^1) = \mathbb{P}(Y_t^1 \mid H_{t+1})$ for $t = 0, \ldots, T$
- A3. P_t^m, the conditional distribution of Y_t^m, is full rank for $m \in \{1, 2, 3\}$
- A4. Cross-sectional distribution of underlying state, π_t, is s.t. $\pi_t(c) > 0$ for each $c \in \{1, \ldots, r\}$
- A5. $\exists i, m^*$ known by the researcher s.t. for row i of matrix P^{m^*} we have $P^{m^*}(i,j) \neq P^{m^*}(i,j')$ for all columns and $P^{m^*}(i,j)$ is monotone in j

Identification

Suppose Assumptions A1–A5 hold. Then the model is identified.

Identification of ME Model for Health

- Identification idea:
  1. *Cross-sectional step* identifies cross-sectional parameters $\pi_t, \{P_t^m\}$
  2. *Longitudinal step* identifies transition matrices for state, $\{K_t\}_{t=0}^{T-1}$

- **Proof**. Identification argument for $K_0, \pi_0, \pi_1, \{P^m\}_{m=1,2,3}$. Back
  - From Bonhomme et al 2016, Theorems 2–3: $\pi_0, \pi_1, \{P^m\}_{m=1,2,3}$ are identified
    WTS: K_0 is identified
  - Note that the joint distribution of noisy measure $m = 1$ at $t = 0, 1$ is:
    $$\mathbb{P}(Y_0^1, Y_1^1) = P^1 \Pi_0 K_0 \Pi_1^{-1} (P^1 \Pi_0)',$$
    where $\mathbb{P}(Y_0^1, Y_1^1)(i, j) = \mathbb{P}(Y_0^1 = i, Y_1^1 = j)$
  - Let $\Omega_0 = P_1 \Pi_0$ and $\Omega_1 = \Pi_1^{-1}(P^1 \Pi_1)'$. Since P^1 is full rank (A3):
    $$K_0 = (\Omega_0' \Omega_0)^{-1} \Omega_0' \mathbb{P}(Y_0^1, Y_1^1) \Omega_1' (\Omega_1' \Omega_1)^{-1}$$
  - Since $\mathbb{P}(Y_0^1, Y_1^1)$ is observable and Π_0, Π_1, P^1 are identified, this completes the proof

ME Model: Estimation Algorithm

- **First Step of Constrained Baum–Welch**:
  - Restrict the sample to observations that are not missing or death (ie, $Y_t^1 \neq \kappa_1, -7$)
  - Use ML and EM algorithm to get $\sqrt{N}$-consistent and asymp. normal estimates of P^1 and $\tilde{\pi}_t = \{\mathbb{P}(H_t = s \mid H_t \neq r\}_{s=1,\ldots,r}$
  - Calculate proportion of people that dies between t and $t+1$:
  $$\text{Prop. of deaths}_{t,t+1} = \hat{\mathbb{P}}(Y_{t+1}^1 = \kappa_1 \mid Y_t^1 \neq \kappa_1, -7).$$
  - Let
  $$\pi_{t+1}^{H_t \neq r} = \left\{\mathbb{P}(H_{t+1} = s \mid H_t \neq r)\right\}_{s=1,\ldots,r}$$
  - A consistent estimate of this object is:
  $$\hat{\pi}_{t+1}^{H_t \neq r} = \left(\hat{\tilde{\pi}}_{t+1}(1 - \text{Prop. of deaths}_{t,t+1}),\ \text{Prop. of deaths}_{t,t+1}\right),$$
  where $\hat{\tilde{\pi}}_{t+1}$ denotes the estimate for $\tilde{\pi}_{t+1}$

ME Model: Estimation Algorithm

- **Second Step of Constrained Baum–Welch**:
  - For each $t = 1, \ldots, T-1$, restrict the sample to observations to those that are non-missing in t and $t+1$ and non-death in t
  - Estimate K_t iterating between an E and a M step until convergence
    - **E step**: Let Q_1 be the emission matrix for Y^1 expanded to include mortality. Given estimates for $\hat{\hat{\pi}}_t, \hat{\pi}_{t+1}^{H_t \neq r}, Q^1, \{Y_{i,\tau}^m\}_{\tau=t,t+1}$ and a guess for $K_t^{(h)}$ calculate the filtered probabilities:

$$\hat{v}_{i,k,j} := \mathbb{P}\big(H_{i,t+1} = j, H_{i,t} = k \mid Y_{i,t}^1, Y_{i,t+1}^1, \{\hat{\pi}_\tau\}_{\tau=t,t+1}, \hat{Q}^1, K_t^{(h)}\big)$$

