NBER WORKING PAPER SERIES

THE LIFETIME COSTS OF BAD HEALTH

Mariacristina De Nardi
Svetlana Pashchenko
Ponpoje Porapakkarm

Working Paper 23963
http://www.nber.org/papers/w23963

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2017, Revised October 2018

We thank Liran Einav, Christopher Flinn, Eric French, Joseph Altonji, Selahattin Imrohoroglu, John Kennan, Naoki Aizawa, Sagiri Kitao, Rasmus Lentz, Vincenzo Quadrini, Yongseok Shin, Matthew White, and all seminar participants at the Keio University, University of Hong Kong, Chinese University of Hong Kong, University of Tokyo, University of Delaware, University of Connecticut, FED Chicago, FED Atlanta, Institute for Fiscal Studies, AMES2017, SED2017, EMES2017, SAET2018 for their comments and suggestions. De Nardi gratefully acknowledges support from the ERC, grant 614328 "Savings and Risks." Porapakkarm gratefully acknowledges support from the JSPS KAKENHI Grant Number 15K03505. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research, the CEPR, any agency of the federal government, or the Federal Reserve Bank of Chicago.

The Lifetime Costs of Bad Health
Mariacristina De Nardi, Svetlana Pashchenko, and Ponpoje Porapakkarm
NBER Working Paper No. 23963
October 2017, Revised October 2018
JEL No. E21,H31,I14

ABSTRACT

What are the cumulative effects of health shocks over the life-cycle? The answer depends on the nature of persistence of bad health and the extent to which it changes individuals' economic circumstances. We measure the lifetime costs of bad health using a rich structural model that can reproduce the short-and long-run dynamics of health and health-related inequality in economic outcomes. We explicitly take into account the fact that this inequality can be partly traced back to factors predetermined early in life. Our model has several implications concerning the pecuniary and non-pecuniary effects of health shocks. The life- time costs of bad health are influenced by ex-ante predetermined factors, and the largest component of pecuniary costs is the loss in labor earnings. Because bad health lowers ones' life expectancy, its non-pecuniary effects are large. In particular, through this channel they are very important for individuals' welfare and non-trivially contribute toward lifetime inequality.

Mariacristina De Nardi
Federal Reserve Bank of Minneapolis
90 Hennepin Ave
Minneapolis, MN 55401
and University College London
and also NBER
denardim@nber.org

Svetlana Pashchenko
University of Georgia
svetlana@uga.edu

Ponpoje Porapakkarm
GRIPS
p-porapakkarm@grips.ac.jp

1 Introduction

How important is health risk over the life-cycle? Bad health can negatively affect individuals through multiple pathways. When markets are incomplete this can translate into significant disparity in economic outcomes and welfare, especially when bad health is persistent.

Several features of the data suggest that health-related inequality in economic outcomes is substantial. First, in additon to higher medical spending, unhealthy people have significantly lower income than healthy people (*income-health gradient*), due to their lower labor supply and lower earnings conditional on working. For instance, among prime age men with a high school degree (ages 45 to 55), participation among the healthy is over 90% while participation among the unhealthy is around 70%. In addition, conditional on participation the healthy earn on average 28% more than the unhealthy.[1]

Second, unhealthy people tend to accumulate substantially less wealth than healthy people (*wealth-health gradient*). The gap in wealth by health starts at relatively young ages and becomes very large by retirement time. For instance, among 65 year old males with a high school degree, the median wealth of the healthy is almost twice that of the unhealthy; \$230,000 for the former versus \$120,000 for the latter (in 2015 dollars).[2] This suggests that the accumulated effects of bad health can be important.

In this paper, we develop a structural framework for measuring the long-run effects of bad health. Our model includes several channels through which bad health can affect individuals: lower productivity, higher disutility from work and medical expenses, and lower survival probability. In addition, the association between health and economic outcomes can be linked to ex-ante heterogeneity across people. Given the goal of our paper, we take particular care that our model can capture not only the cross-sectional but also short- and long-run dynamic aspects of the data, specifically those related to (i) the long-term dynamics of health, and (ii) the joint evolution of health and wealth.

The long-run effects of bad health crucially depend on how persistent bad health is and where this persistence comes from. The data show that the dynamics of health are complex and not consistent with a low-order Markov process. More specifically, health transitions display strong duration dependence in recovery probability: the probability of moving from bad to good health declines monotonically with the number of years that an individual

[1] Own calculations from the Panel Study of Income Dynamics. Individuals are classified into healthy and unhealthy categories based on self-reported health. Details are given in Section 5.1.5.

[2] Own calculations, based on the Health and Retirement Study dataset (HRS). Wealth is total net worth after controlling for family sizes. Details are in Section 5.2. See also Smith (1999) who documents the large disparities in wealth between the healthy and unhealthy in the HRS data; and Poterba et al. (2017) who show that, in the same data set, individuals' health status between the ages of 51 and 61 has a significant impact on the subsequent evolution of their assets.

has been unhealthy. We estimate a parametric model of health shocks that allows for both history-dependence and fixed ex-ante heterogeneity, and that matches the described patterns. Both history-dependence and ex-ante heterogeneity can generate persistence in health, but distinguishing among them allows us to better understand what account for long episodes of bad health: bad luck or (ex-ante) permanent differences across individuals.

People who draw bad realizations from a persistent health process save less because they have less net income and face a shorter life expectancy. However, the gap in wealth between the healthy and the unhealthy far exceeds that implied by the difference in earning, medical spending, and life expectancy.[3] This suggests a systematic difference in saving propensities between the two groups. To capture this in our life-cycle model, we allow the aforementioned fixed heterogeneity in health process to be correlated with the rate of time preferences, i.e., individuals who are predisposed to experience long episodes of bad health can also be less patient. We estimate this correlation in our structural model using the Method of Simulated Moments. This can be viewed as the first attempt to merge together two strands of literature. First is macroeconomic studies that point out the importance of heterogeneity in rates of time preferences to explain wealth inequality (Carroll et al., 2016, Hendricks, 2007a, 2007b; Krueger et al., 2016; Krusell and Smith, 1998, Samwick, 1998). Second is a growing microeconomic literature showing the important role of genetic endowments and childhood circumstances in determining saving behavior (Cronqvist and Seigel, 2015; Barth et al., 2017) and health outcomes in adulthood (Case et al., 2002, 2005, 2010)

We estimate our model using three datasets: the Health and Retirement Study, the Panel Study of Income Dynamics, and the Medical Expenditure Panel Survey. We focus on a relatively homogeneous group of high school men to avoid the confounding effect of education and gender on health and economic outcomes.

Our estimated model is consistent with three sets of important facts. First, it captures the dynamics of health, including its duration dependence. Second, it matches the observed impact of bad health on earnings and labor supply (income-health gradient), medical spending, and life expectancy. Finally, and importantly, it also captures the wealth-health gradient by matching the large difference in wealth *levels* between the healthy and unhealthy over the lifespan, and the disparity in wealth *changes* among people with different number of years spent being unhealthy.[4]

[3] This was pointed out by Poterba et al. (2010) in a reduced form analysis; we confirm this finding in a structural framework.

[4] The literature commonly documents the wealth-health gradient as a large difference in wealth levels between healthy and unhealthy individuals after controlling for observables, e.g., age, education, etc. We add to this an additional dynamic aspect of the gradient: the negative relationship between wealth *changes* and the number of years spent being unhealthy.

Our results can be summarized as follows. First, both fixed ex-ante heterogeneity and history-dependence are important in driving health dynamics but they play a different role in how individuals get sick versus how they recover. More specifically, the persistence of bad health is mostly generated by fixed ex-ante heterogeneity while the persistence of good health is mostly due to history-dependence. As a result, long episodes of bad health are concentrated among individuals with a certain (fixed) health type.

Second, our estimates imply a strong correlation between one's health type and rate of time preferences; among the long-term unhealthy a larger fraction is less patient and has a lower propensity to save. This is important for accounting for the wealth-health gradient: when the correlation between patience and health type is shut down, the model significantly underpredicts the wealth gap between the healthy and unhealthy even though it matches the income-health gradient.

Third, the monetary costs of bad health are very concentrated and highly unequally distributed across health types. Our measure of these costs includes both direct (out-of-pocket medical spending) and indirect (loss in labor earnings) costs. We find that the latter component is the largest contributor to the lifetime costs of bad health and arises because unhealthy individuals are less productive and work less than healthy ones. In addition, even though total medical costs are substantial for the long-term unhealthy, the effects of out-of-pocket costs are much smaller due to health insurance coverage.

Fourth, to capture both the monetary and non-monetary effects of health, we evaluate people's willingness to pay to increase the probability of being healthy next period. We find that, overall, individuals are willing to pay about 5% of average income to increase the probability of being healthy by one percentage point. Our decomposition exercise shows that good health is valuable mainly because it extends life expectancy and increases labor earnings. The first channel accounts for 60% and the second for 36% of the computed willingness to pay.

Finally, we ask how much bad health contributes to lifetime inequality measured as variation in lifetime utility. We find that bad health can explain up to 47 percent of the variation in lifetime utilities. The main mechanism behind this result is that the variation in the length of life due to health shocks creates substantial variation in lifetime utility when life is valuable. Thus, ignoring this mechanism will significantly underestimate the effect of health uncertainty on lifetime inequality. This finding compliments the previous studies on lifetime inequality that mainly focus on the monetary aspect of inequality, thus abstracting from the non-pecuniary effect of longevity.[5]

Our study has several contributions. First, we document a number of new facts related

[5] See, for example, Huggett et al., 2010; Keane and Wolpin, 1997; and Storesletten et al., 2004.

to the long-run dynamics of health and use them as additional moments when estimating our health shock process. Our estimates offer a new insight about the nature of health persistence and the asymmetric dynamics of healthy and unhealthy states.

