AI, Machine Learning and Macroeconomic Aspects of Health, Health Policies

Zhigang Feng

University of Nebraska

November 10, 2022
Open Classroom–AI Series–Brown School, WUSTL
Outline

▶ What are the macroeconomic impacts of health and health policies in the US?

▶ How to build a macro model to answer some critical policy relevant questions?

▶ What additional insights AI and Machine learning can discover?
U.S. Healthcare System Facts

**Fact 1**: The United States is the only developed country without a system of universal healthcare.
- 28.9 (46.5) million American uninsured in 2019 (2010).

**Fact 2**: On average, American spends more than 18% of income on health.

**Fact 3**: US health insurance is employment-based (**EHI**). Over 90% of insured working-age people are covered by EHI.

**Fact 4**: EHI premium is deductible from employee’s taxable income.
- This exclusion reduced federal tax revenue by $268 billion in 2011.
- Because it reduces taxable income, the exclusion benefits taxpayers in higher tax brackets more than those in lower tax brackets.
Macro-aspects of Health and Health Policies

- A research agenda that studies the macro-aspect of health and health policies.
  - Labor supply: Feng ’09, Feng and Zhao ’18,
  - Entrepreneurship: Chivers, Feng and Villamil ’17, Feng and Villamil ’21
  - Job creation: Feng and Bernhart ’22
  - Inequality: Chen, Feng and Gu ’22

- Growing attention on income inequality and health disparity
  - Chetty et al. ’16: strong correlation between income and health.

- Literature well documents that health affects income inequality:
  - The unhealthy have lower income and higher medical expenditure.
Fact: Unhealthy individuals have lower income and higher medical expenditures.
Health and Inequality

Growing attention on income inequality and health disparity

Literature well documents that health affects income:

- The unhealthy have lower income and higher medical expenditure.

What determines health? What are implications on inequality?

- Income affects health through health insurance choice.

\[
\text{Income} \xrightarrow{\text{(1) ?}} \text{Insurance Coverage} \xrightarrow{\text{(2) ?}} \text{Health}
\]
Why Health Insurance

- The evidence to date suggests that health insurance is an element in the health production function, c.f. Finkelstein et.al. ’19.

- Insurance choice is observable and thereby amenable to parameterization and estimation.

- The focus on health insurance choice reflects the well-documented health benefit of insurance coverage.
  - preventive care
  - health-care services once sick
  - wellness program

- Government’s health care policies, such as the Affordable Care Act, are generally designed to target health insurance coverage.
Insurance Coverage by Income

Income → \( (1) \) EHI, affordability, etc → Insurance Coverage \( (2) \) → Health

Graph showing the relationship between income and insurance coverage, divided into total insured, public, private, and EHI categories.
## Health Insurance and Health

Income → Insurance Coverage → Health

1. EHI, affordability, etc
2. Preventive care, etc.

<table>
<thead>
<tr>
<th>Check-up Items</th>
<th>%, Insured</th>
<th>%, Uninsured</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within last year:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall health assessment</td>
<td>70.6</td>
<td>36.7</td>
</tr>
<tr>
<td>Dental</td>
<td>61.9</td>
<td>31.7</td>
</tr>
<tr>
<td>Blood cholesterol</td>
<td>68.1</td>
<td>34.3</td>
</tr>
<tr>
<td>Flu shot</td>
<td>41.2</td>
<td>15.1</td>
</tr>
<tr>
<td>Prostate specific antigen (male only)</td>
<td>42.7</td>
<td>12.7</td>
</tr>
<tr>
<td>Pap smear test (female only)</td>
<td>54.8</td>
<td>41.1</td>
</tr>
<tr>
<td>Breast exam (female only)</td>
<td>63.9</td>
<td>40.8</td>
</tr>
<tr>
<td>Mammogram (female only)</td>
<td>47.2</td>
<td>23.7</td>
</tr>
<tr>
<td><strong>Ever had:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood stool test</td>
<td>20.7</td>
<td>6.8</td>
</tr>
<tr>
<td>Sigmoidoscopy or colonoscopy</td>
<td>29.0</td>
<td>7.3</td>
</tr>
</tbody>
</table>
Health Disparity and Income Inequality

To study the interaction between health disparity and income inequality.