These filtered probabilities can be computed as:

$$\hat{v}_{i,k,j} = \frac{\hat{Q}^1(y_{i,t}^1, k)\hat{\pi}_t(k)K^{(h)}(k,j)\hat{Q}^1(y_{i,t+1}^1, j)}{\sum_{j=1}^r \sum_{k=1}^r \hat{Q}^1(y_{i,t}^1, k)\hat{\pi}_t(k)K^{(h)}(k,j)\hat{Q}^1(y_{i,t+1}^1, j)}$$

ME Model: Estimation Algorithm

- **M step**: Calculate the new guess $K_t^{(h+1)}$ as:

$$K_t^{(h+1)} = \arg\max_{K} \quad \sum_{i=1}^{N} \left\{ \sum_{k=1}^{r} \sum_{j=1}^{r} v_{i,j,k} \log \big(K(k,j) \big) \right\}$$

$$\text{s.t.} \quad \sum_{j=1}^{r} K_t(k,j) = 1, \quad \forall k,$$

$$\sum_{j=1}^{r} K_t(j,c) \hat{\hat{\pi}}_t(j) = \hat{\pi}_{t+1}^{H_t \neq r}(c)$$

Back

Prediction Power of Health Measures

- We analyze the performance of estimated health (correcting for ME) *vis-a-vis* SRHS
  - Binary measure of estimated health: mode of posterior distribution of estimated health (posterior distr. with Bayes' rule, emission matrices, and distribution of health types at each age)

- We regress outcome vars (annual hours worked, annual income, LFP, institutionalization) on age, gender, the health measure, and interaction terms of health with age and gender

- Findings:
  1. Estimated health weakly better than SRHS at explaining and predicting econ outcomes
  2. Health gradient of estimated health significantly larger than that of SRHS in OLS regs
     $\rightarrow$ SRHS is a noisy measure
  3. Health gradient of estimated health significantly larger ($\times 4$) than that of SRHS in FE regs
     $\rightarrow$ Only health shocks and ME left to account for variation in health since FE absorbs health types

OLS & Logit regressions | FE regressions | Back

Prediction Power of Health Measures: OLS

Gradient of estimated health on annual hours worked 20–30% larger than that of SRHS

	(1) Today		(2) 2 years from now		(3) 4 years from now	
Estimated health	572.40***		529.13***		478.37***	
	(15.16)		(18.89)		(23.49)	
SRHS		448.05***		402.76***		389.77***
		(14.38)		(18.26)		(24.11)
R-squared	0.41	0.40	0.40	0.39	0.38	0.38
Observations	44,833	44,833	28,970	28,970	18,196	18,196

Table Notes. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is actual hours worked per year at different points in time, as indicated in the columns. All regressions include controls for age and gender, and interaction terms of health with age and gender.

Prediction Power of Health Measures: OLS

Gradient of estimated health on annual income 20–30% larger than that of SRHS

	(1) Today		(2) 2 years from now		(3) 4 years from now	
Estimated health	8949.68***		2628.51		12277.91***	
	(2804.59)		(3488.42)		(4202.05)	
SRHS		7334.71***		10986.50***		9326.86***
		(2090.74)		(2528.53)		(3220.94)
R-squared	0.04	0.04	0.06	0.06	0.07	0.07
Observations	12,710	12,710	7,455	7,455	4,102	4,102

Table Notes. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is annual income in GBP at different points in time, as indicated in the columns. All regressions include controls for age and gender, and interaction terms of health with age and gender.

Prediction Power of Health Measures: Logit

Odds of participating in labor market are 9 $(= e^{2.21})$ times higher for individuals in good estimated health vs. 5 $(= e^{1.59})$ times higher for individuals with good SRHS

	(1)		(2)		(3)	
	Today		2 years from now		4 years from now	
Estimated health	2.21***		2.09***		1.96***	
	(0.06)		(0.08)		(0.10)	
SRHS		1.59***		1.49***		1.49***
		(0.05)		(0.07)		(0.09)
Pseudo R-squared	0.36	0.34	0.33	0.32	0.31	0.30
Observations	44,757	44,757	28,897	28,897	18,169	18,169

Table Notes. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is a dummy variable for labor-force participation at different points in time, as indicated in the columns. All regressions include controls for age and gender, and interaction terms of health with age and gender.

