On The Rise of Health Spending and Longevity

Raquel Fonseca
Pierre-Carl Michaud
Titus Galama
Arie Kapteyn

December 2009

I Z A

On The Rise of Health Spending
and Longevity

Raquel Fonseca
RAND

Pierre-Carl Michaud
RAND, Netspar and IZA

Titus Galama
RAND

Arie Kapteyn
RAND, Netspar and IZA

ABSTRACT

On The Rise of Health Spending and Longevity[*]

We use a calibrated stochastic life-cycle model of endogenous health spending, asset accumulation and retirement to investigate the causes behind the increase in health spending and life expectancy over the period 1965-2005. We estimate that technological change along with the increase in the generosity of health insurance may explain independently 53% of the rise in health spending (insurance 29% and technology 24%) while income less than 10%. By simultaneously occurring over this period, these changes may have lead to a "synergy" or interaction effect which helps explain an additional 37% increase in health spending. We estimate that technological change, taking the form of increased productivity at an annual rate of 1.8%, explains 59% of the rise in life expectancy at age 50 over this period while insurance and income explain less than 10%.

Corresponding author:

Raquel Fonseca
RAND Corporation
1776 Main Street
Santa Monica, CA 90407-2138
USA
E-mail: fonseca@rand.org

1. Introduction

According to a recent report from the Congressional Budget Office (CBO, 2008), per capita health spending in 2005 was 6 times what it was in 1965. As a fraction of per capita income, health spending in the U.S. has grown from 4% to 16%. Newhouse (1992) concludes that more than half, perhaps even 65%, of the change is probably attributable to technological change, with more generous health insurance and the growth in income accounting for at most one third of the rise in health spending.

Over the same period, life expectancy has increased considerably, particularly among the elderly. In 2005, a 50 year old men could expect to live 28.9 years compared to 22.95 in 1965. Cutler, Deaton and Lleras-Muney (2006) argue that technological change is the leading explanation for the increase in longevity since evidence from the Health Insurance Experiment (HIE, Manning et al., 1987) and other recent studies on the SES-health gradient (Adams et al., 2003; Smith, 2007) suggest that neither growth in income nor the spread of health insurance can explain changes in life expectancy of that magnitude.

In a recent study, Hall and Jones (2007) show that one can generate the growth in health spending and the resulting change in life expectancy solely as a result of growth in income using a model of endogenous health spending (Grossman, 1972; Ehlrich and Chuma, 1990). Instead of modeling other changes simultaneously, Hall and Jones show that with reasonable parameter estimates and implied value of longevity gains, one can generate the rise in health spending and life expectancy, assuming it is an optimal response to the growth in income.

We follow a similar modeling approach, generating health spending growth as an optimal response to changing circumstances. We introduce in our framework health insurance and technological change as explicit possible competing alternatives to income growth. We calibrate a stochastic life-cycle model of endogenous health spending, asset accumulation and retirement with detailed modeling of health insurance, Social Security and taxation using micro-data from 2000 to 2005. Our calibration approach involves fitting life-cycle profiles of assets, medical expenditures, retirement and mortality. We also calibrate the parameters of the model such that behavioral responses to changes in health insurance and income are similar to those reported in the literature. We then perform counterfactual simulations where we first find the necessary productivity gains such that we reach 1965 health spending levels. We find that a 1.8% annual growth in productivity is necessary to achieve the 1965 health spending levels. Once we obtain simulated profiles for 1965, we introduce sequentially each of the competing explanations for the rise in health spending to understand their relative importance. We find that far from competing, these are reinforcing forces: as much as 37% of the overall increase in health spending is due to the fact that these occurred simultaneously. We get a somewhat different story when we study the factors jointly regarding the relative importance of each factor in independently explaining the rise in health spending. We find that the increase in health insurance copayment has been as important as technological change while growth in real income over this period explains at most 10% of the rise in spending. Moreover, we find that both income and health insurance copayments can only explain up to 10% of the rise in longevity while productivity gains from technology account for up to 60% of the gain in life expectancy.

The relative importance of income in our study differs from Hall and Jones (2006) for two important reasons. First, Hall and Jones (2006) use real GNP per capita growth while we use real earnings growth, the first one being much larger over this period. For workers in our sample, most of the growth in real GNP was not reflected in earnings growth. The average worker saw a growth in earnings of roughly 25% over the period we cover compared to growth in GNP per capita nearing 200% between 1965 and 2005. Second, we obtain a lower income elasticity of medical care. Hall and Jones need an income elasticity of medical care close to 4 in order to explain, solely using income growth, that medical expenditures rose by 800% while GNP per capita rose by roughly 200% between 1950 and 2000. Our estimate of the income elasticity (to a permanent change in income) is close to unity which is in line with the cross-national estimates from the literature and larger than those from the micro-literature (between 0.2 and 0.4) (see Newhouse, 1992). Hence, in our model income does not explain the rise in the *share* of health spending. Insurance and technology do, and income facilitates the spillover effect.

In section 2 we present the model. Section 3 presents the data and the calibration. In section 4 we perform the counterfactual simulations. We conclude in section 5.

2. Model

We build a model where an agent makes decisions over the entire life-cycle regarding health spending, consumption and retirement. The agent faces health risk, job risk and insurance coverage risk (up to the age of Medicare eligibility). The framework is closest to the one considered by French and Jones (2007) with three important differences. First, we endogenize health spending. Health spending can affect health which in turn affects directly future health, utility and earnings. In that sense, our modeling of health

4

spending is closest to the human capital framework of Grossman (1972).[2] Second, an important difference is that we do not allow for uncertainty in earnings and medical expenditures given health, job and insurance status. Once medical expenditures are endogenous, there is little reason to think of the realization of health spending as stochastic conditional on a given health insurance, job and health status. In fact, they are assumed stochastic, *ex ante*, as the individual does not know yet his health, job and health insurance status. Finally, another important difference is that we assume labor supply is inelastic up before the agent chooses the age of retirement and that retirement is an absorbing state. We do this because labor supply plays a secondary role in our analysis.

2.1. Decisions and Preferences

The agent makes decisions from age $t = t_0, ..., T$ and dies with certainty at age. In other words, the terminal age T is the limit on human life. At each age *t*, we assume the agent makes three decisions $d_t = (c_t, m_t, r_t)$. She chooses non-medical expenditures c_t, medical expenditures m_t and her retirement status r_t. If she retires, let $r_t = 1$ and $r_t = 0$ if not. For our purposes, we will assume retirement is an absorbing state and is possible after age 54. At age 70, we assume everyone is retired (few in the data work beyond that age). At age *t*, she has a health state h_t assumed to take 6 values $\{0,1,2,3,4,5\}$ where zero denotes death and values 1 to 5 denote the self-reported health status scale {poor, fair, good, very good, excellent}. We assume preferences are represented by the following utility function

---

[2] Other studies investigating endogeneous health investment include Blau and Gilleskie (2008), Galama et al. (2008), Khwaja (2009), Yogo (2009), Bajari, Hong and Khwaja (2009), Halliday, He and Zhang (2009), Juergen and Tran (2007), Suen (2009)

$$u(c_t, m_t, r_t) = \frac{c_t^{1-\sigma}}{1-\sigma} + \psi_h h_{t,>2} + e^{\psi_0 + \psi_t t} r_t. \tag{1}$$

The parameter σ captures the degree of relative risk aversion, ψ_h measures the direct

utility of good health where $h_{t,>2} = I(h_t > 2).$[3] Finally ψ_0, ψ_t control the direct utility of

leisure in retirement, which may vary with age.

We assume she has a bequest motive. When she dies at age t, she leaves a bequest

in terms of accumulated assets, a_t, which give utility

$$b(a_t) = \phi_b \log(a_t). \tag{2}$$

We assume that the bequest function is logarithmic and the parameter ϕ_b measures the

strength of the bequest motive.

2.2. Assets

At each age, the agent can decide to forgo consumption and invest instead in a composite

asset a_t with annual return τ. We specify a life-time budget constraint following Hubbard,

Skinner and Zeldes (1995). At each age, cash-on-hand, x_t, is the sum of last-period assets

a_t and total net income y_t (including asset income) while total expenditures is the sum of

non-medical expenditures and out-of-pocket medical expenditures oop_t (more detail is

provided in section 2.5). We assume assets must be non-negative at all ages (the agent

cannot borrow, particularly against pension wealth). End of period assets a_{t+1} are defined

as

$$a_{t+1} = x_t + tr_t - c_t - oop_t, \quad a_{t+1} \geq 0 \text{ for all } t. \tag{3}$$

---

[3] For simplicity, we assume being in excellent, very good and good health yields the same utility.

Government transfers are given by $tr_t = \max(c_{min} - x_t, 0)$ where c_{min} is a minimum consumption level guaranteed by the existence of substance programs such as, temporary assistance for needy families (TANF), food stamps, housing benefit and Supplemental Security Income (SSI).

2.3. Income and Pensions

An agent's total net income is given by

$$y_t = tax(y_t^e, y_t^s, y_t^o, \tau a_t) \tag{4}$$

where $tax()$ is a tax function that depends on annual earnings y_t^e, social security benefits y_t^s, other (spouse) income y_t^o and asset income τa_t.

Earnings - Earnings are assumed deterministic, conditional on age, employment and health status. If the individual is employed, we specify the log earnings equation as

$$\log y_t^e = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \sum_{s=2}^5 \alpha_{3,s} h_{t,s} \tag{5}$$

where $h_{t,s} = I(h_t = s)$. Note that future earnings are uncertain due to uncertainty in health and employment. If unemployed or retired, earnings are zero.