- build a heterogeneous-agent framework
  - featuring endogenous health through insurance choice
  - accounting for the joint distribution of health, income and the observed pattern of health insurance choice

- empirically identify the causal effects of health insurance on health
  - IV approach to address the endogeneity issue of insurance
  - assess parametric, semi-parametric, and partial identification frameworks

- use the model to quantify
  - the persistent effect of insurance on health disparity and income inequality over the life cycle
  - macroeconomic implications of providing universal health insurance coverage
Insured individuals are more likely to stay healthy or recover from unhealthy status.
Model Overview

- OLG + idiosyncratic shocks + incomplete-markets.

- Income process: varies with education and health.

- Health process: endogenously depending on insurance coverage.
  - Health expenditure shock: depends on age, health, and insurance coverage.
  - Endogenous mortality: depends on age and health.

- Market arrangements:
  - Financial market: one-period, risk free bond
  - Health insurance: EHI, private insurance, Medicaid, Medicare
  - Allow medical bankruptcy as an implicit insurance

- Government collect taxes to finance Medicare, medical bankruptcy, etc.

- Production side: a representative firm.

- Health and income distributions are jointly determined in equilibrium.
Workers’ Problem

- State variables:  $s_w = (j, h, a, m, i^{hi}, i_E, z, \iota)$

- The problem of workers:

\[
V^j(s_w) = \max_{c, n, a', i_{hi}', \iota'} \left\{ u_h(c, n) + \beta p h, j \pi E V^{j+1}(s_w') \right\}
\]

subject to

\[
(1 + \tau_c)c + a' + i_{hi}' \tilde{\pi} + (1 - i_{hi}\phi(m))m = e + (1 + r)a + \hat{w}n - \text{Tax} + \text{TR}, \text{ if } \iota = 0,
\]
\[
(1 + \tau_c)c + i_{hi}' \tilde{\pi} = e - \text{Tax} + \text{TR} - \lambda \text{ and } a' = 0, \text{ if } \iota = 1.
\]

where

\[
\tilde{\pi} = \begin{cases} 
\pi_E(1 - \psi), & \text{if } i_{hi} = 1, i_E = 1, \text{ and } e > \Theta_e \text{ or } a > \Theta_a; \\
\pi_P(h), & \text{if } i_{hi} = 1, i_E = 0, \text{ and } e > \Theta_e \text{ or } a > \Theta_a; \\
0, & \text{if } i_{hi} = 0 \text{ or } e \leq \Theta_e \text{ and } a \leq \Theta_a.
\end{cases}
\]
Retirees’ Problem

- State variables: \( s_r = (j, h, a, m) \)
- The problem of retirees:

\[
V^j(s_r) = \max_{c, a', \iota'} \left\{ u(c) + \beta \rho_{h,j} \mathbb{E} V^{j+1}(s'_r) \right\}
\]

subject to

\[
(1 + \tau_c) c + a' + \pi_{mr} + (1 - \phi_{mr}(m)) m = ss + (1 + r) a - \text{Tax} + \text{TR}, \text{ if } \iota = 0, \]
\[
(1 + \tau_c) c + \pi_{mr} = ss - \text{Tax} + \text{TR} - \lambda \text{ and } a' = 0, \text{ if } \iota = 1.
\]
Equilibrium Conditions

▶ Labor market:

\[ N = \int z(s)g(h(s))n(s)d\Phi(s) \]

▶ Capital market:

\[ K = \int a(s)d\Phi(s) \]

▶ Good market:

\[ C + K' - (1 - \delta)K + M + G = Y \]