Second, we develop a quantitative model specifically designed to gauge the long-run or accumulated effects of bad health.[6] To the best of our knowledge, the accumulated effects of bad health have not been assessed before due to the lack of data or an appropriate structural framework.

Third, we use our model to evaluate the effects of health uncertainty from three distinct perspectives: monetary losses, individuals' valuations, and lifetime inequality. While previous studies focus on only one of these perspectives, we are able to unite them into one framework.[7] Moreover, our unified framework makes it possible to quantify the relative importance of various channels through which health affects individuals over the life course.

Forth, even though the negative relationship between health and economic outcomes is well-documented (see Cutler et al., 2011 for a review), the mechanisms behind it are still not well-understood. We show that this relationship, both in cross-section and over time, can be well accounted for in a life-cycle model once taking into account the fact that some factors determined early in life or at birth can influence both health and economic outcomes in adulthood. This serves as an important building block toward further analyzing health-related inequality through the lens of a structural model.[8]

The rest of the paper is organized as follows. We summarize the data used in our study in Section 2. Section 3 documents empirical facts related to health dynamics and estimates the health shock process. Section 4 introduces our life-cycle model and Section 5 describes its estimation. We present the results and conclusions in Section 6 and Section 7, respectively.

[6] A number of studies use a life-cycle model with health uncertainty, but not focusing on its accumulated effects. For example, Ameriks et al. (2017), De Nardi et al. (2010, 2016), Lockwood (2014), Nakajima and Telyukova (2011) focus on the role of medical spending after retirement. Cornesa et al. (2018), Kitao and Jeske (2009), Pashchenko and Porapakkarm (2017) analyze various aspects of health insurance system. Close to ours are French (2005) who focuses on labor supply and retirement decisions, and Capatina (2005) who quantifies the effects of health uncertainty on income inequality and precautionary saving.

[7] Dobkin et al. (2016), Parro and Pohl (2017), Lundbord et al. (2015), Blundell et al. (2016) show that health shocks affect earnings or employment while Capatina et al. (2017) study the interaction between health and human capital accumulation, thus focusing on monetary consequence of health deterioration. Morphy and Topel (2006), Hall and Jones (2007) emphasize the effects of health on longevity, thus focusing on its non-monetary aspect.

[8] Health-related inequality can also be studied in a structural framework where the evolution of health is endogenous. This framework has been recently used in a number of studies to examine the effects of health insurance policies (e.g. Cole et al., 2018; Ozkan, 2014). These studies typically abstract from ex-ante heterogeneity across individuals that potentially have a long-run effect on health; thus, the disparity in health outcomes mainly result from realized health shocks and economic circumstances during adulthood. We propose an alternative structure which takes into account the fact that people are ex-ante different, e.g. due to genetic endowments and/or childhood circumstances, and this heterogeneity can contribute toward the joint evolution of health, income, and wealth in adulthood.

2 Data for the study

The ideal dataset to measure the lifetime effects of bad health is a lifelong panel that tracks a large number of individuals starting from a young age and until their death and that contains information on their health, total and out-of-pocket medical spending, health insurance status, labor earnings, labor supply, and wealth. However, a dataset of this kind does not exist for the U.S.

To obtain the best possible measurement, we use three different datasets to estimate our health shock process and our life-cycle model: the Panel Study of Income Dynamics (PSID), the Health and Retirement Study (HRS), and the Medical Expenditure Panel Survey (MEPS).

The PSID is a nationally representative panel that surveys individuals and their families. It started in 1968, on an annual basis, and has been administered bi-annually since 1997. The PSID tracks individuals over a long period of time but the number of individuals is relatively small and does not contain information on all of the variables that we need. We use the PSID to construct the moments for health, labor supply, labor earnings, and wealth that are the key targets in our model estimation. For health, labor supply, and earnings we use the 1984-1997 waves because in our model, a period is one year. The PSID collected wealth information every five years before 1997 and every two years after that. To construct the wealth moments, we use the 1994 and 1999-2011 waves.

The MEPS is a nationally representative survey of households that focuses on measuring medical spending and health insurance. It contains individuals of all ages but age is top-coded at age 85 and has a short panel dimension: each individuals is interviewed at most five times over a two-year period. The medical spending reported in MEPS is cross-checked with insurers and providers, which improves their accuracy.[9] We use waves 1998-2012 of MEPS to estimate medical spending and parameters related to health insurance.

The HRS is a bi-annual panel that surveys a nationally representative sample of individuals over the age of 50. The advantage of the HRS over the PSID is a larger number of older individuals. We use the RAND Version O (waves 1994-2012) of this data set to estimate the health-dependent survival probabilities. In addition, we use the HRS to construct several additional moments related to health and wealth for the external validation of our estimated model.

For each dataset, we use a sample of male household heads with education at the high school level. We normalize all nominal variables to the base year (1996) using the Consumer Price Index (CPI).

[9] Pashchenko and Porapakkarm (2016b) provide more details on the MEPS dataset.

3 Health dynamics estimation

We first document the cross-sectional and dynamic moments of self-reported health status of high school males in the PSID data. We then estimate a model for health dynamics that is consistent with these moments and discuss its implications.

We use self-reported health for two reasons. First, this variable is available in all three datasets that we use and is consistently measured across them.[10] Second, a number of studies find that self-reported health is highly correlated with other subjective and objective measures of health and also has significant explanatory power in predicting future mortality, even after controlling for many other factors (See Idler and Benyamini (1997) for a review, and van Doorsaler and Gerdtham (2002), and Pijoan-Mas and Ríos-Rull (2014) for a more recent examination). In addition, Poterba et al. (2017) use a principle component analysis to construct a continuous single measure of health index from the HRS and find that the weights on subjective health measures are relatively high and the highest weight is assigned to the self-reported health variable.

3.1 Data patterns

We construct our measure of health as follows. In the PSID (and the HRS), individuals are asked to rank their health as *excellent*, *very good*, *good*, *fair* or *poor*.[11] We aggregate these answers into a binary measure of health: individuals who report their health to be in the first three categories are classified as healthy or in good health, while individuals who report being in fair or poor health are classified as unhealthy or in bad health.[12]

The top panel of Figure (1) displays the percentage of unhealthy individuals by five-year age brackets. The dots correspond to the statistics constructed from the PSID while the crosses refer to the statistics from the HRS. The percentage of unhealthy individuals over the age of 50 computed from the HRS is similar to that computed from the PSID. The bottom panel of Figure (1) displays the health transition probabilities between two consecutive years by five-year age bracket.[13] These figures show that conditional on survival, as people age, they are more likely to become unhealthy and less likely to recover from bad health.

[10] The top panel of Figure 1 shows that for individuals over 50, which is the age group observed in both PSID and HRS, the measure of self-reported health is consistent in these two datasets. Attanasio et al. (2011) compare this variable in the HRS and MEPS, and document that the two datasets are consistent.

[11] There are 2,368 individuals without missing self-reported health status in the PSID, or 19,503 individual-year observations. On average individuals are observed for 8 consecutive years.

[12] This classification is common in the literature. See, for example, French (2005), Capatina (2015), Rust and Phelan (1997).

[13] These transition probabilities were constructed as follows. Denote health status of individual i at age t as h_{it}. For a group of individuals aged 20 to 24, the probability of moving to good health conditional on currently being in bad health can be expressed as