Retirement Income - Prior to the age of 62, individuals do not receive retirement income if they decide to retire. They finance consumption through their accumulated savings until they reach the age of 62. After the age of 62, they are assumed to be eligible for Social Security at which point they can start receiving benefits if they retire. As for those who elect not to retire prior to 62, they can delay retiring/claiming Social Security

until they reach the age of 70 at which point everyone is assumed to retire (and claim).
Social Security benefits are a function of lifetime earnings and the age of retirement. The
measure of lifetime earnings used is the average indexed monthly earnings (AIME)
denoted ae_t, which represents the average of the 35 highest years of earnings. Before
retirement, we approximate the evolution of AIME using the function

$$ae_{t+1} = G(ae_t, y_t^e, t) \tag{7}$$

where $G()$ is an age-specific non-linear regression of next period ae_{t+1} on current ae_t and
earnings, y_t^e. This function is estimated on Social Security earnings records from the
Health and Retirement Study (section 3.1.1).[4] To calculate Social Security benefits at the
normal or full retirement age (NRA), one simply translates the AIME into the Primary
Insurance Amount (PIA) by applying a concave function which is piece-wise linear. If the
individual retires prior to the NRA, his benefit is reduced by an actuarial adjustment factor
(arf). The same holds if the individual retires beyond the NRA. We follow French (2005)
and compute, at the time where someone claims, a new AIME such that

$$pia(ae_t^*) = arf(t) pia(ae_t) \tag{6}$$

holds. The new AIME is then kept permanently until death. Social Security benefits are
then given by $y_t^s = pia(ae_t^*)$ once an individual has claimed.

Other Income – Other income is composed of the spouse's income. It is assumed to be a
function of age and earnings, $y_t^o = o(t, y_t^e)$.

---

[4] The approximation performs well in practice. The main advantage is that we avoid having to include the
entire earnings history in the state-space (French 2005).

2.4. Health

An individual's future health status is modeled as a dynamic ordered response process, which depends on current health status, medical expenditure, through a health investment function, and age. We first discuss the health investment function and then discuss the dynamic health process.

Health Investment – There is little guidance from the literature on the form of the health investment function. As Elrich and Chuma (1990) argue, a function featuring diminishing returns is probably realistic. We use the following production function

$$I(m_t, h_t, t) = \theta_0(t, h_t) \frac{(1 + m_t)^{1-\theta_1} - 1}{1 - \theta_1}. \tag{8}$$

The elasticity of this function with respect to medical expenditures is negative and decreases with m_t if $\theta_1 > 1$. As Hall and Jones (2006) show, the rate at which the elasticity of the health production function to medical expenditures decreases is directly related to the responsiveness of the optimal investment to income changes. Intuitively, with constant life-cycle consumption and mortality rate, an additional dollar spent on health care is more valuable if the life extension it buys is valued more than the utility today from spending that dollar on consumption. The more concave the utility function is relative to the production function, the more valuable is the dollar spent on health. Hall and Jones (2006) consider a constant elasticity production function. We use a more flexible form in order to have additional degrees of freedom when fitting to the data.

The productivity term is given by $\theta_0(h_t, t) = e^{\theta_{00} + \theta_{0h} h_{t,<3} + \theta_{0t} t}$. We allow the productivity term to increase in bad health (fair and poor) and to increase with age. We

interpret the age coefficient in the productivity term as an interaction between depreciation and productivity. As the body depreciates, there is more that technology can do to "restore" health. At high levels of health and younger ages, there is less knowledge in the medical field how to protect the body against rapid depreciation.

Health Process - Define next period's health index as

$$h^*_{t+1} = \gamma_0(t) + I(m_t, h_t, t) + \sum_{h=2}^{5} \gamma_{h-1} I(h_t = h) + \varepsilon_t \tag{10}$$

where ε_a is a standard normal health shock and $\gamma_0(t)$ is a 4th order polynomial in age. The latent health index is transformed into health status through the following rule,

$$h_t = h \quad \text{if} \quad \delta_{h-1} \leq h^*_t < \delta_h, h = 0, ..., 5.$$

where $\{\delta_h\}_{h=1,...,5}$ is a set of thresholds. Because of the normal error assumption, this defines a dynamic ordered probit process.

2.5. Employment and Health Insurance

Employment and health insurance are stochastic and out of the control of the agent. We assume that four states are possible: employed/unemployed with or without health insurance. The states are defined as $e_t = \{0, 1, 2, 3\}$ where zero is unemployed without insurance, 1 is unemployed with health insurance, 2 is employed without health insurance and 3 is employed with insurance. We assume the probabilities of transiting from one state to the next are first-order Markov and depend only on age. Upon retirement, the agents who had insurance coverage on their job retain it while those without have to wait until they become Medicare eligible. If the individual has health insurance ($e_t = 1, 3$), we use a standard health insurance contract with a deductible and co-insurance rate to transform total into out-pocket expenditures.

Hence, we assume the availability of retiree health insurance for those with insurance on the job at the time of retirement.[5]

Out-of-pocket medical expenditures when the agent is not Medicare eligible are given by

$$oop_t(m_t, e_t) = \left[\begin{array}{c} \min(m_t, \mu_{1t}) + \mu_{2t} \max(m_t - \mu_{1t}, 0) \text{ if } e_t = 1, 3 \\ m_t \text{ if } e_t = 0, 2 \end{array} \right. \tag{11}$$

where $(\mu_{1,t}, \mu_{2,t})$ are the corresponding deductible and co-insurance rate respectively if covered at age t. If Medical eligible (age 65 and older), we use a similar insurance contract as when $e_t = 1, 3$. Because we follow annual medical expenditures and not doctor visits and hospital stays it is impossible to incorporate payment rules under Medicare Part A other than through an approximation.[6] But Medicare Part B coverage which covers outpatient procedures is very similar to the median employer provided health insurance contract. We provide more details in section 3.2.

2.6.   Solution of the Dynamic Problem

The agent maximizes the discounted sum of utility flows using a discount factor β. At each age, the agent can choose d_t from a set of possible decisions $D(s_t)$ given the state space $s_t = (t, e_t, h_t, r_{t-1}, a_t, ae_t)$. The indirect utility of the agent at age t is written as

---

[5] See French and Jones (2007) and Blau and Gilleskie (2008) for models where retiree health insurance is also stochastic. Both of these studies take insurance as given and out of the control of the individual (except for the fact that the individual can change/quit a job with insurance). Marquis and Long (1995) analyze the price and income responsiveness of the demand for health insurance in the non-group market. They find a price (premium) elasticity of -0.3 to -0.4 and an income elasticity of 0.1. See Bajari, Hong and Khwaja (2009) for a study looking at health insurance decisions in a life-cycle framework.

[6] See Blau and Gilleskie (2008) for a model where doctor visits rather than medical expenditures are modeled.

$$v(s_t) = \max_{d_t \in D(s_t)} u(d_t) + \beta[\sum_{h!=0}\sum_e p_h(s_t, m_t)p_e(s_t)v(s_{t+1} \mid s_t, d_t) + p_0(s_t, m_t)b(a_{t+1})] \qquad (12)$$

where $p_h(s_t, m_t)$ denotes the probability of being in state $h_{t+1} = h$ at age $t+1$ and $p_e(s_t)$

denotes the probability of employment/health insurance status. Equation (12) is subject to

preferences and constraints from (1) to (11).

This problem can be solved for the optimal solution: $[d^*(s_t), v^*(s_t)]_{t=t_0,\dots,T}$ by

backward recursion. We assume a terminal age of 105 and a starting age of 25. We assume

everyone who is still alive at age 105 dies. The currency used throughout is thousands of

$USD 2004. We provide details on the solution method in the Technical appendix.

2.7.   Simulation

We simulate the life-cycle trajectories of 1,500 hypothetical individuals starting at age

$t_0 = 25$ by first drawing initial conditions from the joint distribution of AIME, assets and

health in the data (see section 3.3). We then apply the decision rules and update the state-

space at each age until individuals reach age 105 (or die). We draw health and

employment/health insurance shocks from their respective distributions. Since the state at

each age might not fall on the grid, we use linear interpolation for consumption, medical

expenditure and retirement decisions.

3. Data and Calibration

We focus on describing the behavior of a relatively homogeneous group because the model

allows for limited heterogeneity. We chose to calibrate the model so that it matches the

behavior of non-hispanic white men without a college education. We could not find

evidence that growth in health spending or life expectancy has been different across socio-demographic groups. Many studies document preference heterogeneity across groups and growth in income has been very different for college graduates relative to the rest of the population.

We use three main sources of data. The first is the Panel Study of Income Dynamics (PSID), which we use for labor force and insurance status, health status, earnings and assets from age 25 till death. The PSID has information on medical expenditures at the household level. A second and more adequate source of data on individual medical expenditures is the Medical Expenditure Panel Survey (MEPS) for the years 2000 to 2003. Finally, we use the HRS for two purposes. First, we use Social Security earnings records of respondents to the Health and Retirement Study to estimate the relationship between AIME and earnings as described in equation (7). Second, the PSID seriously underestimates mortality in old age and has sparse cell counts over age 80. Hence, we use health status (which includes mortality) of respondents over age 50 from the HRS. Since cohort effects may exist, we select where possible data from the cohort born between 1936 to 1940 which is of retirement age (age 60 to 64) in the year 2000.

We use self-reported health as our measure of health. Although we are aware that it is an imperfect measure of health, there are no other categorical measure that provide an all encompassing measure of health over the life-cycle. We reverse the original scale so that 1 equals poor health and 5 excellent health. Furthermore, we add the category 0 which is defined as death. Death is inferred from exit interviews in the HRS and PSID as well as a matching procedure to the National Death Records for HRS respondents.

We first discuss the estimation of auxiliary processes used in solving the model, such as the earnings, other income, AIME, employment/health insurance and finally health processes. We then follow with institutional details such as taxes, computation of Social Security benefits and health insurance premiums and co-insurance levels. We then discuss how initial conditions are drawn (the initial distribution of individuals at age 25). Finally, we discuss the estimation of profiles from the data used to calibrate parameters. This allows us to pin down a configuration of structural parameters matching the data.