Prediction Power of Health Measures: Logit

Odds of being institutionalized are 5 $(= e^{1.56})$ times lower for individuals in good estimated health vs. 4 $(= e^{1.59})$ times lower for individuals with good SRHS Back

	(1) Today		(2) 2 years from now		(3) 4 years from now	
Estimated health	-1.56**		-2.17***		-1.39***	
	(0.77)		(0.71)		(0.49)	
SRHS		-1.46*		-2.11***		-1.32***
		(0.77)		(0.71)		(0.50)
Pseudo R-squared	0.04	0.04	0.06	0.06	0.05	0.05
Observations	5,771	5,771	5,985	5,985	6,689	6,689

Table Notes. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is a dummy variable for institutionalization (i.e., entry in nursing homes, mental health facilities, prisons) at different points in time, as indicated in the columns. All regressions include controls for age and gender, and interaction terms of health with age and gender.

Prediction Power of Health Measures: FE

Gradient of estimated health on annual hours worked 5 times larger than that of SRHS

	(1)		(2)		(3)	
	Today		2 years from now		4 years from now	
Estimated health	107.69***		13.27		18.74	
	(19.53)		(25.33)		(31.76)	
SRHS		19.95**		1.13		-20.09
		(10.14)		(12.72)		(15.97)
Age interacted with health	✓	✓	✓	✓	✓	✓
Age interacted with gender	✓	✓	✓	✓	✓	✓
Individual fixed effects	✓	✓	✓	✓	✓	✓
R-squared	0.16	0.16	0.13	0.13	0.13	0.13
Observations	44,833	44,833	28,970	28,970	18,196	18,196

Table Notes. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is actual hours worked per year at different points in time, as indicated in the columns.

Prediction Power of Health Measures: FE

Gradient of estimated health on annual income larger than that of SRHS

	(1)		(2)		(3)	
	Today		2 years from now		4 years from now	
Estimated health	9840.45**		-360.88		5649.00	
	(4825.49)		(5434.23)		(5364.18)	
SRHS		-823.18		-2061.81		65.20
		(2318.84)		(2550.88)		(2596.47)
Age interacted with health	✓	✓	✓	✓	✓	✓
Age interacted with gender	✓	✓	✓	✓	✓	✓
Individual fixed effects	✓	✓	✓	✓	✓	✓
R-squared	0.01	0.01	0.01	0.01	0.03	0.03
Observations	12,710	12,710	7,455	7,455	4,102	4,102

Table Notes. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is annual income in GDP at different points in time, as indicated in the columns.

Prediction Power of Health Measures: FE

Gradient of estimated health on participation larger than that of SRHS Back

	(1) Today		(2) 2 years from now		(3) 4 years from now	
Estimated health	0.74***		-0.05		-0.57	
	(0.19)		(0.27)		(0.47)	
SRHS		0.36***		0.04		-0.25
		(0.08)		(0.11)		(0.17)
Age interacted with health	✓	✓	✓	✓	✓	✓
Age interacted with gender	✓	✓	✓	✓	✓	✓
Individual fixed effects	✓	✓	✓	✓	✓	✓
Log likelihood	-3,270	-3,259	-1,758	-1,758	-883	-882
Observations	13,549	13,549	7,193	7,193	3,797	3,797

Table Notes. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is a dummy variable for labor-force participation at different points in time, as indicated in the columns.

Identification of Wage-shock Parameters

- Let $\varepsilon_{it} = \eta_i + u_{it} + m_{it}$ denote the residuals from the wage equation (in levels)
- Using recursion on u_{it} $(= \rho u_{it-1} + \xi_t)$, we can write:

$$\varepsilon_{it} = \eta_i + m_{it} + \sum_{\tau=1}^{t} \rho^{t-\tau} \xi_\tau, \qquad t = 1, \ldots, T$$