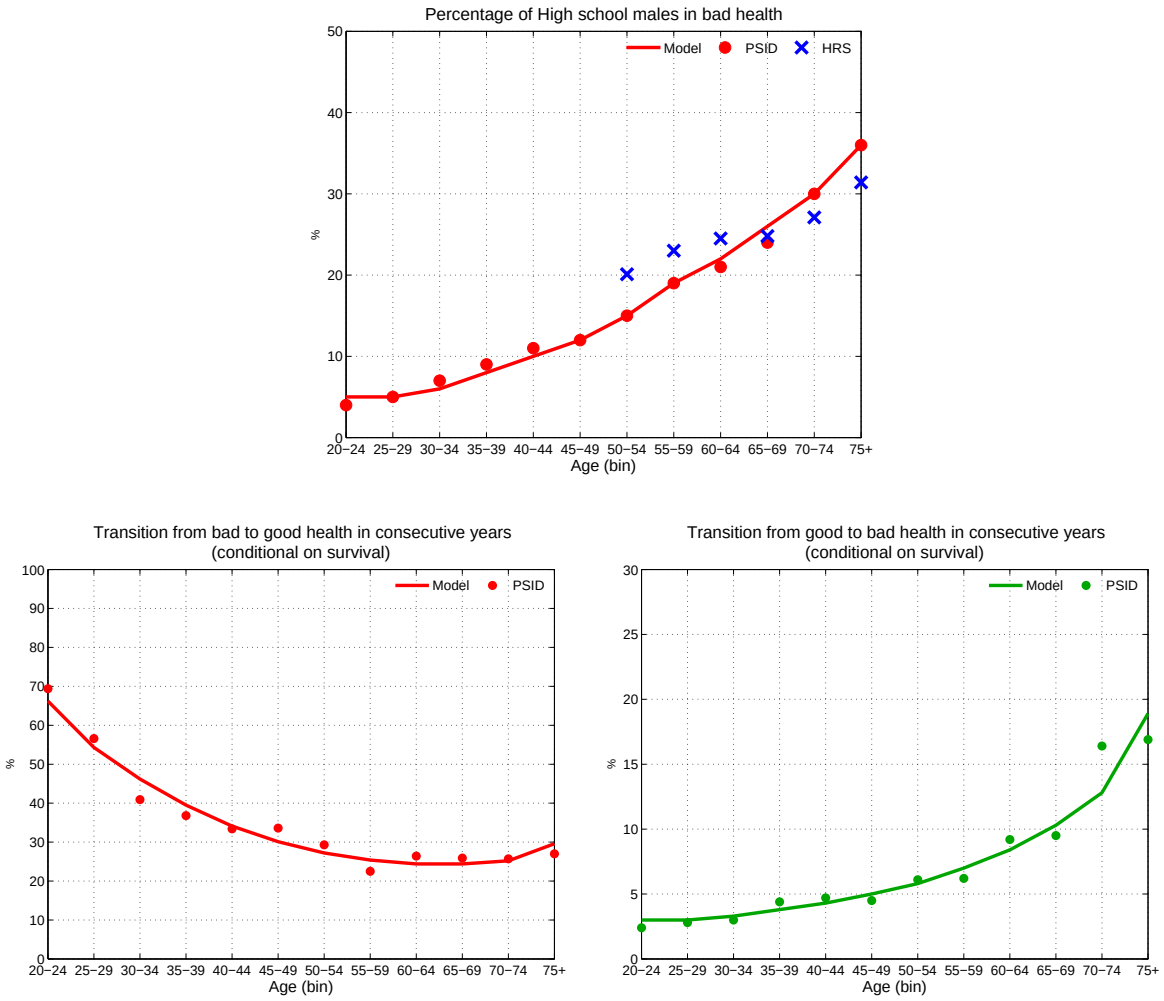


Figure 1: Moments related to health status. Top panel: percentage of individuals in bad health by age. Bottom left panel: percentage of individuals moving from bad to good health. Bottom right panel: percentage of individuals moving from good to bad health. (Dots: PSID. Crosses: HRS. Solid lines: model.)

To better understand the dynamics of health, we next analyze how the transition probabilities to good and bad health depend on the duration of the current health status. Specifically, we compute the transition probability of moving to good (bad) health conditional on being in bad (good) health for at least τ consecutive years.[14] Due to the small sample size we group observations into three larger age groups: 30-54, 55-69, and older than 70. Figure

$$\frac{\sum_{t=20}^{24}\sum_{i}\mathbf{1}\left(h_{it}=B\cap h_{it+1}=G\right)}{\sum_{t=20}^{24}\sum_{i}\mathbf{1}\left(h_{it}=B\cap h_{it+1}=\{B,G\}\right)},$$

where $\mathbf{1}(\cdot)$ is the index function equal to one if its argument is true; otherwise it is zero.

[14] Denote the sequence of health status in the past τ years up to age t as h_{it}^{τ}. For age group 30-54, we compute the probability of moving to good health conditional on being unhealthy for at least τ consecutive

(2) plots (in shaded bars) the resulting duration-dependent transition probabilities from bad to good health (top panel) and from good to bad health (bottom panel).

A key feature of the probability of recovering from bad health is that it declines monotonically with duration: the longer an individual has been unhealthy, the less likely he is to become healthy, and this pattern holds for all age groups.[15] It is important to note that this decline cannot be captured by the low-order Markov process for health that is commonly used in the literature (e.g., French, 2005; French and Jones, 2011; and Capatina, 2015). For example, a first order Markov process implies that the transition probability does not depend on how long one has been in bad health, while a second order Markov process would imply that this probability is the same for durations longer or equal to two years. In the next section, we discuss how this observation motivates our parametrization of the health process.

In contrast to the transition from bad to good health, the transition from good to bad health does not display noticeable duration dependence, especially at younger ages, as can be seen in the bottom panel of Figure (2). More specifically, there is a noticeable difference between the probability of moving into bad health after having been healthy for at least one and two years, but after that the probability profile is rather flat. In other words, individuals who are healthy for two years have almost the same probability of becoming sick compared to individuals who are healthy for more than two years. This lack of duration dependence suggests that the probability of becoming sick can be well described by a low-order Markov process.

3.2 Health process specification and estimation

The negative duration dependence in the probability of recovering from bad health shown in the top panel of Figure (2) can be generated by two different mechanisms. First, the effects of bad health can be compounding, i.e., individuals who stay sick for a long period of time might have a smaller recovery probability than those who are sick for a short period of time.

years as follows:

$$\frac{\sum_{t=30}^{54}\sum_{i}\mathbf{1}\left(h_{it}^{\tau}=B\cap h_{it+1}=G\right)}{\sum_{t=30}^{54}\sum_{i}\mathbf{1}\left(h_{it}^{\tau}=B\cap h_{it+1}=\{B,G\}\right)}.$$

[15] This negative duration dependence is a robust pattern even when we exclude those ever receiving Social Security Disability Insurance or when we use smaller age groups, for example, based on a 10-year age bracket. As an additional robustness check, we also compute the transition probability from bad to good health, where we include in the bad health category only people who report their health being fair, thus excluding individuals with poor health (the worst self-reported health status) who are less likely to recover. The declining pattern still holds when using this more homogeneous measure of bad health.

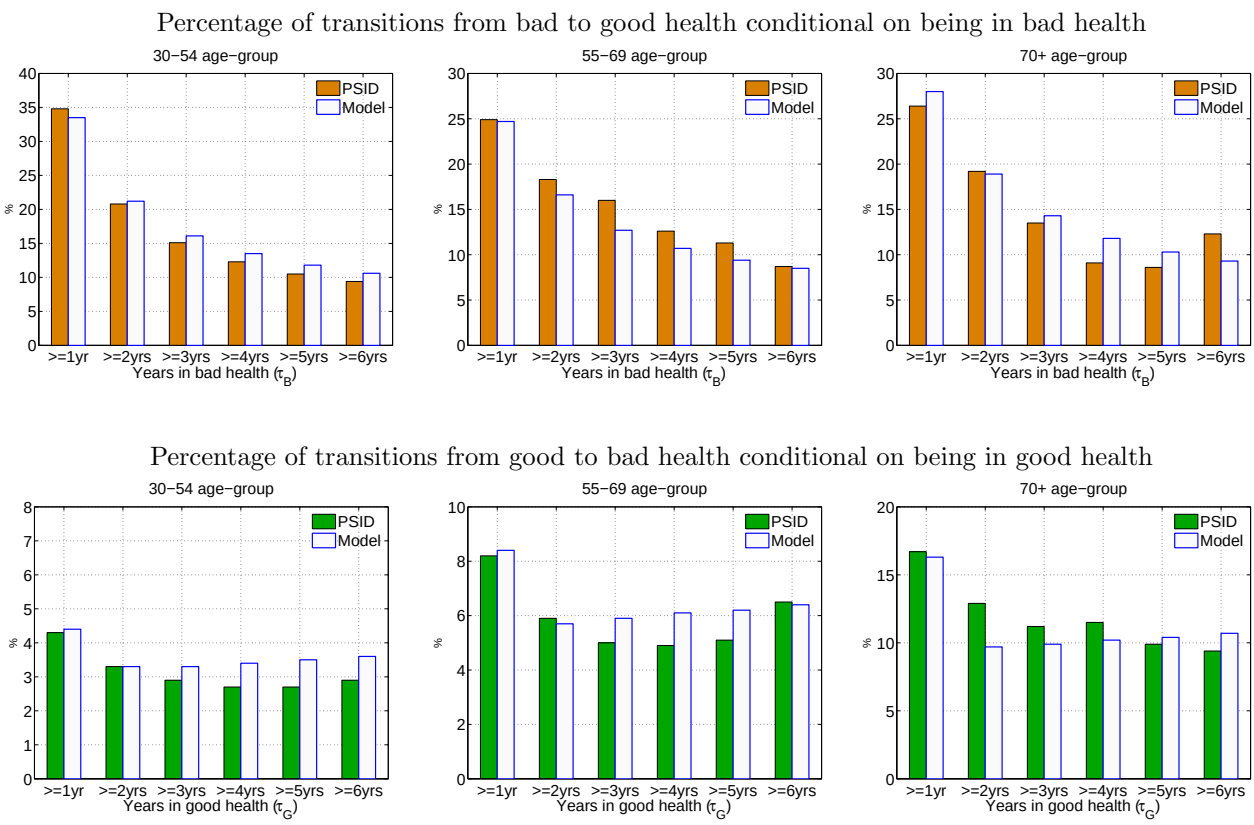


Figure 2: Dynamics of health conditional on duration.

This mechanism is consistent with a high-order Markov process. Second, individuals may differ in terms of their ability to recover, i.e., some individuals have lower recovery probability than others. In the latter case, people who are more likely to recover move out of the bad health state faster, hence the pool of the long-term unhealthy is predominantly composed of individuals who are inherently less likely to recover. The latter mechanism is consistent with fixed heterogeneity in health transition probabilities.

We choose our model for health dynamics based on two criteria. First, the model must capture the cross-sectional and dynamic moments of health that we document. Second, the model must be parsimonious, so that a structural life-cycle model augmented with this health shock process is computationally manageable. Based on these criteria, we formulate our health shock process as a second-order Markov process with fixed heterogeneity. Specifically, the probability of being in good health at age $t+1$ conditional on surviving to age $t+1$ and being in bad health for τ_B years, denoted $\pi_{it}^G(\tau_B)$, is formulated as the following logit function:

$$logit\Big(\pi_{it}^G(\tau_B)\Big) = \Big(a_1\mathbf{1}(\tau_B = 1) + a_2\mathbf{1}(\tau_B \geq 2)\Big) + \Big(b_1 t + b_2 t^2\Big) + \eta_i. \tag{1}$$

The first bracket is a second-order Markov process, the second bracket is a second-degree polynomial in age, and η_i is the fixed heterogeneity or *health type*.[16] We assume that η_i is uniformly distributed over five discrete points that are symmetric around zero, i.e., there are five distinct health types. Note that an individual with low η_i has a lower probability of recovering.