3.1. Auxiliary Processes

3.1.1. Earnings, AIME and Other Income

We estimate the earnings profile from the PSID using waves from 1980 to 2005. We use the sample of respondents aged 21 to 70 with positive earnings. After deletion of extreme cases, this leaves records on 3714 respondents with on average 6.4 observations per respondent.[7] We use log earnings as the dependent variables and control a quadratic in age, indicators for health states (2 to 5) and fixed effects. Upon estimation, we use the average of the fixed effects for the 1940 cohort to trace out the earnings profile. Estimates are reported below along with standard errors (in parenthesis).

$$\log(y_t^e) = \underset{(0.226)}{0.809} + \underset{(0.009)}{0.099}t - \underset{(0.001)}{0.001}t^2 + \underset{(0.035)}{0.098}h_{t,2}$$
$$+ \underset{(0.036)}{0.14}h_{t,3} + \underset{(0.037)}{0.166}h_{t,4} + \underset{(0.037)}{0.158}h_{t,5} \tag{13}$$

Our estimates report a substantial penalty from being in poor health ($h_t = 1$) relative to other health states. Relative to being in poor health, being in fair health is

---

[7] We drop observations with earnings larger than \$200 thousand per year and less than \$5 thousands.

associated with a 9.8% increase in earnings, being in good health 14% and being in

excellent health 15.8%. The estimated earnings profile peaks at age 46 with average

earnings of just over $32 thousand dollars.

The average indexed monthly earnings is the average of the highest 35 years of

earnings where each year of earnings is indexed to age 60 wage levels using the National

Wage Index. We use earnings histories to estimate that process. The process for AIME is

estimated from Health and Retirement Study Data merged with Social Security earnings

histories. We use data from those born between 1936 and 1940. We postulate the following

age-specific log-log regression:

$$\log ae_t = \varsigma_{0t} + \varsigma_{1t} I(y_t^e > 0) + \varsigma_{2t} \log y_t^e + \varsigma_{3t} \log ae_{t-1} + \upsilon_t \tag{14}$$

where υ_t is a prediction error. We construct the AIME on the left-hand side using the

computed AIME from earnings histories (35 highest years) and regress on current

earnings, whether earnings are greater than zero. The R-square from such regressions

comes close to 0.99 for most ages. The coefficients at each age are displayed in the

Technical appendix.

We estimate the process for other income using the 1980 to 2005 waves of the PSID. Other

income is defined as spousal income. We estimated the following median regression

$$Q_{0.5}(y_t^o \mid t, y_t^e) = -\underset{(0.319)}{7.97} + \underset{(0.013)}{0.345}t + -\underset{(0.0001)}{0.003}t^2 + \underset{(0.0009)}{0.076}y_t^e + \underset{(0.127)}{5.55}I(y_t^e > 0) \tag{15}$$

which yields an age profile peaking at age 44 with a value of $11.51.

3.1.2. Employment and Insurance Transition Probabilities

We use the 1999 to 2005 waves of the PSID to estimate employment and insurance

probabilities. We define a non-employment as being unemployed and drop observations

where respondents do not participate in the labor force.. We define someone as insured (prior to Medicare eligibility) if he is either covered by employer provided health insurance (own or through spouse) or private insurance. Figure 1a reports the fraction in each state as a function of age. On average 17.9% of those aged 25 to 64 do not have health insurance. The fraction of uninsured workers steadily decreases with age until age 50 and then increases slightly. A small fraction, 3.2%, are unemployed. The fraction of unemployed decreases slightly with age.

We estimate two-year transition probabilities as a function of age and the origin state using a multinomial logit model. The probability of transition to each of the four states is defined by

$$p_e(t, e_{t-1}) = \frac{\exp(\gamma_{0e} + \gamma_{te}t + \sum_{j=1}^{3} \gamma_{ej}I(e_{t-1} = j))}{1 + \sum_{k=1}^{3} \exp(\gamma_{0k} + \gamma_{tk}t + \sum_{j=1}^{3} \gamma_{ek}I(e_{t-1} = k))} \cdot \qquad (16)$$
$$e = 1, 2, 3$$

These are estimated using the sample of respondents age 25-64. We compute annual transition probabilities by taking the square root of the biennial transition matrix at each age.

3.1.3. Health Process

To estimate the health process, we ideally need panel data on both health and individual level medical expenditures. MEPS is a short panel (on average 2 years) that does not allow to estimate annual transition probabilities across health states reliably. On the other hand, the PSID does not have individual level data on medical expenditures (only at the household level). Finally, the PSID does not capture mortality well, particularly in old age. The HRS provides better coverage in old age but lacks data on the population younger than

50 and has very noisy medical expenditure data compared to the MEPS. Our strategy is to combine all three data sources for the purpose of calibrating the health process. Self-reported health is defined in the same way in all three surveys. Hence, we resort to the following strategy:

- Estimate the distribution of medical expenditures as a function of self-reported health status and age in the MEPS. We assume the distribution has two-parts, first whether expenditures are positive or not and then log normally distributed when positive. We keep both the fraction of positives, mean and variance of the log normal distribution conditional on having positive expenditures for each age and health state.

- Pool together both PSID data for those age <50 (waves 1999 to 2005) and HRS data for those age >50 (waves 1992 to 2006). The distribution of self-reported health is remarkably similar at the cutoff ages of 50 as can be seen from Figure 1b.

- Estimate parameters of the health process (γ) other than those of the investment function (θ) by maximizing the following likelihood:

$$
L(\gamma \mid \theta, h_{it}, t_i) = \prod_{i,t \in PSID} \int p_{h,PSID}(t, h_{it}, m_{it}; \theta) f_{MEPS}(m_{it} \mid h_{it}, t) dm_{it}
$$

$$
\times \prod_{i,t \in HRS} \int p_{h,HRS}(t, h_{it}, m_{it}; \theta) f_{MEPS}(m_{it} \mid h_{it}, t_i) dm_{it}
$$

(17)

where the probabilities $p_{h,x}(t_i, h_{it}, m_{it}; \theta)$ are ordered probit probabilities for observing in dataset x an individual at age $t+1$ in state $h_{i,t+1} = h$ given age, current health status, medical expenditures and conditional on the productivity parameters θ. The integrals are evaluated using simulation by drawing from the distribution of medical expenditures in the

MEPS. We use 25 draws per respondent/wave. The resulting maximum simulated likelihood estimator $\hat{\gamma}(\theta)$ provides us with a calibration that fits the health process given a choice of productivity parameters. We perform this conditional procedure because we calibrate θ using other moments from the data. Based on various diagnostic tests, it became apparent that the age-invariant assumption on the thresholds of the dynamic ordered model was too restrictive, particularly for mortality ($h_{it} = 0$). Hence, we allowed the threshold defining the upper bound for mortality to depend on a fourth order polynomial in age.

3.2.   Institutional Details

There are three important institutions represented in the model. First, there is the income tax schedule. Second, the Social Security benefit calculation formula and finally the Medicare and Non-Medicare health insurance out-of-pocket schedule.

We use the 2001 Tax function taking into account Federal taxes and Social Security & Medicare taxes. We allow for basic deduction in couples and the personal exemption. Prior to age 65, the deduction is $7.6 and after 65 it is $9.65. We account for the partial tax treatment of Social Security benefits by taxing only 50% of Social Security benefits.

Social Security benefits are basically a function of three factors: a measure of average earnings (AIME), birth year, and age at which benefits are first drawn. For our purposes, the Social Security earnings test is irrelevant since we assume individuals withdraw from the labor force at the time where they claim benefits. Furthermore, we model spouse benefits through other income such that the other spouse's receipt of Social Security does not play a distinct role. We assume individuals are born between 1936 and

1940 and that their Normal Retirement Age (NRA) is 65 years of age (in reality it was 65 for those born in 1936 and 65 and 6 months for those born in 1940). To calculate the monthly benefit, the AIME is transformed into the Primary Insurance Amount (PIA). The PIA is a piece-wise linear concave function of the AIME. The PIA is the benefit that would be paid if claimed at the NRA. If the individual claims prior to the NRA, the PIA is reduced by 7.6%. Similarly, an individual can claim his benefit past the NRA in which case a delayed retirement credit (DRC) is granted. The credit is 6.5% for those born in 1938 (6% for those born in 1936 and 7% for those born in 1940). Upon reaching 70, there is no DRC applied.

For those younger than 65 and insured, we use data reported by Blau and Gilleskie (2008) from the Health Insurance and Pension Provider Survey (HIPPS). The median deductible is found to be $0.2 and the median co-insurance rate is 20%. The median premium is $0.48. We do not impose a stop-loss on the insurance contract (maximum out-of-pocket). Since there is no direct cost to investing in health, individuals in the model who desire to spend slightly more than the stop-loss would then consume an infinite amount of care.

Establishing an overall co-insurance rate and deductible for Medicare is more complicated. Only Medicare Part B, which covers outpatient treatments, has a common deductible-coinsurance structure (part D was not in place for those respondents at the time). There is no premium for part A and the cost-sharing schedule depends on hospital stays, which we do not model. The premium for Medicare part B is $0.49 per year, the deductible $0.2 and the co-insurance rate 0.2 (20%). Hence, we choose to simply use the same price schedule for Medicare and Non-Medicare insurance. Therefore, within the

context of our model, there is very little difference in coverage prior and after Medicare eligibility for those insured. The main difference is the uncertainty in insurance coverage.

We applied the insurance parameters to total medical expenditures and compared with actual out-of-pocket expenditures reported in MEPS. At the mean, the computed out-of-pocket expenditures are very similar to those from the data, both before and after the age of Medicare eligibility.