▶ Government budget:

\[ G + \int_s [ss + \phi_{mr}(m(s))m(s) - \tau_{mr}]d\Phi(s) + \int_s TR(s)d\Phi(s) \]

\[ + \int_s \nu(s)[(1 - \phi(m(s)))m(s) - \lambda(s)]d\Phi(s) + \int_s \phi_{md}(m(s))m(s)\mathbb{1}_{e \leq \Theta_e, a \leq \Theta_a}d\Phi(s) \]

\[ = \int_s [\tau_{c}c(s) + T(y(s))]d\Phi(s) + \int_s [\tau_{mr}(y(s)) + \tau_{ss}(y(s))]d\Phi(s) \]

▶ Stationary: \( \mathbb{E}[\varphi(s)] = \varphi(s) \)
Equilibrium Conditions

- The F.O.C. with respect to individual’s (private) health insurance choice is given as below.

\[
\frac{uc}{1 + \tau_c} \frac{\partial c}{\partial hi} \tilde{\pi} + \beta \frac{\rho E u_c}{1 + \tau_c} \frac{\partial \tilde{\pi} + \partial h_+}{\partial hi} = \frac{uc}{1 + \tau_c} \frac{\partial c}{\partial Tax} \frac{\partial Tax}{\partial hi} + \beta \left\{ \frac{\partial \rho}{\partial h_+} \frac{\partial h_+}{\partial hi} E u_+ \right\} + \frac{\rho E u_c + \partial y_+ + \partial h_+}{1 + \tau_c} \frac{\partial h_+}{\partial hi}
\]

where

\[
\frac{\partial y_+}{\partial h_+} = \mathbf{1}_{(TR \leq 0)} \left[ w_+ z_+ \left( \frac{\partial g(h_+)}{\partial h_+} n_+ + g(h_+) \frac{\partial n_+}{\partial h_+} \right) - (1 - \phi) \frac{\partial m_+}{\partial h_+} \right]
\]

\[
\frac{\partial \tilde{\pi} + \partial h_+}{\partial h_+} = \mathbf{1}_{(e_+ > \Theta_e \text{ or } a_+ > \Theta_a)} \frac{\partial \pi_+}{\partial h_+} \frac{\partial y_+}{\partial h_+}
\]
Tradeoffs of Insurance Purchase

- The marginal cost of purchasing (private) health insurance:
  - lost consumption due to premium: \( \frac{u_c}{1 + \tau_c} \frac{\partial c}{\partial h_i} \pi \);
  - forgone opportunity in qualifying for Medicaid: \( \beta \frac{\rho E u_{c+} \frac{\partial \pi^+}{\partial h_+} \frac{\partial h_+}{\partial h_i}}{1 + \tau_c} \);

- The marginal benefit of purchasing (private) health insurance:
  - income tax deduction for insurance premium: \( \frac{u_c}{1 + \tau_c} \frac{\partial c}{\partial \text{Tax}} \frac{\partial \text{Tax}}{\partial h_i} \);
  - extra year of quality life: \( \beta \frac{\partial \rho}{\partial h_+} \frac{\partial h_+}{\partial h_i} E u_+ \);
  - rising future income: \( \frac{\rho E u_{c+} \frac{\partial y_+}{\partial h_+} \frac{\partial h_+}{\partial h_i}}{1 + \tau_c} \);
Frictions of Insurance Policy

Who are more likely to stay uninsured:

- lower income: with high $u_c$, low $\frac{\partial y_+}{\partial h_+}$;
- young and healthy: low $\frac{\partial \rho}{\partial h_+} \frac{\partial h_+}{\partial h_i}$.
- with access to implicit health insurance: $1_{(TR \leq 0)}$

There are two distortions against the health insurance choices.