In a similar fashion, we model the probability of being in bad health at age $t+1$ conditional on surviving to age $t+1$ and being in good health for τ_G years, denoted $\pi_{it}^B(\tau_G)$, as follows:

$$logit\Big(\pi_{it}^B(\tau_G)\Big) = \Big(a_3\mathbf{1}(\tau_G = 1) + a_4\mathbf{1}(\tau_G \geq 2)\Big) + \Big(b_3 t + b_4 t^2\Big) + b_5\eta_i. \tag{2}$$

We allow the health type to have a different effect on the probabilities of getting sick and recovering by introducing the coefficient b_5 in Equation (2). It should be noted that our specification nests the first-order Markov model of health shock commonly used in the existing literature; this requires the following restrictions on the coefficients: $a_1 = a_2, a_3 = a_4$, and $\eta_i = 0$.

We use the Method of Simulated Moments to estimate our health shock process and target the moments documented in Figures (1) and (2). The transition probabilities in Equations (1) and (2) and the targeted moments are conditional on surviving from age t to $t+1$; so we first estimate the health-dependent survival probabilities by age. Since the sample size of the elderly in the PSID is small, we use the data on males with a high school degree from the HRS (1994-2012) to estimate a probit model of two-year survival probabilities as a function of a cubic polynomial of age interacted with the dummy variable of the current health status.[17] The one-year survival probability is computed as the square root of the estimated two-year survival probability. Since the sample in the HRS is older than 50, we use our estimated probit model to predict the survival probability for the younger age groups. Figure 3 shows our estimated one-year survival probabilities conditional on the current health status.

Given our estimated survival probabilities and parameter values $\boldsymbol{\theta_H} = \{a_{1-4}, b_{1-5}, \eta_{1-5}\}$ for Equations (1)-(2), we can simulate the realized health status over the life-cycle for a large number of individuals. The initial distribution of health status is taken from a sample of people age 19-22 in the PSID, where we assume that the initial health status is orthogonal

[16] The proposed specification is similar to a proportional hazard model commonly used in survival models, where the first bracket is a baseline hazard function.

[17] We do not allow one's health type to affect one's survival probability directly, but there is an indirect effect through the evolution of health. If we were to allow for a direct effect, our estimated life-cycle model would imply that one's'wealth could predict his immediate survival probability, even after controlling for his current health status. This happens because our model implies a strong correlation between wealth and health type (as will be discussed in more details in Section 5.2.2). This is not true in the data: Pijoan-Mas and Ríos-Rull (2014) find using the HRS that after controlling for the current self-assessed health, the effects of education, wealth, and income on the next two-year mortality rate are very small.

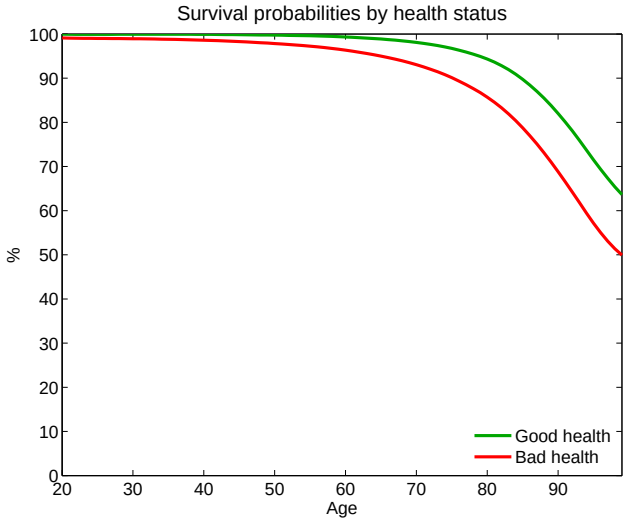


Figure 3: Estimated health-dependent survival probabilities.

to one's health type η_i.[18]

Our algorithm searches for the parameters $\boldsymbol{\theta_H}$ that minimize the following function:[19]

$$\min_{\boldsymbol{\theta_H}} \left(\mathcal{M}_H^D - \mathcal{M}_H^S\left(\boldsymbol{\theta_H}\right)\right)' \left(\mathcal{M}_H^D - \mathcal{M}_H^S\left(\boldsymbol{\theta_H}\right)\right), \tag{3}$$

where $\mathcal{M}_H^D$ and $\mathcal{M}_H^S$ are the vectors of the targeted moments from the PSID and the simulated data, respectively. The targeted moments in our estimation are listed below.

- The percentage of unhealthy individuals in each five-year age group, as shown in the top panel of Figure (1) (12 moments).
- The health transition probabilities between two consecutive years for each five-year age group, as shown in the bottom panel of Figure (1) (24 moments).
- The duration-dependent profiles of the transition probabilities, as shown in Figure (2) (36 moments).

The identification of $\boldsymbol{\theta_H}$ is straightforward, given the relatively simple specification of our health shock process. The percentage of unhealthy individuals and the age-dependent transition probabilities in Figure (1) help pin down the age-dependent coefficients $\{b_1, b_2, b_3, b_4\}$. As discussed in the previous subsection, our Markov process of order two implies a constant transition probability after being in bad (good) health for two years or longer. Thus, $\{\eta_i\}_{i=1}^5$

[18] This assumption is innocuous since the majority (96%) of individuals are healthy at this age. Most people become unhealthy later on, after receiving a health shock.

[19] We first do a grid search over the possible values of $\boldsymbol{\theta_H}$, and then use the simplex method to find the minimum using the parameters obtained from the grid search as our initial guess.

and b_5 are identified from the transition probabilities over the durations longer than two years, as plotted in Figure (2).[20] Finally, the coefficients $\{a_1, a_2, a_3, a_4\}$ are used to capture the difference in the transition probabilities between those in bad (or good) health for at least one year vs. two years. The solid lines in Figure (1) and white bars in Figure (2) show that our parsimonious model of health captures both the cross-sectional and dynamic moments of health over the life-cycle relatively well.

3.3 Estimation results

The implications of our estimated health process are illustrated in Figure (4). The left (right) panel of the figure plots the probability of moving from bad to good (good to bad) health conditional on one's fixed health type and duration of the current health status (bad and good, respectively). Comparing the two panels reveals a striking difference in what generates persistence of good and bad health. The left panel shows that fixed heterogeneity has a large impact on the probability of recovering from bad health. For example, a 60-year-old individual of health type η_1 (the "worst" type), who is in bad health, has about a 5% probability of recovering, while a 60-year-old individual of type η_5 (the "best" type) has about an 80% probability of recovering. At the same time, once fixed heterogeneity is controlled for, duration dependence plays little role: individuals who spend one year being unhealthy have almost the same probability of recovering as individuals who spend more than two years being unhealthy *conditional on being of the same health type* (see the comparison of the dashed and solid lines for each health type).[21]

The right panel of Figure (4) shows that in contrast to the probability of recovering from bad health, the probability of becoming sick is influenced very little by health type: what plays an important role in this case is duration dependence. For example, a 60-year-old individual who has been healthy for two or more years has less than a 10% probability of becoming unhealthy while an individual of the same age who just recovered (has been healthy for only one year) has close to a 50% probability of relapsing back into bad health.

For an external validation of our estimated health process, we turn to the HRS and select a sample of *healthy* males with a high-school degree, age 55-56, and whom we can observe in every survey year until they are 65-66. This leaves us with 828 individuals in the balanced panel data. We then compute the distribution of the number of unhealthy periods that these individuals report over the next ten years. Since the HRS is a bi-annual survey, an individual can only report being unhealthy for at most five periods. We then construct a comparable

[20] Since we assume that η_i is symmetric around zero, we estimate only $\{\eta_1, \eta_2\}$.

[21] We also estimated an alternative model where the second-order Markov processes in Equations (1)-(2) are replaced with third-order Markov processes. The estimations are not much different from Figure (4) and the probabilities of recovering from bad health still depend mostly on health types.

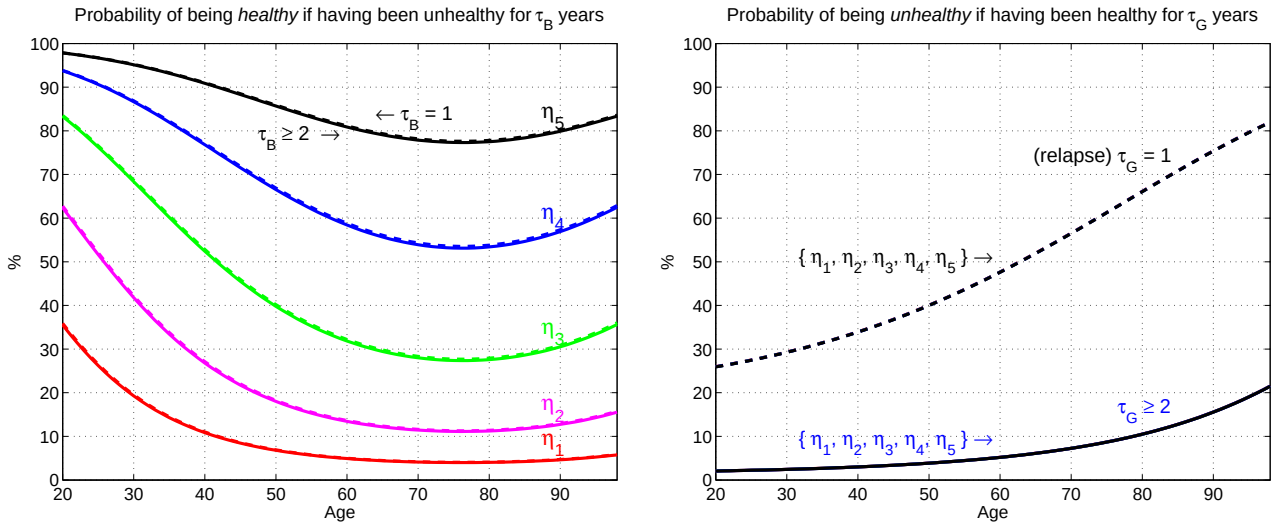


Figure 4: Estimated health process. Dotted line: Conditional on the duration of the current health status being one year ($\tau = 1$). Solid line: Conditional on the duration of the current health status being at least two years ($\tau \geq 2$).