3.3. Initial Conditions

We draw initial conditions for assets, AIME, insurance coverage and health status from the empirical distribution in the PSID data at age 25 (we use age 24 to 28 to boost the sample and replicate observations using frequency weights). However AIME is not measured in the PSID. Since AIME is very close to earnings for those respondents (they have little labor market experience) we apply the following procedure. We first compute someone's percentile in the earnings distribution at age 25. We then attribute the AIME of someone in the same percentile of the earnings distribution from the HRS restricted SSA earnings records. This imputation procedure preserves the correlation between earnings and other outcomes (assets, insurance coverage and health status).

3.4. Moments

To calibrate the model, we use four sets of moments. First, we use median assets by age from the PSID. Second, we use retirement hazard rates by age from the Health and Retirement Study. Third, we use average medical expenditures by age from the MEPS. Fourth, we use mortality rates from ages 50 to 100 computed from the Health and

Retirement Study. We discuss the construction of those moments below. Figure 2 reports estimates of those moments by age.

3.4.1. Assets

We estimate a median asset age-profile using the PSID following the methodology of French (2005). Assets include all real and financial assets minus debt. We express all monetary amounts in $2004 and discard observations with assets larger than $1e+06. We estimate a median regression model of the form

$$Q_{0.5}(a_t \mid t, cohort, hhsize) = \alpha_0 + \alpha_t + \lambda_{cohort} + \beta hhsize$$

where α_t are age fixed effects. We also control for 5-year cohort fixed effects (λ_{cohort}) and household size ($hhsize_a$). Upon estimation we predict for household size of 3 born between 1936 and 1940.

3.4.2. Retirement

Retirement is a complex state to measure. First, some workers return to work after prolonged periods of inactivity in old age while not considering themselves retired. Second, some workers work while claiming Social Security benefits. Since our model considers a simplified definition of retirement, it is not clear which concept should be used. We elected to use self-reports of respondents on whether they are retired in the HRS. We exclude as retired those who report being retired but are still at work. We use the retirement ages from the cohort born between 1931 and 1941 which has almost entirely

retired by 2004. We find peaks in the retirement hazard at age 62 and 65 which are respectively the ages of early and normal retirement age under Social Security.

3.4.3. Medical Expenditures

We use median total medical expenditures from the 2000 to 2003 waves of the MEPS. These expenditures include in-patient and out-patient expenditures as well as drug expenditures. They do not however include nursing home expenditures. To be consistent with other monetary flows in the model, we express those in $2004 dollars using the CPI. Not surprisingly, medical expenditures increase steadily with age.

3.4.4. Mortality

We use the mortality profile from the Health and Retirement Study (older than 50) rather than from official life tables because we are looking at a specific sub-population (non-hispanic white men with high school or less education). We express biannual rates into annual ones assuming a constant hazard within each age interval. Even with large samples, the mortality rates at older ages are quite volatile as can be seen from Figure 3. Smoothing these mortality rates (using a quadratic in log mortality rates) and computing a life-table yields remaining life expectancy of 47.3 years at age 25 and 26.9 years at age 50. We also report the mortality rates for men in 2002. White males without college education have slightly lower life expectancy than average males. Life expectancy at age 25 according to the life table is 49.2 years and 27.6 at age 50.

3.5.   Calibration

3.5.1 Preferences and Production Function

In principle, we can estimate all structural parameters of the model using a method of simulated moments approach. For this paper, we have instead used a calibration approach, which consist of choosing manually parameters such that the fit of the moments is satisfactory for our purposes. However, we wish to emphasize that there is no way for us to claim that this is the unique and optimal set of parameter estimates for this model. There might well be other parameters that provide an equally (or even better) fit of the model. In choosing these parameters, we have borrowed partly from the literature on consumption and labor supply. We should note that no estimates of parameter values exist for a model that simultaneously features endogenous health spending, savings and retirement. Table 1 provides the values of the parameters we have chosen.

Estimates of risk aversion vary substantially across studies. Studies using linearized Euler equations and consumption data tend to estimate risk aversion parameters in the range 1.5 to 2.5 (Attanasio and Weber, 1995). Estimates based on asset data tend to be much larger. Cagetti (2003) estimates on PSID data risk aversion parameters ranging from 2.7 to 4.3. Recent estimates by French (2005), French and Jones (2007) and De Nardi, French and Jones (2006) suggest a risk aversion coefficient in the range 3-7. Each of these models are different and the estimate of risk aversion can be very sensitive to the amount of risk assumed, the sample used (elderly vs. young), the extent of liquidity constraints and the presence of a consumption floor. One of the reasons for the larger values in these recent studies is the role played by the consumption floor in depressing savings (Hubbard, Skinner and Zeldes, 1995). Hence, a larger value of risk aversion is needed to rationalize savings when there is a consumption floor. We chose a value of 2.6 for risk aversion. We

fix the consumption floor at $3000 following the argument that few of the social programs in place apply to married men (accounting for 85% of the sample) and that take-up of those problems is generally well below 50% (De Nardi, French and Jones, 2006).

The real rate of return is set to 3%. There is considerable heterogeneity in estimates of the discount rate. Some studies find very low values, even negative [-0.04, -0.02] (Hurd, 1989; French, 2005) while others find relatively larger values [0.04 to 0.07] (Gourinchas and Parker, 2002; Cagetti, 2003; De Nardi, French and Jones, 2006). We use a value of 0.05. Hence, individuals can be characterized as relatively impatient and precautionary savings play an important role in consumption decisions. We impose a relatively strong bequest motive in order to rationalize the high level of assets in old age (relative to annuity income). The value of θ_b is set to 0.05.

Next, we pin down retirement by choosing the marginal utility of leisure. We select values such that they capture the distribution of retirement ages and a median retirement age of 63.5 years.

Finally, we calibrate the health production function and disutility/utility of poor health status. We use the average medical expenditure profile, which rises almost linearly with age. This, along with the life-cycle path of health and mortality informs on the productivity of medical care.[8] We find that low values of productivity are necessary to rationalize why individuals spend so little on their health at earlier ages given the potentially important returns in terms of wages and consumption benefits. Without an age adjustment, the medical expenditure profile is much flatter as we cannot rationalize why

---

[8] Trivially, at zero productivity, there are no simulated medical expenditures. Increasing the productivity will lift the simulated medical expenditure profile. But if the simulated profile is below the true profile, mortality will be over-estimated and vice versa if productivity is "too high". Because the depreciation adjustment is with respect to the actual medical expenditure profile, it is not until the actual and simulated profile are close that mortality will be matched.

agents delay health investments until old age. Given that the health production function is concave, there is an incentive to smooth health investment over the life-cycle. We calibrate the age-related increase in productivity so that we match the medical expenditure profile. The increase in productivity due to poor/fair health is calibrated by making use of the differences in health expenditures by health status. Total medical expenditures are 8 times larger for those in poor health than those in excellent health in the MEPS. Without the health related productivity parameter, we can only generate a factor of 5.

We use a second piece of information to calibrate the health production function, in particular its curvature. We show in the technical appendix, using a simplified version of the Hall and Jones (2007) model, that both price and income elasticities depend on the curvature of the health production function. The intuition is that in a health production model, the rate at which the marginal product of medical expenditure relative to the marginal utility of consumption increases as income or price increases is directly related to the value of life extension and therefore to the demand for health.

Hence, we calibrate the health production function parameter so that it replicates roughly, among those of working age, the price (co-insurance) and income elasticity found in the literature. The co-insurance elasticity tends to fall between -0.1 and -0.4 depending on the type of services (preventive vs. curative and in-patient vs. out-patient) (Ringel et al., 2000). Estimates from the Health Insurance Experiment (HIE) (Manning et al., 1987) yields an estimate around between -0.17 and -0.22 which varies across co-insurance rates and is slightly larger for preventive care (-0.17 to -0.43). For calibration, we consider a change from a co-insurance rate of 0.2 to 0.5 and calculate the resulting change in medical expenditures.

As for the income elasticity, cross-sectional studies tend to obtain estimates in the range of 0.2 to 0.4 while studies using cross-country or time-series variation find an income elasticity close to unity. Cross-sectional studies face the problem of measurement error and reverse causality. Income is in part determined by health and other common factors, such that regressing medical expenditures on income, even controlling for imperfect measures of health, leads to a negative bias in the income elasticity. Classical measurement error in income might add to this downward bias. One possible reason for the larger elasticity at the cross-national level is that such analyses omit differences in technology which are correlated with income (higher income countries have better technology). This leads to an upwards bias which is likely more severe in studies looking at time-series data (as income and technology are increasing simultaneously) (Newhouse, 1992). There is yet another important difference between cross-sectional and aggregate cross-national studies which results from the nature of income changes. If the nature of income shocks is transitory, the response is likely to be smaller than if the shock is permanent. A permanent income shock raises the value of living longer if the value of life is a normal good. Viscusi and Aldy (2003) find that the value of life increases with income when performing a meta-analysis of value of life estimates. Hence, the permanent responses should be larger than transitory responses as the latter will not lead to a rise in the value of life.[9] Essentially, cross-sectional studies involve a mix of evolutionary, transitory and permanent income differences while cross-national differences involve to a large extent permanent differences in income across countries. Therefore, we will attempt

---

[9] Dustmann and Windmeijer (2000) make a similar point but looking at time allocation jointly with medical investment. In that case the transitory wage response can be negative as workers spend less time investing in health when their wage rises temporarily while the positive income effect dominates for permanent wage changes. They find support for such effects in German panel data.

to replicate an income elasticity to a permanent income change of 10% and a transitory (anticipated) change in income. We expect a permanent income elasticity of roughly unity while the transitory response should be in line with cross-sectional studies, perhaps larger due to absence of measurement error in our simulated data.

We set the curvature of the health production function to $\theta_1 = 2.1$. Hence the production function is assumed to be less concave than the utility function. We show in section 3.5.3 the resulting elasticities.