- Identification:

$$1 + \rho = \frac{\text{Cov}(\varepsilon_{i4} - \varepsilon_{i2}, \varepsilon_{i1})}{\text{Cov}(\varepsilon_{i3} - \varepsilon_{i2}, \varepsilon_{i1})}, \qquad \sigma^2_{\xi,1} = \frac{\text{Cov}(\varepsilon_{i4} - \varepsilon_{i2}, \varepsilon_{i1})}{\rho(\rho^2 - 1)},$$
$$\sigma^2_m = (\rho - 1)\sigma^2_{\xi,1} - \text{Cov}(\varepsilon_{i2} - \varepsilon_{i1}, \varepsilon_{i1}), \qquad \sigma^2_{\xi,t} = \text{Var}(\varepsilon_{i2} - \varepsilon_{i1}) - 2\sigma^2_m - (\rho - 1)^2 \sigma^2_{\xi,1},$$
$$\sigma^2_\eta = \text{Var}(\varepsilon_{i2}) - \rho^2 \sigma^2_{\xi,1} - \sigma^2_m - \sigma^2_{\xi,t}$$

- Target (+80) additional statistics of wage-shock process to increase MDE precision

Back

Spousal Earnings: Identification & Estimation with ME

- Empirical counterpart of the spousal income function:

$$\mathsf{ys}(t, H) = \mathbb{E}\left[\mathsf{ys}_{it} \mid H_{it}\right]$$

- Health is not directly observable, but ys can be estimated using Minimum Distance. **Identifying assumption**:

Exclusion restriction

$$\mathbb{E}\left[\mathsf{ys}_t \mid Y_t^1, H_t = c\right] = \mathbb{E}\left[\mathsf{ys}_t \mid H_t = c\right].$$

This amounts to saying "*Given true health, suffering from problems, say, with mobility conditions does not predict spousal income*"

Spousal Earnings: Identification & Estimation with ME

- Under exclusion restriction, we can write expectation of spousal earnings given Y_t^1 as:

$$\mathbb{E}[\mathrm{ys}_t \mid Y_t^1 = y] = \sum_{c=1}^{2} \mathbb{E}[\mathrm{ys}_t \mid H_t = c]\mathbb{P}(H_t = c \mid Y_t^1 = y), \qquad y = 1, \dots, \kappa_1$$

- This can be written as the following linear system:

$$\mathbb{P}(Y_t^1)^{-1} P^1 \Pi_t \mathbb{E}[\mathrm{ys}_t \mid H_t] = \mathbb{E}[\mathrm{ys}_t \mid Y_t^1],$$

  where:
  - $\mathbb{E}[\mathrm{ys}_t \mid H_t]$: column vector whose i-th element is $\mathbb{E}[\mathrm{ys}_t \mid H_t = i]$
  - $\mathbb{E}[\mathrm{ys}_t \mid Y_t^1]$: column vector whose i-th element is $\mathbb{E}[\mathrm{ys}_t \mid Y_t^1 = i]$
  - $\mathbb{P}(Y_t^1)$: diagonal matrix with cross-sectional distribution of Y_t^1 at each t in the diagonal

- This system has at most one solution if P^1 is full rank and if π_t has non-zero elements for all t (already required for ME identification)

- Estimation imposing this linear-system of restrictions in finite sample using MD Back

Average Earnings

Since ELSA starts at age 50, we need retrospective information on employment and earnings to construct measure of average earnings

- **Ideally.** Administrative data to obtain average earnings
  - Restricted to UK-affiliated researchers
- **Go-around #1.** ELSA Life History + ELSA surveys to construct employment spells and earnings histories since job market entry
  - Not very effective in practice
  - Imputed average earnings do not exhibit the relationship with pension benefits documented by many others → Very noisy imputation procedure
- **Go-around #2.** ELSA data + data simulated by O'Dea 2018
  - O'Dea 2018 provides a good fit of earnings profiles (admin data)
  - Implicit assumption: our cohort (1950–57) is similar to his (1935–50)

Imputation of Average Earnings with O'Dea Data

1. Obtain parameters $\{\hat{\beta}_0^j, \hat{\beta}_1^j, \hat{\beta}_2^j\}_{j=1}^4$, where j indexes household type, from OLS regression in data simulated by O'Dea 2018:

$$
\begin{aligned}
ae_{i,64}^j =& \beta_0^j \left(1 - 1\{\text{pbb}_{i,65}^j \geq 29.13k\}\right) \text{pbb}_{i,65}^j + \beta_1^j 1\{\text{pbb}_{i,65}^j \geq 29.13k\} \\
&+ \beta_2^j \left(1 - 1\{\text{pbb}_{i,65}^j \geq 29.13k\}\right) \text{privben}_{i,65}^j + \varepsilon_i^j . \qquad (1)
\end{aligned}
$$