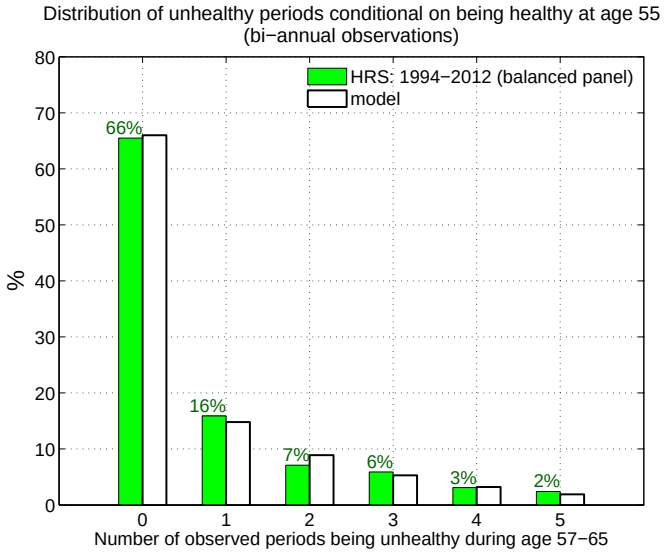


Figure 5: Distribution by unhealthy periods: HRS vs model.

distribution using simulated data from our model. Figure (5) shows that our simulated data and the data from HRS are very close.

3.4 What accounts for the long spells of bad health?

Using our estimated model, we can construct the lifetime distribution of unhealthy years over the working period. The left panel of Figure (6) plots the distribution of individuals by the total number of years that they have spent being unhealthy between ages 20 and 64,

conditional on being alive at age 64. Most people are relatively healthy during their working life: 72% of individuals experience fewer than 5 years of bad health. However, a non-trivial number of individuals spend more than a third of their working period being unhealthy. For instance, 6% of individuals experience 16 or more years in bad health. The right panel of Figure (6) illustrates how this distribution differs across health types by comparing two extreme groups: individuals born with the best health type (η_5) and those born with the worst health types (η_1 and η_2). Among individuals with η_5 type, 91% spend fewer than 5 years being unhealthy and almost none of them experiences more than 11 unhealthy years. In contrast, among η_1- and η_2-type individuals, 21% endure between 11 and 20 unhealthy years, and 8% are unhealthy for 20 years or longer. Thus, long spells of bad health are primary concentrated among individuals with the worst health types. In other words, long spells of bad health are mostly due to fixed heterogeneity rather than repeated draws of bad realizations from a persistent stochastic process.

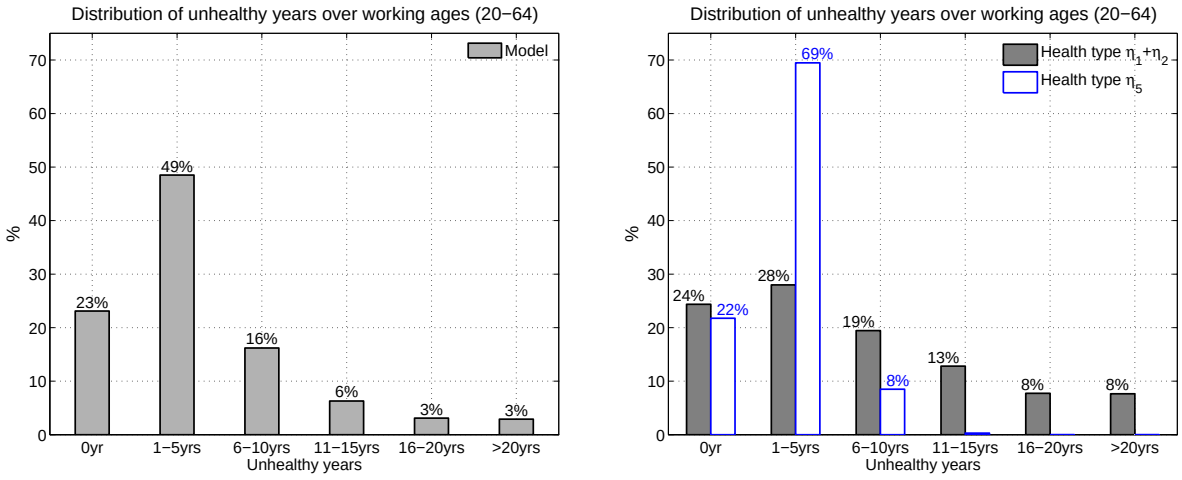


Figure 6: Distribution by lifetime unhealthy years. Left panel: all individuals. Right panel: individuals with $\{\eta_1, \eta_2\}$ and η_5 health types.

3.5 How should the health type be interpreted?

As our previous discussion shows, the health type (η) plays an important role in the persistence of bad health: our specification allows for the possibility that people have different abilities to cope with illness, and our estimation shows that this heterogeneity is substantial.[22]

[22] Halliday (2008) uses the PSID to estimate a dynamic model of health status with fixed heterogeneity and heterogeneous persistent coefficients. He finds that for a large part of his sample, persistence is mostly driven by fixed heterogeneity. Lange and McKee (2012) estimate a dynamic latent health model using multiple health measures available in the HRS. They also find that heterogeneity across individuals (random effects) is important in capturing the high persistence of objective and self-reported health measures.

Individuals can recover differently from sickness due to the variation in genetic predisposition and/or lifestyle, where the latter can be partly due to habits developed in childhood. In addition, a number of studies find that childhood circumstances can have a long lasting effect on adult health. To look for evidence supporting these mechanisms we resort to the HRS, which has a large sample size and more detailed information on individuals' characteristics. We use the same sample of individuals used to construct Figure (5); that is, a balanced panel of healthy individuals aged 55-56 and whom we observe until they are 65-66.

# unhealthy periods	Individuals' characteristics[a] (HRS)					% $\{\eta_1, \eta_2\}$ in model
	% smoking	BMI[b]	% father alive	% mother alive	parents' educ (yrs)[c]	
0-1	23.2	27.8	21.2	49.5	10.0 / 10.4	29.6
2-3	25.9	29.2	20.2	46.7	8.9 / 9.5	39.7
4-5	43.5	30.0	15.2	36.9	8.2 / 9.0	71.1

[a] All variables are reported at age 55-56.

[b] BMI is the average Body Mass Index.

[c] The first and second numbers are the average educational years of father and mother, respectively.

Table 1: Characteristics of individuals at age 55-56 by the number of unhealthy periods between ages 57-58 and 65-66. All individuals are healthy at age 55-56. The sample size for individuals with 0-1 periods being unhealthy is between 597 and 674, depending on the variable. The sample size for 2-3 periods and 4-5 periods are 97-108 and 42-46, respectively.

# unhealthy periods	Polygenic scores (HRS)		
	educational attainment	smoking	BMI
0-1	-0.104	-0.014	-0.015
2-3	-0.179	-0.018	0.066
4-5	-0.645	0.269	0.234

Table 2: The average Polygenic scores of individuals by the number of unhealthy periods between ages 57-58 and 65-66. All individuals are healthy at age 55-56. The sample size for individuals with 0-1, 2-3, and 4-5 periods being unhealthy are 501, 69, and 27, respectively.

Table 1 sorts the HRS sample based on the total number of unhealthy periods that they report over the ten-year interval. An interesting observation is that there is a correlation between the *future* number of unhealthy periods and factors that can be linked to lifestyle *recorded at age 55-56* or genetics. In particular, individuals who report being unhealthy for four to five periods between ages 57-58 and 65-66, are much more likely to smoke, to have a higher body mass index (BMI), and less likely to have living parents.

The next correlation worth noting is between the number of unhealthy periods and parental education: individuals with longer unhealthy spells have less educated parents. This is consistent with the findings of Case et al. (2002) who show that parental income and education have a significant impact on child's health and thus on the subsequent health evolution during adulthood.

Overall, Table 1 shows that even in a relatively homogeneous sample of healthy males (aged 55-56) with the same educational attainment there is heterogeneity in some fixed or long-lasting factors, which in turn are correlated with their health evolution during the next ten years. These features of the data are consistent with our stylized model of health dynamics: the last column of Table 1 shows that in a comparable sample simulated from our model, 71% of individuals who experience 4-5 unhealthy periods have the worst health types (η_1,η_2).

In Table 2 we further document the relationship between the number of unhealthy periods (our proxy for health types) and behavioral genetic variables recently found to be a robust predictor of economic outcomes. Specifically, Barth et al. (2017) find that wealth at retirement is positively associated with the Polygenic scores for educational attainment and negatively associated with the Polygenic scores for smoking and BMI (after controlling for education, labor income, and other observables). These scores are indices created by combining various genetic markers that robustly predict individuals' educational attainment, smoking behavior, and BMI. The scores are normalized to have mean of zero and variance of one. A higher score predicts a higher education level, a higher propensity to be a smoker, and a higher BMI, respectively.[23] Table 2 reports the average scores of the same HRS sample as in Table 1 sorted by the number of unhealthy periods. Notice that individuals reporting 4-5 unhealthy periods between the ages of 57-58 and 65-66, on average, have a noticeably lower score for educational attainment and higher scores for smoking and BMI comparing with those who report at most one unhealthy period. Appendix A shows that this finding is robust when we compare the $25^{th}, 50^{th}$, and 75^{th} percentiles of the scores distribution. This suggests that, to a certain extent, the wealth-health gradient can be traced back to factors determined early in life. We formalize this mechanism in the next section when we discuss our model.