3.5.2 Fit of Moments

In Figure 3, we compare moments of the data with those calculated on simulated data. Given the admittedly coarse calibration, the fit is relatively good. First, mortality is relatively close to the HRS data. In particular, we simulate a remaining life-expectancy of 27.1 years at age 50 while it is 26.6 in the HRS for this group. The progression of mortality is very well approximated before age 75 while we under-estimate mortality in the ages 75 to 90 years old. The simulated average medical expenditure profile tracks the data relatively well. We over-estimate expenditures in middle ages and under-estimate at older ages. The asset profile follows the hump-shape found in the data and the slow decline after retirement. However, we tend to over-estimate assets just before retirement. Finally, the simulated retirement pattern captures well the spike at age 62 and the increasing hazard beyond age 65. However, it misses the peak at age 65. One reason is that we have assumed everyone with insurance at the time of retirement had access to retiree health insurance. Hence, those retiring at 62 and 63 do not have an incentive to delay until age 65.

3.5.3 Co-Insurance and Income Elasticities

Table 2 presents various price (co-insurance rate) and income elasticities from the simulation. We report those for 3 different age groups (25-35, 35-45 and 45-55). We focus on individuals younger than 55 (few studies of price and income responses exist among the elderly). In the first column, we compare total medical expenditures for those insured in the baseline scenario with total medical expenditures for the same people in a scenario where we increase the co-insurance rate over the entire life-cycle. This yields co-insurance elasticites ranging from -0.26 to -0.35 which falls in the range of estimates reported in the literature. The elasticity falls with age primarily because medical expenditures become more of a necessity with age, hence less elastic. Individuals are generally in better health at age 25 such that health expenditures more often take the form of preventive or precautionary health investment. These expenditures are more likely to be price sensitive since there is a substitute (save until health deteriorates). Results from the HIE show that elasticities can be as high as -0.43 for preventive care services.

In the second column, we report results for a permanent income change of 10%. The elasticity estimate is large in the earlier age group (1.56) while just below unity in the older age groups (0.88 for 35-45 and 0.909 for those 45-55). Overall, the elasticity is 1.02 at the aggregate level. The last column shows that a large component of this response is not due to the current change in income but to the fact that the change is permanent. The elasticity for a transitory change ranges between 0.5 and 0.65. The permanent income change raises the value of life by a large amount in the age group 25-35. We take this as evidence that our calibration of the production function is broadly consistent with existing evidence on the sensitivity of the demand for health to co-insurance rates and income.

4. Simulations

We use the model to look at the relative contribution of various factors to the growth in health spending and life expectancy in the last 50 years. According to a recent report by the CBO (2008), per capita medical expenditures are roughly 500% larger today than they were in 1965 which represents an average growth rate of 3.6% . Figure 4 shows, using data from Cutler and Meara (1998), average medical expenditure by age in 2004 relative to 1963 (adjusted in 2004 dollars using CPI). The largest change is found for the elderly (65+). Over this period, there has been an increase in life-expectancy at age 50 of 6 years for men (from 22.9 to 28.9 years (6 years)). We investigate whether our model can replicate these changes as a result of growth in income, generosity of insurance and technological change.

Newhouse (1992) reviews the reasons behind the increase in health spending. The generosity of health insurance has increased over the last 50 years and the price elasticity of the demand for health is negative. The average co-insurance rate has fallen from an average of 0.6 in the 1960s to 0.2 in 2005 (Newhouse, 1992). On the income side, real earnings have increased by 27.3%. Social Security has become more generous, which has also raised the income of retirees by an average of 20% for medium and low earners (Diamond and Gruber, 1997). [10] Given the positive income elasticity of demand, this should also explain a share of the rise in health spending as argued by Hall and Jones (2006). Finally, technological change has lead to large improvements in the treatment of several conditions (Deaton, Cutler and Lleras-Muney, 2006) and some have argued

---

[10] Other factors that could explain the rise include population structure (aging), physician induced demand and finally lagging factor productivity in the health sector. Newhouse (1992) reviews the evidence on each of those and finds little ground for the hypothesis that these explain a large share of what actually happened over the period.

(Newhouse, 1992) that the bulk of the increase is due to technological change, which increases productivity. There is little evidence on the magnitude of productivity improvements. Cutler (2004) looks at cardio-vascular diseases, the leading cause of death among men over this period, and concludes that medical advancements explain 2/3 of the improvement in mortality over the last half century, the other third likely explained by the decrease in smoking among men. Lichtenberg (2003) provides evidence that a large part of the gain in life expectancy in recent decades could be attributed to pharmaceutical innovation.

We first implement income and insurance changes at the same time to obtain a 1965 scenario. This means reducing average earnings by 27%, reducing the generosity of social security benefits by 20% and increasing co-insurance rates to 0.6. This should not get us to 1965 spending level unless technological change plays a role. We model technological change as a scalar that varies the productivity term in equation (8) as a function of calendar time s:

$$\theta_0(z_s, h_t, t) = z_s e^{\theta_{00} + \theta_{0h} h_{t,<3} + \theta_{0t} t}, \quad z_s = e^{\theta_z s}$$

where $s = 0$ refers to 2005 ($z_0 = 1$). We calibrate the change in productivity so that matches the level of medical expenditures in 1965. We find an annual productivity decline $\theta_z = -0.018$ roughly fits the data such that $z_{50} = 0.4$. Table 3 reports a number of outcomes in the 1965 and 2005 scenario. The simulated 1965 scenario yields spending levels close to those that would be predicted if we used the ratio of 2005 to 1963 spending levels (spending is 6 times higher in 2005) produced by CBO. The predicted average spending would by $0.697 while we obtain $0.740. The simulated average annual rate of increase in health spending is 3% in real terms or 400% total. We simulate that the share of

net income devoted to health spending is 15% in 2003 while it is 3% in 1965. Given minor differences in the definition of income (we use net income rather than gross income or GNP), this rate of growth is consistent with the data complied by the CBO (4.5% in 1965 and 14.5% in 2005). Another important finding is that the age pattern of medical expenditures is much flatter in 1965 than in 2005, except at very young ages. Cutler and Meara (1998) report data on the age distributions of total medical expenditures from National Health Surveys between 1963 and 1987. Although we replicate the much more rapid growth in old age (relative to the age 35-45), our simulation also generates high growth for those aged 25-34. Cutler and Meara (1998) report very little growth in that age group. We take this as suggestive evidence that technological change has been concentrated on the older population or on curative rather than preventive care. Nevertheless, our calibration of technological change is probably a good first-order approximation to the scale of productivity improvements over this period.

We simulate that life expectancy at age 50 has grown by 16% between 1965 and 2005. The relative change from 2005 to 1965 was 25.8% (for all men). Hence, this explains roughly 60% of the improvement in life expectancy between 1965 and 2005. Note that we do not model other changes in health, such as the reduction in smoking over this period. Hence, it is not surprising that we cannot explain the entire gap.

We also look at the average retirement age. Interestingly, the average retirement age does not change much. However, as we will see next, this is likely the result of the fact that the various changes (e.g. income, technology, health insurance copayment) lead to cancel out different effects on retirement.

Next we look at the relative contribution of each of these changes (insurance, income and technology) in explaining the rise in spending and life expectancy. We sequentially introduce changes starting from the situation in 1965. Results are presented in Table 4.

As can be seen from the second column of Table 4, increasing the generosity of health insurance increases substantially medical expenditures. The average rises from $0.742 to $1.725 which represents a change of 132%. The effect is larger among younger individuals who we have found are more price sensitive (Table 2). Increasing the generosity of health insurance raises the income share of medical expenditures from 3.1% to 7.2%. There is no change in the average age of retirement. An important finding is that increasing the generosity of health insurance has a very small effect on life expectancy, less than 0.3 years. This implies that the additional production of health has little health benefit and would not be consumed if individuals had to pay for it. This is consistent with the evidence from the HIE that generous insurance had little effect on health outcomes within the time frame of the study.

Increasing real earnings and the generosity of social security has surprisingly little effect on health spending. This is shown in column 3 of Table 4. Total expenditures increase by 44% from $0.742 to $1.071. The average change in net income is 16.6%. Part of the reason why expenditures do not rise sharply is due to the responsiveness of the retirement age to the income change. We find that it reduces the retirement age by 2.4 years. This effect, along with the progressivity of income taxes, mitigates the income effect. Furthermore, the ratio of medical expenditures to income rises only by 1 percentage point. Finally, there is a negligible gain in life expectancy of 0.19 years.

In column 4, we increase productivity at 1.8% a year until 2005. This implies an improvement of 150% in productivity over the period. This increases total medical expenditures by about as much as they increase with the change in insurance generosity. The magnitude of the change is similar across age groups which is most likely the result of our assumption that productivity gains were distributed equally over the life-cycle and across health status. The average retirement age increases by 1 year, in most part due to the fact that individuals can now expect to live 3 more years due to technological improvements. Of the three changes (insurance, income and technology), technological change is the only one that is capable of generating appreciable gains in life expectancy. By itself, it explains roughly half of the increase in life expectancy from 1965 to 2005.

None of the scenarios are able to generate health expenditure levels of the magnitude witnessed in 2005, even if we sum their relative contribution. For example, insurance leads to a 132%, income a 44% and technology a 110% change in spending. This sums to 287% , i.e. well under the 464% that we reach when we implement all changes together. Hence, we must conclude that there are important interaction effects. This exercise would imply these interactions are responsible for a rise of 176% in health care spending over the last 50 years.

In Figure 4, we show the sources of increase in health spending. Increased generosity of health insurance accounts on its own for 29% of the insurance in expenditures while income only for 10%. Technological change accounts for 24% of the change. But the largest component is the interaction effect, which amounts to 37% of the total change.

As for life expectancy, we report in Figure 5 the sources of the increase as we estimate them from the model. Technological change, taking the form of productivity gains accounts for the bulk of the increase, 59%. Income and insurance explain a very small share of the total increase (4% and 6% respectively). Interaction effects are smaller for life expectancy than they are for health spending, amounting to 6%. This leaves a residual category of 25% of the total increase in life expectancy. Several candidates are plausible to explain the remaining part, one is the reduction in smoking, the second is general improvements in living conditions not captured by the model (air quality, working conditions, etc).