- social insurance puts a lower bound on MC of being unhealthy:
  $\frac{\partial y_+}{\partial h_+} = 1_{(TR \leq 0)} \left[ w_+ z_+ \left( \frac{\partial g(h_+)}{\partial h_+} n_+ + g(h_+) \frac{\partial n_+}{\partial h_+} \right) - (1 - \phi) \frac{\partial m_+}{\partial h_+} \right]$;

- Medicaid affects the effective cost of insurance purchase:
  $\frac{\partial \tilde{\pi}_+}{\partial h_+} = 1_{(e_+ > \Theta_e \text{ or } a_+ > \Theta_a)} \frac{\partial \pi_+}{\partial y_+} \frac{\partial y_+}{\partial h_+}$.
Modern macroeconomic theory rests on the idea of Rational Expectation:

- **Representative agents:**
  - Characterized by: \( V(k, K), \ g(k, K), \ K_+ = G(K), \) and \( k = K \).
  - Fixed point, contraction mapping, monotonic operators.
  - Projection, perturbation, etc.

- **Heterogenous agents:**
  - No aggregate uncertainty and existence of stationary distribution.
    - a fixed point for prices: \( V(k, \epsilon; r(\Gamma)), \ g(k, \epsilon; r(\Gamma)), \) and \( \Gamma_+ = \Gamma = \Gamma^* \)
  - with aggregate uncertainty
    - distribution enters the state space: \( V(k, \epsilon, Z; r(\Gamma)), \ g(k, \epsilon, Z; r(\Gamma)), \) and \( \Gamma_+ = \mathcal{H}(\Gamma, Z) \)
    - the evolution of distribution \( \mathcal{H} \) summarizes each individual’s \( i \) behavior \( g^i \),
    - while individual’s behavior rests on: their expectations on other’s actions \( g^{-i} \), and hence the distribution and its evolution.
 Curse of Dimensionality (CoD) # 1: dimension of state space
  - $k \rightarrow \{k, a, h, \ldots\}$
  - $\Gamma$ is an infinite dimensional object

 Curse of Dimensionality (CoD) # 2: unfolding of uncertainty
  - $\mathbb{E}$ over future contingencies.
  - the space for $\Gamma$ is unknown
Machine Learning and Macroeconomics

- Curse of Dimensionality (CoD) #1: dimension of state space
  - Traditional methods
    - sparse grids, local approximation, etc.
    - construct grids to generate informations and to solve parameters
  - Machine learning
    - neural network, grid free
    - recover the equilibrium functions via Monte Carlo sampling
Figure 1: Grids $G(3, 2)$, $G(4, 2)$, $G(5, 2)$, and $G(6, 2)$.
Machine Learning and Macroeconomics

- Curse of Dimensionality (CoD) # 2: unfolding of uncertainty
  - Traditional methods
    - discretization and reduction of shock process
    - impose Markovian
  - Machine learning
    - re-produce the ergodic state via direct simulation Monte Carlo
    - recover the equilibrium functions based on the ergodic state
Machine Learning and Macroeconomics

```
<table>
<thead>
<tr>
<th>t=0</th>
<th>t=1</th>
<th>t=2</th>
<th>t=3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.1,1,1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.1,1,0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.1,0,1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.1,0,0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0,1,1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0,1,0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0,0,1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0,0,0)</td>
</tr>
</tbody>
</table>
```
The equilibrium conditions for the canonical optimal growth model can be characterized by the generalized Euler equation.

\[ u'(c) = \beta \cdot \left[ u'(c') \cdot (1 - \delta_k + f'(k')) \right], \tag{1} \]

where \( \frac{\partial K(k')}{\partial k'} \) is the derivative of saving function.
In order to recast it as a machine learning problem, we will define a slightly modified one as follows:

$$\min_{g^k} \mathcal{L} \left( g^k \right) \equiv \frac{1}{N} \sum_{i=0}^{N} \left\{ \left[ u'(c_i) - \beta \cdot u'(c'_i) \cdot (1 - \delta_k + f'(k'_i)) \right] - 0 \right\}^2,$$

where \( i \) indexes different initial condition \( k_i \) coming from the stationary distribution, \( \mathcal{L} \left( g^k \right) \) is the loss function and \( g^k \) is the unknown (to-be-determined) equilibrium policy function. This boils down to find \( g^k \) such that

$$\frac{d\mathcal{L} \left( g^k \right)}{dg^k} = 0,$$

where the derivative is with respect to functional equation and hence it is Frechet derivative.
Optimal Growth Model--Euler Equation