4 Our life-cycle model

In this section, we construct a life-cycle model with health uncertainty, where health affects individuals through multiple channels and evolves according to the process described in the previous section.

4.1 Demographics, preferences, and labor income

A model period is one year and each individual lives at most T periods. During the first $R-1$ periods of life an individual chooses whether to work or not. At age R all individuals

[23] For more details on the construction of this score, see Papageorge and Thom (2016).

retire. We denote the health-dependent survival probability from age t to $t+1$ as ζ_t^h (this probability is plotted in Figure 3).

At age t an agent's health, h_t, can be either good ($h_t = 1$) or bad ($h_t = 0$). Health evolves according to the process defined in Equations (1) and (2), i.e., one's current health status depends on one's health status in the previous two periods and health type $\eta_i \in \{\eta_1, ..., \eta_5\}$.

Health and economic outcomes in our model are linked via two mechanism. First, health directly affects medical spending, productivity, disutility from work, access to health insurance, and survival probabilities.[24]

Second, individuals differ in their rates of time preferences which, in turn, can be correlated with their health type.[25] This correlation captures the fact that some factors determined early in life can influence both health and economic outcomes. The mechanism creates a compositional difference between the healthy and unhealthy; among the latter there can be more impatient people, thus partially accounting for the observed disparity in economic outcomes between the two groups.

Formally, we assume that individuals' discount factor (β_i) can take two values, $\beta_i \in \{\beta_{low}, \beta_{high}\}$, where $\beta_{low} < \beta_{high}$. At age 20 (when an individual enters the model) the joint distribution of $\{\beta_i, \eta_i\}$ is captured by $Pr(\beta_j|\eta_m) \in [0,1]$, where $j \in \{low, high\}$ and $m \in \{1, ..., 5\}$.

An individual is endowed with one unit of time that can be used for either leisure or work. Labor supply (l_t) is thus indivisible; $l_t \in \{0, 1\}$. Work implies a fixed utility cost ϕ_W for healthy individuals and $\phi_W + \phi_B$ for unhealthy ones. We assume that the preferences of individuals over consumption and leisure take the following form:[26]

$$u(c_t, l_t, h_t) = \frac{(c_t/\overline{n}_t)^{1-\rho}}{1-\rho} - \phi_W \mathbf{1}_{\{l_t>0\}} - \phi_B \mathbf{1}_{\{h_t=0, l_t>0\}} + \overline{b}, \tag{4}$$

[24] These direct effects have been used in other structural models with health uncertainty, e.g., Capatina (2015), French (2005), French and Jones (2011), Pashchenko and Porapakkarm (2013, 2016a, 2017), Rust and Phelan (1997).

[25] Individuals' heterogeneity in the rates of time preferences is supported by empirical studies, e.g., Epper et al. (2018), Lawrance (1991), Warner and Pleeter (2001). In addition, Cronqvist and Siegel (2015) find that genetic differences explain a significant fraction of the variation in saving propensities across individuals, potentially through its link to the rate of time preferences or self-control.

[26] In our model, health affects utility by increasing disutility from work but does not affect the marginal utility of consumption. Theoretically, the health-dependent marginal utility can help explain the wealth-health gradient. Specifically, if the marginal utility of consumption is significantly higher in the unhealthy state, the precautionary saving among the healthy will be higher. We find that the quantitative effect of this mechanism is small because the probability of falling sick among working-age people who are healthy for at least two years is rather small. (See Figure 4.) We also tried estimating an alternative model where the marginal utility of consumption in the unhealthy state is increased by 30%. This modification does not affect our estimated parameters, including the correlation between β_i and η_i. In addition, the effect of health deterioration on marginal utility is inconclusive. (For a brief review, see Finkelstein et al., 2009).

where $\mathbf{1}_{\{.\}}$ is an indicator function which is equal to one if its argument is true and zero otherwise, ρ is risk-aversion, and $\overline{n}_t$ is an age-specific household size. We follow Hall and Jones (2007) by adding a positive term $\overline{b}$ to ensure that individuals in our model value their life; i.e., the continuation value of being alive exceeds the utility when deceased. This is important because otherwise individuals would welcome higher mortality that comes from worsening health. As explained in Section 5.2.2, we calibrate $\overline{b}$ to match the empirical statistical value of life.

Individuals also have bequest motives and derive utility from leaving a bequest of size k as follows:

$$\theta_{Beq}\frac{(k+k_{Beq})^{1-\rho}}{1-\rho},$$

where θ_{Beq} determines the strength of the bequest motive and k_{Beq} is a parameter shifter that determines to what extent bequests are a luxury good. In this approach we follow De Nardi (2004).

The earnings of individuals are equal to $z_t^h l_t$, where z_t^h is an idiosyncratic productivity component that takes the following form:

$$z_t^h = \lambda_t^h \Upsilon_t. \tag{5}$$

Here λ_t^h is a deterministic function of age and health, while Υ_t is the stochastic shock that we specify in Section 5.1.5.

4.2 Medical expenditures and health insurance

During each period every agent receives a medical expenditure shock $\left(x_t^h\right)$ which depends on age and health. We denote the distribution of medical shocks as $\mathcal{G}_t(x_t^h|h_t)$.

A working-age individual receives an offer to buy employer-sponsored health insurance (ESHI) with probability $Prob_t$, which depends on age (t), productivity (z_t^h), and health (h_t). The variable $g_t^{h,z}$ characterizes the status of the offer: $g_t^{h,z}=1$ if an individual gets an offer, and $g_t^{h,z}=0$ otherwise. Only working individuals with an offer ($l_t=1, g_t^{h,z}=1$) can purchase the ESHI insurance. We assume that an employer who offers ESHI fully covers the premium, i.e., the employer contribution is 100%.[27] In addition, every working-age individual can buy health insurance in the individual health insurance market. The price of health insurance in the individual market depends on one's age and health. We denote the individual market

[27] On average, employers who offer ESHI contribute around 80% of the premium for single coverage and around 70% for family coverage (Kaiser Family Foundation, 2004); we abstract from workers's contribution for simplicity, this assumption does not affect our results but helps to lower the computational costs since working individuals with an ESHI offer always buy insurance.

price as $p_I(h_t, t)$. All retired individuals are covered by public health insurance, Medicare. We denote the Medicare premium as P_{MCR}.

We index the insurance status of an individual by using i_H: $i_H = 0$ corresponds to being uninsured, $i_H = 1$ corresponds to individual insurance, $i_H = 2$ corresponds to group (or ESHI) insurance, and $i_H = 3$ corresponds to Medicare. All types of insurance only provide partial medical expenses coverage. We denote by $cvg\left(x_t^h, i_H\right)$ the fraction of medical expenditures covered by insurance which is a function of the medical shock and insurance type. Note that $cvg\left(x_t^h, 0\right) = 0$.

4.3 Taxation and social transfers

All individuals pay an income tax $\mathcal{T}(y_t)$ that consists of two parts: a progressive tax and a proportional tax. Taxable income y_t includes labor and capital income. Working households also pay payroll taxes, which include the Medicare tax (τ_{MCR}) and the Social Security tax (τ_{ss}). The Social Security tax rate for earnings above $\overline{y}_{ss}$ is zero. Consumption is taxed at a proportional rate of τ_c.

We also assume a public safety-net program, $T^{SI}(\overline{c})$. This program guarantees every household a minimum consumption level $\overline{c}$, which is a simple way to represent several means-tested programs in the U.S., such as Medicaid, food stamps, and Supplement Security Income. In addition, the consumption floor captures the existence of uncompensated care or medical bankruptcy.[28]

Retirees receive Social Security benefits ss. In practice, these payments depend on an individual's history of earnings. To reduce computational costs, however, we allow ss to depend only on one's health type η_i and fixed productivity type, which is part of the productivity Υ_t (see Section 5.1.5). Since the labor supply decisions of individuals over the life-cycle are affected by fixed productivity and health types, this approach allows us to capture the resulting heterogeneity in pension benefits without introducing an additional state variable.

4.4 Timing of the model

The timing of the model is as follows. At the beginning of the period, individuals learn their productivity, health and ESHI offer status. Based on this information, an individual decides his labor supply (l_t) and insurance choice (i_H). At the end of the period, the medical expenses shock (x_t^h) is realized. After paying the out-of-pocket medical expenses, an individual chooses his consumption (c_t) and savings for the next period (k_{t+1}). The problem of retirees is simpler; they only choose consumption and savings for the next period.

[28] In 2004, 85 percent of the uncompensated care was paid by the government.

4.5 The optimization problem

Working age individuals ($t < R$)**.** At the beginning of each period, the state variables for an individual i are capital ($k_t \in \mathbb{K} = R^+ \cup \{0\}$), health status in the current and previous periods ($h_t, h_{t-1} \in \mathbb{H} = \{0, 1\}$), idiosyncratic labor productivity $\left(z_t^h \in \mathbb{Z} = R^+\right)$, ESHI offer status $\left(g_t^{h,z} \in \mathbb{G} = \{0, 1\}\right)$, age ($t \in \mathbb{T} = \{1, 2, ..., R-1\}$), health type ($\eta_i \in \{\eta_1, ..., \eta_5\}$) and discount factor ($\beta \in \{\beta_{low}, \beta_{high}\}$).[29] We denote the vector of state variables as $\mathbb{S}_t : \mathbb{S}_t \in \mathbb{K} \times \mathbb{H} \times \mathbb{H} \times \mathbb{Z} \times \mathbb{G} \times \mathbb{T} \times \{\eta_1, ..., \eta_5\} \times \{\beta_{low}, \beta_{high}\}$.