5. Conclusion

We calibrate a stochastic life-cycle model of health spending, asset accumulation and retirement which allows us to look at the determinants of the rise in health spending and longevity over the last 50 years. We find that with reasonable parameter values, the model fits the data relatively well and is able in particular to replicate the income and price elasticity of the demand for medical care. We then perform a simulation where we attempt to go back to 1965 by simultaneously decreasing the generosity of health insurance and income to 1965 levels. We back out the rate of productivity gains necessary to achieve 1965 spending levels and longevity and find that a rate of 1.8% fits the data relatively well.

We find that both insurance and technological change independently explain respectively 29% and 24% of the total change in health spending and that income explains roughly 10%. The remainder, 37%, is due to what we call synergy effects or interactions created by simultaneously increasing income, the generosity of health insurance and productivity. As for life expectancy, we find that the bulk of the gain is due to productivity

growth while income growth and increased generosity of health insurance explain a

relatively small fraction of the overall increase (less than 10%).

References

Adams, P., M. D. Hurd, D. McFadden, et al. (2003). "Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status." *Journal of Econometrics* **112**(1): 3-56.

Attanasio, O. P., and Weber, G. (1995), "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence From the Consumer Expenditures Survey," *Journal of Political Economy*, 103, 1121-1157.

Bajari, P. , H. Hong, and A. Khwaja (2009): "Moral Hazard, Adverse Selection and Health Expenditures: A Semiparametric Analysis", Mimeo Duke University.

Blau, D.M. and D. Gilleskie, (2008), "The Role of Retiree Health Insurance in the Employment Behavior of Older Men" *International Economic Review,* 49(2): 475-514.

Cagetti, M. (2003), "Wealth Accumulation over the Life Cycle and Precautionary Savings" *Journal of Business & Economic Statistics*, Vol. 21, No. 3 (Jul., 2003), pp. 339-353.

Congressional Budget Office (2008) Technological Change and the Growth of the Health Care Spending, A CBO paper, January 2008.

Cutler, D.M. (2004), *Your Money or Your Life*, Oxford: Oxford University Press, 2004.

Cutler, D. M. and E. Meara, (1998). "The Medical Costs of The Young and Old: A Forty Year Perspective" in Ed. D.A.Wise (1998) *Frontiers in the Economic of Aging*, pages 215-246, University of Chicago Press.

Cutler, D., A. Deaton and A. Lleras-Muney (2006). "The Determinants of Mortality," *Journal of Economic Perspectives*, American Economic Association, vol. 20(3), pages 97-120.

Diamond, P. and J. Gruber, 1997. "Social Security and Retirement in the U.S." NBER Working Papers, 6097, National Bureau of Economic Research, Inc.

De Nardi, M., French, E. and Jones, J.B. (2006), "Differential Mortality, Uncertainty Medical Expenses, and the Saving of Elderly Singles", Working Paper 12554, NBER.

Dustmann, C. and Windmeijer, F., 2000. "Wages and the Demand for Health - A Life Cycle Analysis," IZA Discussion Papers 171, Institute for the Study of Labor (IZA).

Ehrlich, I. and H. Chuma (1990), "A Model of the Demand for Longevity and the Value of Life Extension", The *Journal of Political Economy*, Vol. 98, No. 4, pp. 761-782

French, E. and J.B. Jones (2007): "The Effects of Health Insurance and Self-Insurance on Retirement Behavior", MRRC working paper 2007-170.

French, E. (2005). "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour." *Review of Economic Studies* 72(2): 395-427.

Galama, T., A. Kapteyn, R. Fonseca and P.-C. Michaud (2008): "Minimally Productive Health and Retirement", WP -2008 RAND.

Gourinchas, P.-O. and J.A. Parker (2002). "Consumption Over the Life Cycle," *Econometrica*, Econometric Society, vol. 70(1), pages 47-89, January

Grossman, M. (1972), "On the concept of health capital and the demand for health" *Journal of Political Economy* 80: 223-255.

Habermann, C. and F. Kindermann (2007): "Multidimensional Spline Interpolation: Theory and Applications," *Computational Economics*, Springer, vol. 30(2), pages 153-169.

Halliday, T.J., He, Hui and Zhang, H. (2009) "Health Investment over the Life Cycle", IZA DP No. 4482 October.

Hall, R.E. and Jones, C.I. (2007), "The Value of Life and the Rise in Health Spending" *The Quarterly Journal of Economics*, February 2007.

Hubbard, R. G., Skinner, J., & Zeldes, S. P. (1995): "Precautionary saving and social insurance*". Journal of Political Economy*, 103(2), 360-399.

Hurd, Michael D, (1989) "Mortality Risk and Bequests," *Econometrica*, Econometric Society, vol. 57(4), pages 779-813, July.

Jung, J. and C. Tran (2007): "The Macroeconomics of Health Savings Accounts." CAEPR Working Paper 2007-023

Khwaja, A. (2009) "Estimating Willingness to Pay for Medicare Using a Dynamic Life-Cycle Model of

Demand for Health Insurance," Fuqua School of Business, Duke University, forthcoming, *Journal of Econometrics.*

Lichetenberg, F. (2003) "Pharmaceutical Innovation, Mortality Reduction, and Economic Growth," in *Measuring the Gains from Medical Research: An Economic Approach*, ed. by Kevin M. Murphy and Robert H. Topel (Chicago: University of Chicago Press 2003), pp. 74-109.

Manning W.G., Newhouse, J. P., Duan, N., Keeler, E., Benjamin, B., Leibowitz, A., Marquis, M.S. and Zwanziger, J. (1987), Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment. *The American Economic Review*, 77:251-277. June 1987.

Marquis, S.M., S.H. Long (1995): Worker demand for health insurance in the non-group market, *Journal of Health Economics*, Volume 14, Issue 1, May, Pages 47-63,

Newhouse, J. P., (1992), "Medical Care Costs: How Much Welfare Loss?" *The Journal of Economic Perspectives*, Vol. 6, No. 3, pages 3-21.

Newhouse, J. P., and the Insurance Experiment Group (1993), *Free For All? Lessons from the Health Insurance Experiment*, Harvard University Press, Cambridge.

Ringel, J., Hosek, S.D., Vollaard, B.A. and Mahnovski, S. (2000): "The Elasticity of Demand for Health Care A Review of the Literature and Its Application to the Military Health System, RAND Report.

Rust, J. and C. Phelan (1997). "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets." *Econometrica* 65(4): 781-831.

Smith, J. (2007): "The Impact of Socioeconomic Status on Health over the Life-Course," *Journal of Human Resources*, 42(4):739-764, Fall.

Suen, R. (2009): "Technological Advance and the Growth in Health Care Spending", Conference Health and Macroeconomics in Santa Barbara.mimeo.

Viscusi, W.K. and J. Aldy (2003). " The Value of a Statistical Life: A Critical Review of Market Estimates throughout the World," *Journal of Risk and Uncertainty*, Springer, vol. 27(1), pages 5-76, August.

Yogo, M. (2009): "Portfolio Choice in Retirement: Health Risk and the Demand for Annuities, Housing, and Risky Assets" *mimeo*

Technical Appendix

Solution Method

Because some of the state variables are continuous (assets and AIME), we define a bi-dimensional equally-spaced grid over that space. We select 35 points for assets and 35 points for AIME. The maximum for assets is set to 300 and the minimum 0. For AIME, the minimum is zero and the maximum 6. Since decision rules are likely more non-linear at low values of assets and AIME, we set the grid in terms of equally spaced points using the square root of assets and AIME ($a_t^{1/2}$ and $ae_t^{1/2}$). We solve for value function at each point on that grid starting at terminal age T.

       We have two continuous decision variables. We define a grid for each of these decisions (45 points for consumption and 45 for medical expenditures). We select bounds of 0 and 25 for medical expenditures (25 is the 99th percentile in the data) and define the grid for consumption in terms next period's assets using the boundary conditions along with cash on hand. For consumption, we center the grid around current cash on hand and define an equally spaced grid away from that point. We use interpolation to calculate next period's value function when the solution for consumption and medical expenditure does not fall on the grid given the current state and retirement choice. Since we use an equally-spaced grid, we use the bi-dimensional cubic spline approximation proposed by Habermann and Kindermann (2007). When retirement is an option, we compute the optimal solution for each retirement path and compare value functions to calculate $r^*(s_t)$.

       The decision rules are generally insensitive to the number of grid points. We solve for decision rules using a program written in Ox from OxMetrics (with sub-programs written in C++) and the parallel message parsing interface MPICH on a 16 processor core cluster with 8 GB of RAM. A typical solution takes roughly 12 minutes.

AIME Approximation

The following table provides age-specific coefficients for the AIME regression. The average R-squared is 0.991. For ages above 60, the age 60 coefficients are used.