   (Reference group is those receiving more than 29.13k GBP in pbb; for others, avg. earnings are linear in pbb and $privben$ at 65)

2. Use $\{\hat{\beta}_0^j, \hat{\beta}_1^j, \hat{\beta}_2^j\}_{j=1}^4$ and similar household classification to generate $ae_{i,64}$ for individuals in our sample according to (1) without ε_i^j term

3. Recover average earnings at age 50 from:

$$
ae_{i,64}^j = \frac{\text{Emp years}_{i,50}^j \cdot ae_{i,50}^j + \text{Earnings ELSA}_i^j}{\text{Emp years}_{i,50}^j + \text{ELSA empl years}_i^j}
$$

Back

Initial Distribution of States

1. **Ignoring ME**: Obtain initial distribution of states directly from data (assets, average earnings, health), simulating initial wage-shocks according to estimated initial distribution of wage shocks

2. **Acknowledging ME**:
   - Use previously-estimated initial probability distribution of true health
   - Estimate joint distribution of assets & avg. earnings given true health
     - Assume $(\log a_0, \log ae_0)$ are jointly log-normal given true health

$$\begin{pmatrix} \log a_0 \\ \log ae_0 \end{pmatrix} \sim \mathcal{N}\left(\mu_H, \Sigma_H\right)$$

   - Use previously-estimated initial distribution of wage shocks (assuming these are independent of the rest of states)

Back to Estimation Back to Model Fit

Data Profiles

- We pool individuals from different birth cohorts and assume that differences in profiles across cohorts driven solely by cohort effects
  - This responds to data limitations
- Classify individuals in:
  - 19 two-year age groups (from 50–51 to 86–87)
  - 4 cohort groups (born <1935, born 1935–1943, born 1943–1950, born >1950)
  - 2 health groups (healthy and unhealthy)
- Run regression

$$y_{i,w} = \gamma_0^{y,m} + \eta_a^{y,m} + \eta_c^{y,m} + \gamma_a^{y,m}\left(a_{iw} \times 1_{\{\text{Health}_i^m = \text{Good}\}}\right) + u_{i,w}^{y,m}$$

y: targeted variable $\quad$ i: individual $\quad$ w: wave $\quad$ m: health indicator $\quad$ a: age $\quad$ c: cohort

to obtain estimates $\left\{\hat{\gamma}_0^{y,m}, \{\hat{\eta}_a^{y,m}, \hat{\gamma}_a^{y,m}\}_{a=1}^{19}, \{\hat{\eta}_c^{y,m}\}_{c=1}^{4}\right\}$ for each y

Data Profiles

- Use estimates $\{\hat{\gamma}_0^{y,m}, \{\hat{\eta}_a^{y,m}, \hat{\gamma}_a^{y,m}\}_{a=1}^{19}, \{\hat{\eta}_c^{y,m}\}_{c=1}^{4}\}$ for each (y, m)-pair to generate y profiles by health status according to:

$$
\begin{aligned}
y_a^{\text{good health}(m)} &= \hat{\gamma}_0^{y,m} + \hat{\eta}_a^{y,m} + \hat{\eta}_{c=4}^{y,m} + \hat{\gamma}_a^{y,m} a, \qquad && a \in \{1, \ldots, 19\}, \\
y_a^{\text{bad health}(m)} &= \hat{\gamma}_0^{y,m} + \hat{\eta}_a^{y,m} + \hat{\eta}_{c=4}^{y,m}, && a \in \{1, \ldots, 19\}.
\end{aligned}
$$

y: targeted variable a: age m: health indicator $c = 4$ (cohort born in 1950–57)

- The vectors

$$
\begin{aligned}
\mathbf{y}^{\text{good health}(m)} &= \left(y_1^{\text{good health}(m)}, \ldots, y_{19}^{\text{good health}(m)}\right), \\
\mathbf{y}^{\text{bad health}(m)} &= \left(y_1^{\text{bad health}(m)}, \ldots, y_{19}^{\text{bad health}(m)}\right),
\end{aligned}
$$

give the y profiles for our cohort where $m \in \{\text{SRHS}, \text{DLs}\}$.

Back