The value function of a working age individual at the beginning of period t is:

$$V_t^i\left(\mathbb{S}_t\right) = \max_{l_t, i_H} \sum_{x_t^h} \mathcal{G}_t(x_t^h | h_t) W_t^i(\mathbb{S}_t; l_t, i_H, x_t^h) \tag{6}$$

where

$$W_t^i(\mathbb{S}_t; l_t, i_H, x_t^h) = \max_{c_t, k_{t+1}} u\left(c_t, l_t, h_t\right) + \beta_i \left[\zeta_t^h E_t\Big(V_{t+1}^i\left(\mathbb{S}_{t+1}\right)\Big) + \left(1 - \zeta_t^h\right) \theta_{Beq} \frac{(k_{t+1} + k_{Beq})^{1-\rho}}{1-\rho}\right] \tag{7}$$

subject to

$$k_t\left(1 + r\right) + z_t^h \; l_t - x_t^h \left(1 - cvg(x_t^h, i_H)\right) - P_t^h - Tax + T^{SI}\left(\overline{c}\right) = (1 + \tau_c)c_t + k_{t+1} \tag{8}$$

$$P_t^h = \begin{cases} 0 & ; \text{ if } i_H \in \{0, 2\} \\ p_I\left(h_t, t\right) & ; \text{ if } i_H \in \{1\} \end{cases} \tag{9}$$

$$T^{SI}(\overline{c}) = max\Big(0, (1 + \tau_c)\overline{c} + Tax + P_t^h + x_t^h\big(1 - cvg(x_t^h, i_H)\big) - k_t(1 + r) - z_t^h l_t\Big)$$

$$Tax = \mathcal{T}\left(y_t\right) + \tau_{MCR} z_t^h l_t + \tau_{ss} \min\left(z_t^h l_t, \overline{y}_{ss}\right) \tag{10}$$

$$y_t = k_t r + z_t^h l_t \tag{11}$$

$W_t^i(\mathbb{S}_t; l_t, i_H, x_t^h)$ is the interim value function conditional on the labor supply and insurance choices after the medical shock is realized. The conditional expectation on the right-hand side of Equation (7) is over $\left\{h_{t+1}, z_{t+1}^h, g_{t+1}^{h,z}\right\}$. Equation (8) is the budget constraint; in this constraint P_t^h is the insurance premium, which is described in Equation (9). In Equation (10), the first term is the income tax and the last two terms are payroll taxes. Equation (11) describes taxable income.

We assume that an individual's insurance premium is based on his expected medical costs

[29] To make our expression less clustered, we omit the subscript i for all state variables. Also, to simplify the notation, we use z_t^h as the state variable for labor productivity even though z_t^h is composed of an AR(1) and fixed productivity component as will be discussed in Section 5.1.5.

and administrative loads:

$$p_I\left(h_t,t\right)=\xi^h EM_t\left(h_t,t\right)+\varphi^h. \tag{12}$$

The term ξ^h is a proportional load, while φ^h is a fixed load. We allow the loads to depend on health to capture the fact that unhealthy individuals may face more frictions when purchasing insurance through the individual market, for example, through search costs or a larger probability of being denied coverage due to pre-existing conditions. The expected medical costs covered by insurance are determined as follows:

$$EM_t\left(h_t,t\right)=\sum_{x_t^h} x_t^h cvg\left(x_t^h,1\right)\mathcal{G}_t(x_t^h|h_t).$$

Retired individuals ($t \geq R$). The state variables for retired people are assets (k_t), health in the current and previous periods (h_t, h_{t-1}), medical shock $\left(x_t^h\right)$, age $\left(t \in \mathbb{T}^R = \{R, ..., T\}\right)$, health type ($\eta_i$), and discount factor ($\beta_i$).[30] We denote the vector of state variables as $\mathbb{S}_t^R:\ \mathbb{S}_t^R \in \mathbb{K} \times \mathbb{H} \times \mathbb{H} \times \mathbb{T}^R \times \{\eta_1, ..., \eta_5\} \times \{\beta_{low}, \beta_{high}\}$.

The value function of a retired household is:

$$V_t^i\left(\mathbb{S}_t^R\right)=\sum_{x_t^h}\mathcal{G}_t(x_t^h|h_t)W_t^i(\mathbb{S}_t^R;x_t^h) \tag{13}$$

where

$$W_t^i(\mathbb{S}_t^R;x_t^h)=\max_{c_t,k_{t+1}}\ u\left(c_t,0,h_t\right)+\beta_i\left[\zeta_t^h E_t\Big(V_{t+1}^i\left(\mathbb{S}_{t+1}^R\right)\Big)+\left(1-\zeta_t^h\right)\theta_{Beq}\frac{(k_{t+1}+k_{Beq})^{1-\rho}}{1-\rho}\right] \tag{14}$$

subject to

$$k_t\left(1+r\right)+ss-x_t^h\left(1-cvg(x_t^h,3)\right)-P_{MCR}-\mathcal{T}\left(y_t\right)+T^{SI}\left(\overline{c}\right)=(1+\tau_c)c_t+k_{t+1} \tag{15}$$

$$T^{SI}(\overline{c})=max(0,(1+\tau_c)\overline{c}+\mathcal{T}\left(y_t\right)+P_{MCR}+x_t^h(1-cvg(x_t^h,3))-k_t(1+r)-ss)$$

$$y_t=k_t r+ss \tag{16}$$

$W_t^i(\mathbb{S}_t^R;x_t^h)$ is the interim value function conditional on medical shock realization. The conditional expectation on the right-hand side of Equation (14) is over h_{t+1}. Equation (15) is the budget constraint.

[30] As explained in Section 4.3, Social Security payments ss depend on the fixed productivity type; thus, fixed productivity is also part of the state variables for retired households. We omit it from the description of the optimization problem to simplify the notation.

5 Model parameters estimation

In this section, we explain our strategy to estimate the model parameters and exogenous shocks. The information related to earnings, labor supply, and wealth is taken from the PSID, while the information about medical expenses, health insurance coverage, and ESHI status is taken from the MEPS.[31]

We adopt a two-step estimation strategy. In the first step, we set parameters related to demographics, taxes, social security benefits, and health insurance and estimate the shock processes directly from the data. In the second step, we estimate the remaining parameters using our structural model to match the targeted moments from the data.

5.1 First step estimation/calibration

In the following, we describe parameters and the shock processes set/estimated during the first step. The survival probability and health shock process are taken from Section 3.3.

5.1.1 Demographics

An individual enters the model at the age of 20, retires at age 65 and lives at most until the age of 99. The age-dependent family size $\overline{n}_t$ is the average family size from the PSID. The average family size is 1.8 at age 20 and gradually increases to its largest size (3.2) at age 38. At retirement, the average family size is 2.1. We set the risk aversion ρ to 3, which is within the range of 1 to 5 commonly used in macro and structural life-cycle models.

5.1.2 Taxes and social security benefits

In specifying the tax function $\mathcal{T}(y)$ we use a combination of the nonlinear functional form formulated by Gouveia and Strauss (1994), and a linear income tax τ_y:

$$\mathcal{T}(y) = a_{\tau 0}\left[y - (y^{-a_{\tau 1}} + a_{\tau 2})^{-1/a_{\tau 1}}\right] + \tau_y y.$$

The first term captures the progressive income tax; in this functional form, $a_{\tau 0}$ controls the marginal tax rate faced by the highest income group, $a_{\tau 1}$ determines the curvature of marginal taxes, and $a_{\tau 2}$ is a scaling parameter. Following Gouveia and Strauss (1994) we set

[31] A household unit in the MEPS includes all members who would be covered under a typical family health insurance plan. We use the sample of household heads with a high school education, where household heads are males with the highest earnings. The MEPS has five rounds of interviews over two years. In each round, individuals are asked to rank their health as *excellent*, *very good*, *good*, *fair*, or *poor*, similar to the PSID and HRS. We classify an individual as healthy in a certain year if his self-reported health falls in the first three categories for at least two rounds in that year.

$a_{\tau 0}$ and $a_{\tau 1}$ to 0.258 and 0.768, respectively. The parameters $a_{\tau 2}$ and τ_y are set to 0.6160 and 0.066 percent, respectively, following Pashchenko and Porapakkarm (2017). The Medicare, Social Security and consumption tax rates were set to 2.9 percent, 12.4 percent and 5.67 percent, respectively. Using the social security rule in 1996, the maximum taxable income for Social Security ($\overline{y}_{ss}$) is set to \$62,700.

To compute the Social Security pension payments ss, we use the Social Security benefit formula in 1996 and apply it to the average labor income over working ages. Specifically, we first group individuals in our model based on their health types and fixed productivity defined in Section 5.1.5. Then for each group the average labor income over working ages is defined as the average labor income over the 35 highest-earning years.

5.1.3 The medical expenses shock

The medical costs in our model correspond to total paid medical expenditures in the MEPS dataset. The medical expense shock is approximated by a 3-state discrete health- and age-dependent stochastic process. To capture the highly skewed distribution of medical expenses, for each age and health status, these three states correspond to the average medical expenses of three groups: those with medical expenses below the 50^{th}, 50^{th}-97^{th}, and above the 97^{th} percentiles, respectively.[32] More details on the estimation of our medical shock process are provided in Appendix B.

5.1.4 Health insurance and ESHI offer probability

We define a person as having employer-based insurance in the MEPS if he reports having ESHI for at least eight months of the year. The same criterion is used when defining a person as having individual insurance.[33] Due to the small sample size of those with individual insurance, we assume that ESHI and individual insurance provide the same coverage; $cvg\left(x_t^h, 1\right) = cvg\left(x_t^h, 2\right)$. We combine information on individuals insured by ESHI and individual insurance to estimate the fraction of medical expenses covered by these insurance policies. We report these estimates in Appendix B. For retired households, we set $cvg\left(x_t^h, 3\right)$ to 0.5 following Attanasio et al. (2011).