Table A.1 AIME Approximation Coefficients

age	earnings>0	log(earnings)	log(AIME(t-1))	Constant
26	0.0848	0.1380	0.7425	-0.6538
27	0.0706	0.0793	0.8654	-0.2817
28	0.0713	0.0794	0.8360	-0.3334
29	0.0221	0.0768	0.8792	-0.2243
30	0.0511	0.0578	0.8719	-0.2093
31	0.0072	0.0541	0.9255	-0.1109
32	0.0075	0.0665	0.8884	-0.1772
33	0.0605	0.0447	0.9012	-0.1542
34	0.0104	0.0584	0.8906	-0.1521
35	0.0014	0.0485	0.9140	-0.1015
36	0.0588	0.0208	0.9410	-0.0589
37	0.0513	0.0260	0.9312	-0.0719
38	0.0269	0.0301	0.9430	-0.0575
39	0.0269	0.0208	0.9636	-0.0265
40	0.0166	0.0234	0.9645	-0.0259
41	0.0033	0.0228	0.9777	-0.0128
42	0.0259	0.0225	0.9610	-0.0323
43	0.0158	0.0181	0.9818	-0.0117
44	0.0048	0.0175	0.9871	-0.0013
45	0.0058	0.0146	0.9916	0.0048
46	0.0060	0.0165	0.9874	-0.0036
47	0.0033	0.0185	0.9807	-0.0046
48	0.0047	0.0170	0.9870	-0.0051
49	0.0059	0.0133	0.9912	0.0018
50	0.0056	0.0129	0.9931	0.0008
51	0.0040	0.0119	0.9956	0.0019
52	0.0000	0.0120	0.9964	0.0031
53	0.0163	0.0121	0.9757	-0.0022
54	0.0135	0.0189	0.9567	-0.0080
55	0.0060	0.0123	0.9840	-0.0016
56	0.0056	0.0104	0.9880	0.0004
57	0.0158	0.0076	0.9853	-0.0008
58	0.0134	0.0068	0.9875	0.0006
59	-0.0019	0.0144	0.9674	0.0065
60	-0.0056	0.0176	0.9578	0.0105

Price and Income Elasticities

We demonstrate how the price and income elasticity of demand are related to the concavity of the health production function. We add a co-insurance rate (price) to the model used by Hall and Jones (2006). Let $1/h$ be the mortality rate and h life expectancy (health status). Income y can be allocated between consumption c and medical expenditures m. Out-of-pocket medical expenditures are given by μm where μ is the co-insurance rate. Health is produced trough a production function $i(m)$ which is increasing and concave. The problem is given by

$$\max U(c,h) = \int_0^\infty e^{\frac{1}{h}t} u(c)\, dt = hu(c)$$

$$s.t.$$

$$y = c + \mu m$$

$$h = i(m)$$

As in Hall and Jones, define $\eta_m = \frac{i'(m)}{i(m)} m, \eta_u = \frac{u'(c)}{u(c)} c$ and $\lambda = \dfrac{\eta_m}{\eta_m + \eta_u}$. The optimality condition is

$$s = \frac{m}{y} = \frac{1}{\mu}\lambda .$$

The income elasticity of optimal medical expenditures $m*$ is given by

$$\eta_y = 1 + \frac{y^2}{\mu m}\lambda_y$$

where λ_y is the derivative of λ with respect to y. The uncompensated price elasticity is

$$\eta_\mu = -1 + \frac{y}{m}\lambda_\mu.$$

Both λ_μ and λ_y depend on the curvature of the utility and production functions since they involved second derivatives of $i(m)$ and $u(c)$. If both utility and production functions are of the constant elasticity type, the income and price elasticities are 1 and -1. Alternatively as in Hall and Jones, if $\lambda_y > 0$, for example if the production function exhibits constant elasticity while the utility function exhibits decreasing elasticity. In the case we consider both utility and production exhibit decreasing elasticity such that the curvature of the production function, taking the curvature of the utility function as given, is identified by calibrating the price and income elasticities to existing evidence on their magnitude (for example, the HIE).
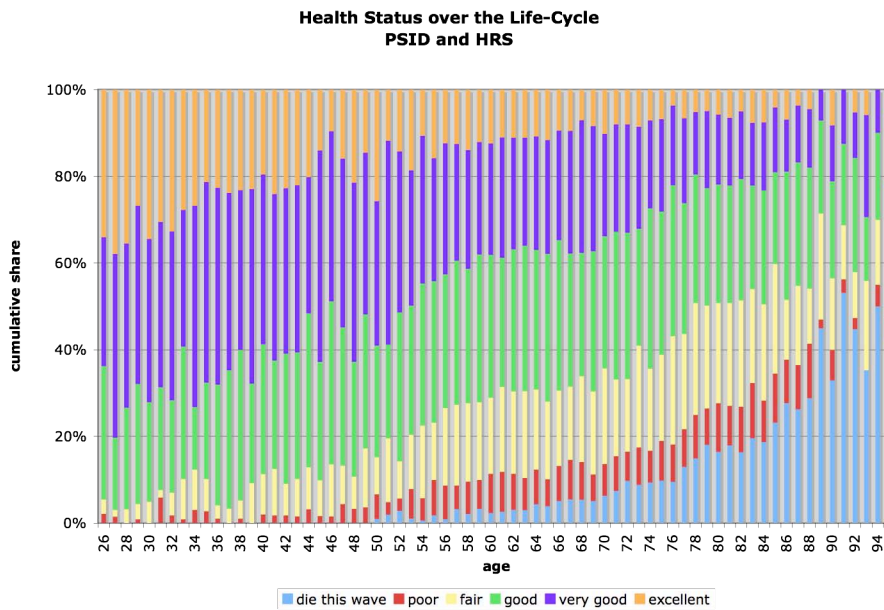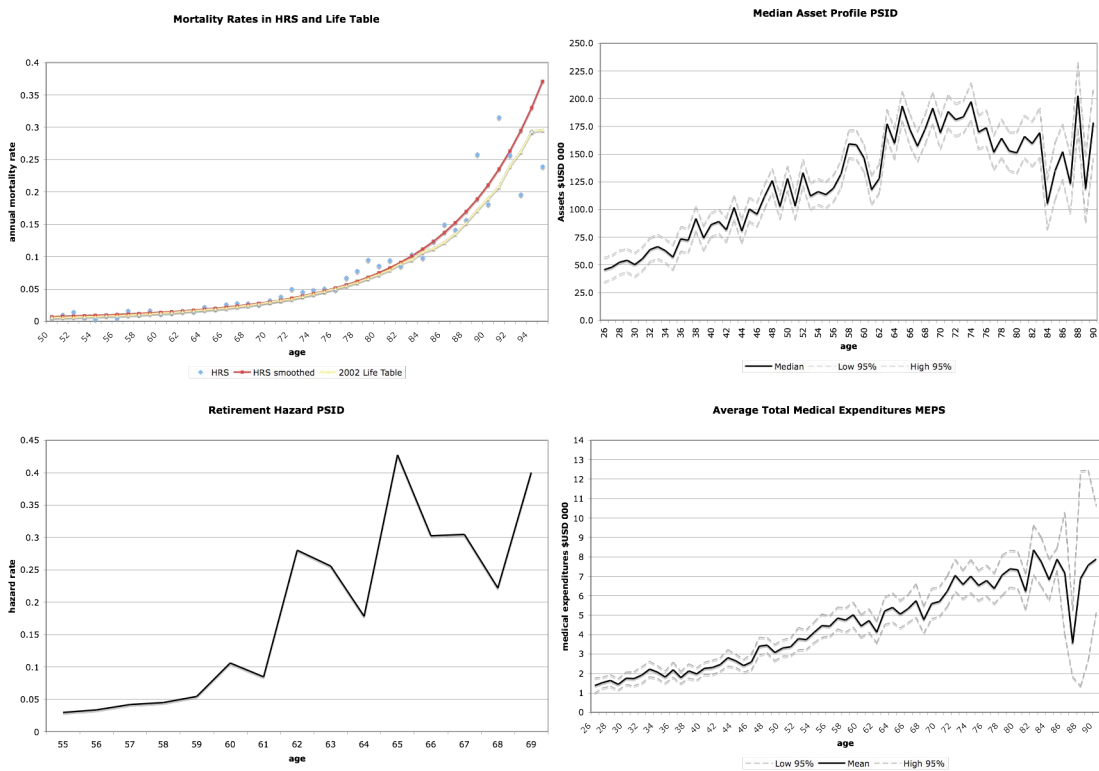
Figures

Figure 1a Employment and Insurance Status

Employment and Health Insurance
PSID
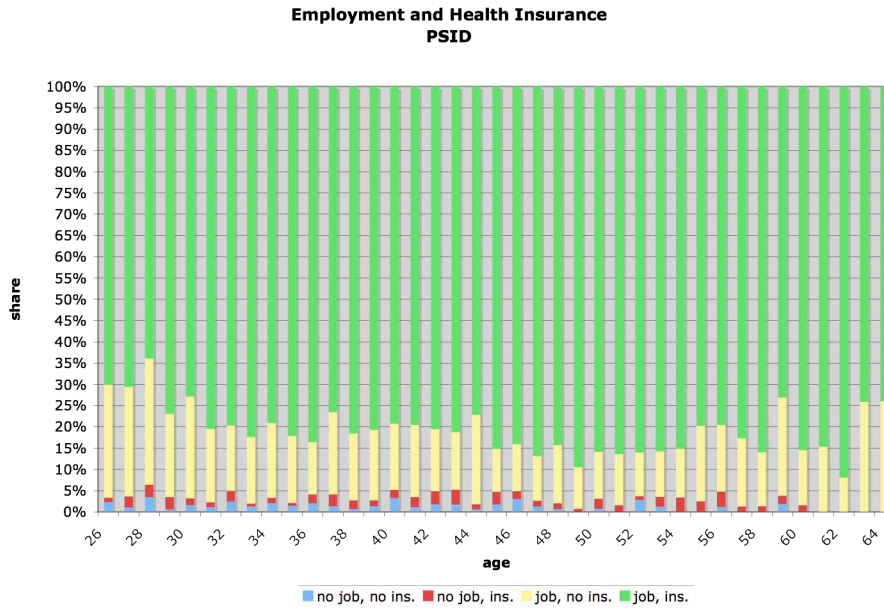
Notes:.raw data from 1999 to 2005 waves of PSID.