The function $g^k(\cdot)$ will be approximated by $2N$ layer neural network

$$k \rightarrow \Gamma_\gamma(k) = \sigma_N (W_N \cdot [\ldots \sigma_2 (W_2 \cdot \sigma_1 (W_1 \cdot k + b_1) + b_2) \ldots + b_N])$$

(4)

where $W_n$ are weight matrixes and $b_n$ are bias vectors. They are used to do linear adjustments. While $\sigma_n$ are nonlinear activation functions.
Optimal Growth Model--Euler Equation

Next, we will use $\Gamma_\gamma$ to approximate $g^k$ and find the optimal $\Gamma_\gamma$ to minimize the loss function via gradient descent.

$$\min_{\Gamma_\gamma} \mathcal{L}(\Gamma_\gamma).$$  \hspace{1cm} (5)

Universal Approximation Theorem: Whenever the activation function is bounded and non-constant, a standard multilayer feedforward network can approximate any (Borel) measurable function arbitrarily well, given that sufficiently many hidden units are considered.

It implies that we can find $\Gamma_\gamma$ by finding the optimal $\gamma^*$ minimizing the loss function, and the corresponding $\Gamma^*_\gamma$ will uniformly approximate $g^k$.

We find $\gamma^*$ via an iterative procedure. Given $\gamma^{(j)}$, $\gamma^{(j+1)}$ will be updated as follows:

$$\gamma^{(j+1)} = \gamma^{(j)} - \alpha^{\text{learning}} \frac{d\mathcal{L}(\Gamma_\gamma^{(j)})}{d\gamma},$$  \hspace{1cm} (6)

where $\alpha^{\text{learning}}$ is the learning rate.
Gradient descent

▶ Stochastic batch gradient descent (BGD) method is a popular approach to reduce the cost of computing \( \frac{d\mathcal{L}(\Gamma_{\gamma(j)})}{d\gamma} \)

\[
\frac{d\mathcal{L}(\Gamma_{\gamma(j)})}{d\gamma} \approx \frac{1}{n} \sum_{i=1}^{n} \frac{d\mathcal{L}(\Gamma_{\gamma(j)}; b_i)}{d\gamma}
\]

▶ \( n = 1 \): stochastic gradient (SGD) method which approximates the expectation function with the value of such a function in one randomly chosen data point;

▶ \( n = N \): the conventional gradient descent (GD) method in which all data points are used for constructing the gradient.
Gradient descent: algorithm

- Make an initial guess on the parameters vector
- For $n = 1, 2, \ldots$, do the following:
  - draw a random realization for $b_n$
  - compute a stochastic vector of gradients $g(b_n; \gamma_n)$
  - choose a learning rate $\lambda_n > 0$
  - compute the new parameters vector as $\theta_{n+1} \leftarrow \theta_n - \lambda_n g(b_n; \gamma_n)$
- End iterations when convergence is achieved.
Gradient descent: convergence

- Assumption 1: the objective function $F$ is continuously differentiable; the gradient of $F$ is Lipschitz continuous with a constant $L > 0$.

- Assumption 2: the objective function $F$ and the algorithm satisfies the following three properties:
  - The sequence $\theta_n$ is in an open set where the objective function is bounded from below by a scalar $F_{\text{inf}}$ for all $n$;
  - There exists scalars $\mu_G$ and $\mu$ such that $\mu_G \geq \mu > 0$ and for all $n$, we have
    \[
    \nabla F(\theta_n) \mathbb{E}_{b_n} [g(b_n; \theta_n)] \geq \mu \|\nabla F(\theta_n)\|^2
    \]
    \[
    \|\mathbb{E}_{b_n} [g(b_n; \theta_n)]\| \leq \mu_G \|\nabla F(\theta_n)\|
    \]
  - There exists scalars $H \geq 0$ and $H_V \geq 0$ such that for all $n$, we have
    \[
    \mathbb{E}_{b_n} \left[\|g(b_n; \theta_n)\|^2\right] - \|\mathbb{E}_{b_n} [g(b_n; \theta_n)]\|^2 \leq H + H_V \|\nabla F(\theta_n)\|^2.
    \]
Gradient descent: convergence