Figure 1b Health Status

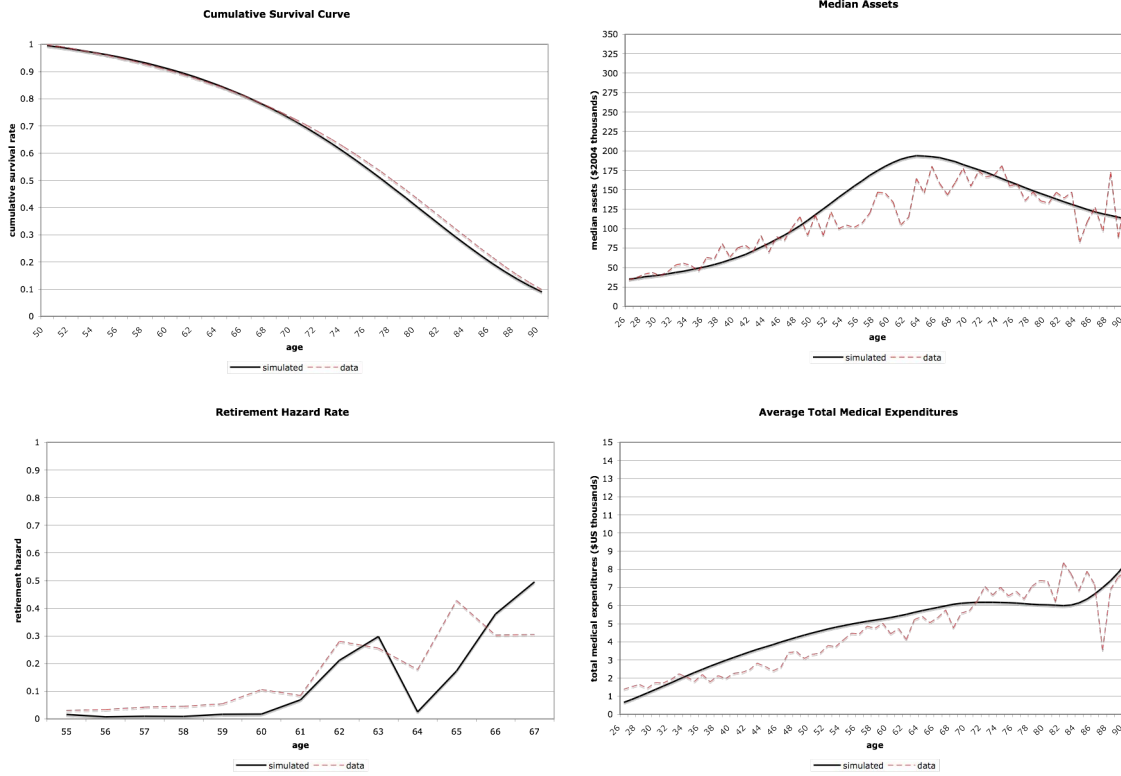**Health Status over the Life-Cycle**
PSID and HRS

Notes: raw data from 1999 to 2005 waves of PSID for those younger than 50 and HRS 1992-2004 for those older than 50.

Figure 2 Moments for Calibration

Notes: The figures present the moments from the data used to calibrate parameters. See the discussion in section 4.4.
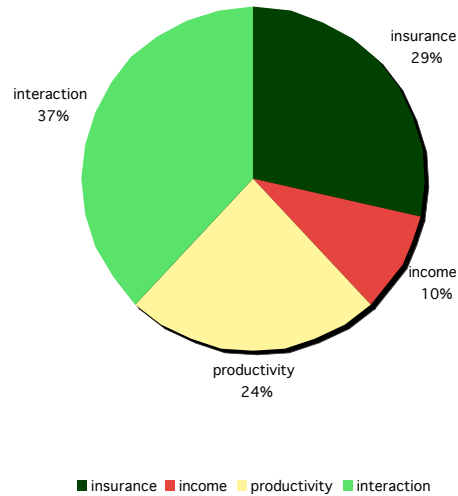
Figure 3 Goodness-Fit of Model Simulations

Notes: Simulated and data profiles based on parameters presented in Table 1.

Figure 4 Increase in Health Spending

Share of Increase in Health Spending 1965-2005

■ insurance ■ income ■ productivity ■ interaction

Notes: Based on simulation output from Table 4. Insurance refers to the induced change brought by an increase in the generosity of health insurance keeping income and productivity at the 1965 level. Income refers to the induced change brought by the increase in real earnings and the generosity of Social Security keeping insurance and productivity at the 1965 level. Productivity refers to the induced change brought by improvements in productivity keeping income and insurance at the 1965 level. Finally, interaction refers to the residual induced change brought by the interaction of income, insurance and productivity change when occurring simultaneously.

Figure 5 Increase in Life Expectancy

Share of Increase in Life Expectancy at age 50, 1965-2005
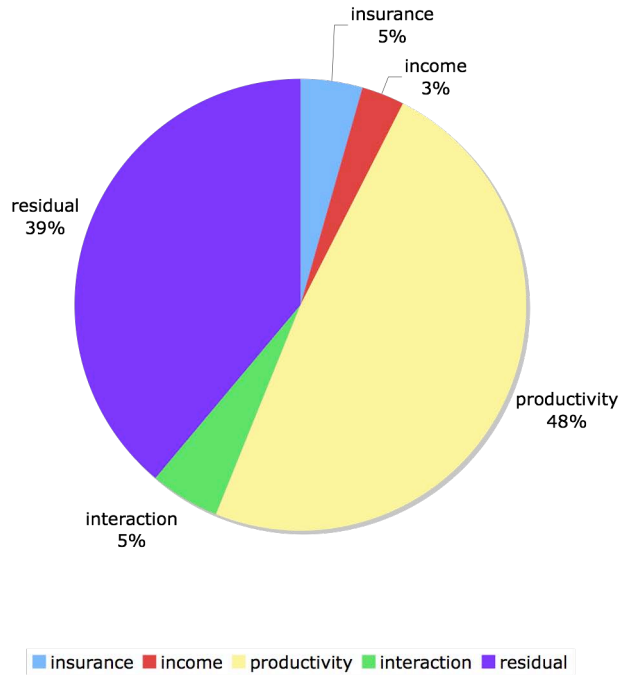
Notes: Based on simulation output from Table 4. Insurance refers to the induced change brought up by an increase in the generosity of health insurance keeping income and productivity at the 1965 level. Income refers to the induced change brought by the increase in real earnings and the generosity of Social Security keeping insurance and productivity at the 1965 level. Productivity refers to the induced change brought by improvements in productivity keeping income and insurance at the 1965 level. Finally, interaction refers to the residual induced change brought by the interaction of income, insurance and productivity change when occurring simultaneously. Since the simulated change in life expectancy is less than the observed relative share. There is another residual category which captures the unexplained component of the increase in life expectancy in the data.

Tables

Table 1 Structural Parameters Chosen by Calibration

Parameter	Definition	Value
σ	Coefficient of relative risk aversion	2.6
ψ_h	Utility benefit if health status >2	0.035
ψ_0	Baseline log utility benefit of retirement	4.65e-03
ψ_a	Incremental log utility benefit of retirement by age	6.15e-04
θ_b	Strength of bequest motive	0.05
θ_{00}	Baseline log productivity (age 25, good health)	-3.352
θ_{0h}	% increment in productivity fair/poor health	1
θ_{0t}	% change in productivity with age	0.02
θ_1	Concavity of production function	2.1
τ	Real interest rate	0.03
β	Discount factor	0.95

Notes: See discussion in section 4.5 for the justification.

Table 2 Price and Income Effects on Medical Expenditures in 2005 Baseline Scenario

Age Group	Co-Pay (0.2 to 0.5)	Permanent Income (+10%)	Transitory Income (+10%)
25-35	-0.345	1.564	0.631
35-45	-0.281	0.888	0.576
45-55	-0.267	0.909	0.583

Notes: Simulation based on calibration in Table 1. The first two columns consider permanent changes. In the first column, we increase the co-payment from 0.2 to 0.5 across all ages for those insured and look at the resulting change in total medical expenditures. In the second column, we increase earnings by 10% over the entire working life. The reported elasticity is the average elasticity within the age group. In the last column, we increase earnings at the mid point in the age interval (30 for the first, 40 for the second and 50 for the last age group).

Table 3 Simulated Outcomes in 1965 and 2005 Scenarios

	Simulation results	
Outcomes	1965	2005
Total Medical Expenditures ($k)		
25-35	0.131	1.376
35-45	0.563	3.210
45-55	1.020	4.535
55-65	1.172	5.363
65-75	1.039	6.149
75-85	0.740	6.052
85+	0.945	7.585
Total ($k)	0.742	4.183
net income ($k)	24.030	28.221
share of net income	0.031	0.148
Average retirement age (years)	64.2	63.9
Life expectancy at age 50 (years)	23.88	27.65

Notes: simulated results in 2005 and 1965 using the insurance, income and technology changes. Amounts in thousands of $US 2004.

Table 4 Simulated Incremental Changes Starting from 1965 Scenario

			Simulation		
Outcomes	1965	increase generosity insurance	increase real earnings and social security generosity	increase productivity	2005
Avg. Medical Expenditures($k)					
Age 25-35	0.131	0.397	0.235	0.398	1.376
35-45	0.563	1.355	0.816	1.201	3.210
45-55	1.020	2.265	1.418	1.941	4.535
55-65	1.172	2.538	1.614	2.233	5.363
65-75	1.039	2.494	1.530	2.198	6.149
75-85	0.740	1.925	1.197	1.887	6.052
85+	0.945	2.378	1.620	2.302	7.585
Avg Medical Expenditures($k)	0.742	1.725	1.071	1.563	4.183
Net income ($k)	24.030	23.991	27.985	23.760	28.221
Share of net income	0.031	0.072	0.038	0.066	0.148
Average retirement age (years)	64.2	64.3	61.9	65.1	63.9
Life expectancy at age 50 (years)	23.88	24.16	24.07	26.86	27.65

Notes: Simulated outcomes. First column represents scenario in 1965. Column 2 to 4 introduce changes one at a time starting from 1965. Finally, the last column presents estimates in 2005 (when all changes are introduced at once). All amounts in thousands $US 2004.