Theorem (Nonconvex objective and a diminishing learning rate):
Suppose Assumptions 1 and 2 hold. Assume that a sequence \( \{ \lambda_n \} \) satisfies

\[
\Delta_n \equiv \sum_{n=1}^{N} \lambda_n = \infty, \sum_{n=1}^{N} \lambda_n^2 < \infty
\]

then the expected sum of squares of the gradients of \( F(\theta_n) \), weighted by \( \lambda_n \), satisfies

\[
\mathbb{E} \left[ \sum_{n=1}^{N} \lambda_n \| F(\theta_n) \|^2 \right] < \infty
\]

and the expected average squared gradients of \( F(\theta_n) \), weighted by \( \lambda_n \), satisfies

\[
\mathbb{E} \left[ \frac{1}{\Delta_n} \sum_{n=1}^{N} \lambda_n \| F(\theta_n) \|^2 N \to \infty 0 \right].
\]

Optimal Growth Model

![Graph showing optimal growth model with two curves and loss vs iteration plots.](image-url)
Implementation: Pseudo Code

1. define parameters;
2. define Neural Network;
3. simulate the economy and generate the stationary distribution (ergodic set);
4. define the objective functions:
   - Euler equation, market clearing;
   - consistency between perceived and the actual equilibrium functions.
5. training the NN by minimizing the loss function.
   - generate data based on the ergodic set from step 3: initial condition and equilibrium functions;
   - feed with minibatchs (within one Epoch) via Dataloader, and train the system to update the NN;
   - repeat the process for many Episodes;
# construction of neural network
n1, n2 = 50, 50
decision_rule = nn.Sequential(nn.Linear(1, n1),
    nn.ReLU(),
    nn.Linear(n1, n2),
    nn.ReLU(),
    nn.Linear(n2, 1),
    nn.Sigmoid())
decision_rule = decision_rule.to(device)
Implementation: Dataloader

# define dataloader:
class Set_data(Dataset):
    def __init__(self, Dataset):
        self.data = Dataset
    def __len__(self) -> int:
        return len(self.data[:, 0]) # find the size of row
    def __getitem__(self, idx: int):
        return self.data[idx, :] # draw samples by row

# load dataset:
ds = Set_data(x_sim_data)
dl = torch.utils.data.DataLoader(ds, batch_size=x_batch_size,
shuffle=True)
for batch_idx, sample in enumerate(dl):
    x_data_minibatch = sample
Implementations: Training of the NN

```python
losses_p_func = []
optimizer_foc = torch.optim.Adam(decision_rule.parameters(), lr=v_lr)
x_loss_p, x_p1_nn = obj_p_func(decision_rule, x_x0_p, x_p1_fit)
optimizer_foc.zero_grad()
x_loss_p.backward,retain_graph=True)
optimizer_foc.step()
x_decision_p.state_dict()
losses_p_func.append(x_loss_p.detach().cpu().numpy())
```
def obj_p_func(x_decision_p, x_x0_p, x_p1_fit):
    x_x0_p = torch.tensor(np.array(x_x0_p), device=device)
    x_p1_fit = torch.tensor(np.array(x_p1_fit), device=device)
    x_p1_nn = torch.sigmoid(x_decision_p(x_x0_p)) * (p_max - p_min) + p_min
    loss_function = torch.nn.MSELoss(reduction='mean')
    loss_p_func = loss_function(x_p1_nn, x_p1_fit)
    return loss_p_func, x_p1_nn.detach().cpu().numpy()
Determinants of Adulthood Health and Income Distribution

Q: How do initial conditions affect adulthood health and income status at later stage of life?

- Simulate a long panel data using equilibrium decision rules from the model

- Estimates using the simulated data should be comparable to real data:
  - the baseline economy replicates the salient features of the real data
  - carefully choose the initial distribution at 25 when agents enter the economy

- Advantages of using the simulated data:
  - much longer in time dimension compared to real data without attrition bias
  - easily control for unobserved heterogeneity in the simulated data
Persistent Effects of Initial Status

Consider the following regression for income at $j \geq 2$

$$h_j^i = \gamma_h h_1^i + \gamma_e e_1^i + \gamma h i_1^i h_i + \sum_k \psi_k 1[(\alpha^i, \sigma^i) = (\alpha_k, \sigma_k)] + u_j^i$$

Initial insurance status has persistent effects on lifetime health and income:

- Initial insurance: health $\uparrow \Rightarrow$ income $\uparrow \Rightarrow$ insurance coverage $\uparrow \Rightarrow$ health $\uparrow$.
- Controlling for initial income and health, insurance affects income even after 10 years.
Persistent Effects of Initial Status

Long-Run Effect of Insurance on Health

Long-Run Effect of Insurance on Income
The Value of Modeling Health Premium

The graphs show the long-run effect of insurance on health and income over years after $t_0$. The graphs are labeled with "Baseline Model" and "No Health Premium."
Universal Health Care

- Universal health care: directly intervene the health market
  - financed through increasing consumption tax.

- Macroeconomic impact of universal health care:
  - level effects on aggregate demographic variables and human capital supply
  - improving health, in addition to hedging medical exp. shocks
  - countervailing effects on income inequality
Universal Health Care: Level Effects

<table>
<thead>
<tr>
<th></th>
<th>Universal Health Coverage (% Change Compared to Baseline)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working-Age Population</td>
<td>+0.8</td>
</tr>
<tr>
<td>Healthy Individuals</td>
<td>+3.8</td>
</tr>
<tr>
<td>Human Capital (Eff. Units of Labor)</td>
<td></td>
</tr>
<tr>
<td>Aggregate</td>
<td>+7.1</td>
</tr>
<tr>
<td>Average</td>
<td>+6.2</td>
</tr>
</tbody>
</table>
# Universal Health Care: Distributional Effects

<table>
<thead>
<tr>
<th>Percentage of Healthy Individuals (% Change Compared to Baseline)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>By Income Quintiles:</strong></td>
</tr>
<tr>
<td>Q1</td>
</tr>
<tr>
<td>Q2</td>
</tr>
<tr>
<td>Q3</td>
</tr>
<tr>
<td>Q4</td>
</tr>
<tr>
<td>Q5</td>
</tr>
<tr>
<td><strong>By Age Groups:</strong></td>
</tr>
<tr>
<td>25–34</td>
</tr>
<tr>
<td>35–44</td>
</tr>
<tr>
<td>45–54</td>
</tr>
<tr>
<td>55–65</td>
</tr>
</tbody>
</table>
Table: Income Inequality

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>UHC</th>
<th>UHC (mortality adj.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 10</td>
<td>33.6</td>
<td>33.2</td>
<td>31.7</td>
</tr>
<tr>
<td>Top 25</td>
<td>58.5</td>
<td>58.2</td>
<td>56.0</td>
</tr>
<tr>
<td>Bottom 50</td>
<td>18.0</td>
<td>18.3</td>
<td>20.7</td>
</tr>
</tbody>
</table>

The difference in income share of the top 10 percent between the U.S. and Canada, where a UHC is in place, is around 5.5 percent. In this regard, the UHC implementation alone explains around one-third of Canada-US difference in inequality.
Additional Insights on Universal Health Care

- Universal health care resolves adverse selection in the insurance market.

- With endogenous mortality, UHC distorts the inter-temporal choice:
  - improves health for the poor and therefore survival rate and life expectancy
  - effectively changes the discount factor for the poor
  - their savings may not be enough for consumption of the longer lifetime

- The poor might be better off if they were offered with a combination of insurance and monetary compensation.