[32] The MEPS tends to underestimate aggregate medical expenditures (Pashchenko and Porapakkarm, 2016b). To bring aggregate medical expenses computed from the MEPS in line with the corresponding statistics in the National Health Expenditure Account (NHEA), the estimated medical expenses were multiplied by 1.60 for people younger than 65 years old and by 1.90 for people 65 or older. These numbers correspond to the ratio of aggregate medical spending in the NHEA divided by aggregate medical spending in the MEPS for people younger and older than 65 years old, respectively, averaged over the years 2002, 2004, 2006, 2008, and 2010 (the years when the NHEA provides the aggregate statistics by age).

[33] If a person reports having both ESHI and individual insurance in one year and each coverage lasts for eight months or less but the total duration of coverage lasts for more than eight months, we classify this person as individually insured.

We assume that individuals receive an ESHI offer with probability $Prob_i$, which is estimated from the following logit regression:[34]

$$logit\left(ESHI_{it}\right)=\sum_{j=0,1}\Big(a_{E0}^{j}+a_{E1}^{j}\log\widehat{inc}_{it}+a_{E2}^{j}\left(\log\widehat{inc}_{it}\right)^{2}+a_{E3}^{j}\left(\log\widehat{inc}_{it}\right)^{3}+a_{E4}^{j}ESHI_{it-1}\Big)\mathbf{1}_{\{h_{it}=j\}}. \quad (17)$$

$ESHI_{it}$ is one if individual i is insured through ESHI, otherwise it is zero; $\widehat{inc}_{it}$ is labor income normalized by the average income of the same year and $ESHI_{it-1}$ is ESHI status in the previous year. We include income in the regression to capture the positive relationship between labor income and ESHI coverage as observed in the data. In the MEPS, unhealthy individuals have a noticeably lower ESHI coverage even after controlling for labor income. To capture this we allow the ESHI offer probability to be different between the healthy and unhealthy. For the initial distribution at age 20, we run a separate logit regression among individuals aged 19-22 without including $ESHI_{it-1}$ in the regression specification.

For individual insurance premiums we use the estimates from Pashchenko and Porapakkarm (2017) and set the fixed load φ^h at zero for the healthy and \$790 for the unhealthy. The proportional load ξ^h is 1.079 for the healthy and 1.135 for the unhealthy.

5.1.5 The labor productivity shock

We specify labor productivity in our model as follows:

$$\begin{aligned} z_t^h &= \lambda_t^h \Upsilon_t = \lambda_t^h \exp(\nu_t)\exp(\gamma), && (18)\\ \nu_t &= \rho_\nu \nu_{t-1} + \varepsilon_t; \quad \varepsilon_t \sim N(0,\sigma_\varepsilon^2), && (19)\\ \gamma &\sim N(0,\sigma_\gamma^2), \end{aligned}$$

where λ_t^h is a deterministic component that depends on age and health. The idiosyncratic component Υ_t consists of a persistent shock ν_t and fixed productivity γ.

We use the PSID to estimate the parameters determining the evolution of z_t^h. First we compute labor income, defined as earnings plus income from business (converted into 1996 dollars). Given that an individual in our model makes a discrete labor supply choice, we define a worker as a person younger than 65 who earns at least \$2,470 per year in base year dollars (this corresponds to working at least 10 hours per week and earning a minimum wage of \$4.75 per hour). The crossed marks in Figure 7 show the fraction of workers (left panel)

[34] We use only individuals who earn more than \$2,470 per year in base year dollars. Note that we use the ESHI status as opposed to ESHI offer status in our logit regression. Since everyone in our model always buys employer-based insurance if offered, there is no difference between being insured through ESHI and receiving an ESHI offer in our model. In the MEPS, 95% of individuals who receive an ESHI offer take it.

and the average labor income conditional on working (right panel) by health status from the PSID. For all age groups, unhealthy individuals work much less and, conditional on working, earn significantly less than healthy individuals.

Because the fraction of unhealthy workers is significantly below 100%, average income conditional on working could be a biased estimate of $\lambda_t^{h=B}$ if there is selection into employment. To account for the potential selection problem, we estimate the parameters of λ_t^h inside the model in the second step.[35]

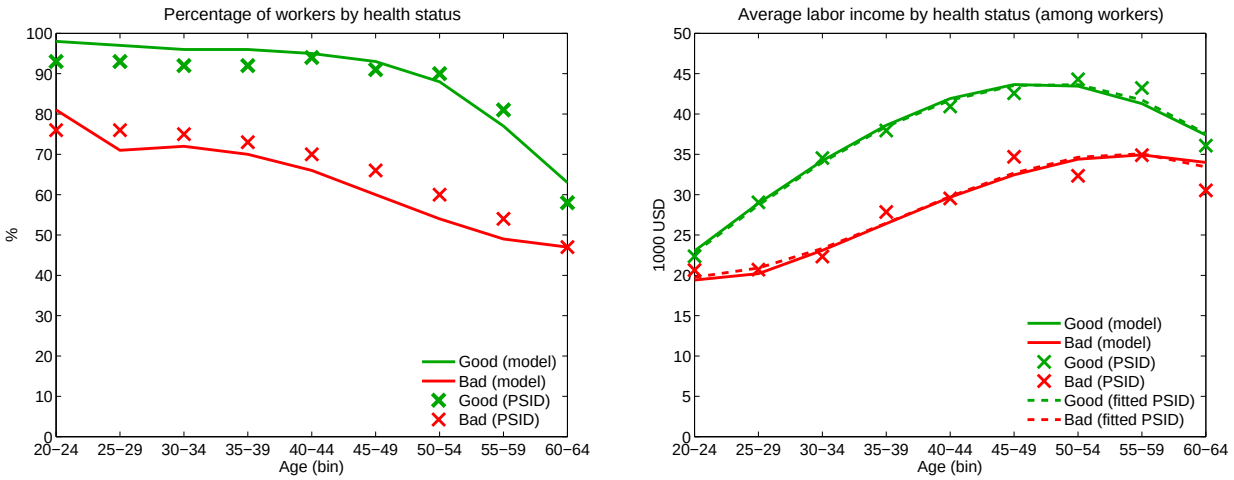


Figure 7: Employment by health (left panel) and average labor income among workers by health (right panel): The crossed marks are from the PSID while the dashed lines in the right panel are the corresponding approximations using a third degree polynomials of age. The solid lines are from our model.

We estimate the stochastic component Υ_t in the first step using the sample of working individuals.[36] We have 1,036 working individuals and 10,778 individual-year observations. Our estimation strategy closely follows that of French (2005). First, we run the following fixed effect regression:

$$log(inc_{it}) = \sum_{j=0,1} \left(b_{I1}^j age_{it} + b_{I2}^j age_{it}^2 \right) \mathbf{1}_{\{h_{it}=j\}} + \sum_{t=1984}^{1997} b_{It} D_t + \gamma_i + u_{it},$$

where inc_{it} is labor income and D_t is a set of year dummy variables. By construction, the residual u_{it} is orthogonal to health and age. Thus, it is the empirical counterpart of the

[35] An alternative method is to apply the two-step Heckman estimation directly to the data. However, this approach requires a valid variable serving as an exclusion restriction. Though more computationally costly, our strategy ensures that the model reproduces the income-health gradient in the data (Figure 7 and Table 6), which is important for evaluating the cost of bad health through the labor market channel.

[36] Note that the parameters in Equation (19) are assumed to be independent of health status and age. Since most workers are healthy and over 90% of healthy people work, we are less concerned about the selection problem when estimating the parameters of Υ_t directly from the data. An alternative approach is to use only the sample of healthy workers younger than 50, but this would reduce our sample size.

AR(1) component in Equation (19).

Next, we construct the empirical autocovariance matrix using the residuals ($\gamma_i + u_{it}$). We estimate the parameters of the productivity shock by minimizing the distance between the empirical autocovariance matrix and the corresponding matrix implied by Equation (19).[37] Our resulting estimates are as follows: $\rho_\nu = 0.9275$, $\sigma^2_\varepsilon = 0.021$, and $\sigma^2_\gamma = 0.042$, which are within the range of values estimated in the literature. In our computation we discretize the shock processes using nine gridpoints for ν_t and three gridpoints for γ. Our grid for ν_t expands with age to capture the observed cross-sectional variance that increases with age. We adopt the method in Floden (2008) for our discretization since it performs well for highly persistent processes, such as our AR(1) process.

5.2 Second step estimation

Given the parameters and the shock processes from the first step, we implement the Method of Simulated Moments to estimate the remaining parameters in our model, i.e., we minimize the distance between the targeted and simulated moments. Below we describe the two sets of moments that we explicitly target.

First set of moments. To capture the effects of health through the labor market channel and to replicate the income-health gradient, we target the fraction of workers and the average labor income conditional on working among the healthy and unhealthy for each age-group (Figure 7).[38] These moments are used to pin down the disutility from work $\{\phi_W, \phi_B\}$ and the health-dependent productivity profile λ^h_t.

Second set of moments. The lifetime costs of bad health depend on how well individuals are insured against labor income and medical expense risks. Because saving is an important instrument to insure against these risks, it is important that our model can reproduce the wealth distribution by health status. To capture the wealth-health gradient, we target the $25^{th}, 50^{th}$, and 75^{th} percentiles of wealth, conditional on health status by age, between the ages of 30 and 85 (dashed lines in Figure 8). We discard the wealth moments below age 30 because we assume that individuals enter the model with zero assets. These moments are informative about the preference parameters $\{\beta_{low/high}, Pr\left(\beta_{low}|\eta_i\right), \theta_{Beq}, k_{Beq}\}$ and the consumption floor $\overline{c}$.

To construct our targeted wealth profiles, we use net worth from the PSID (1994, 1999-2013).[39] Because net worth is measured at the household level and our model abstracts from

[37] This is a standard procedure commonly used in the literature. See for example, Storesletten et al. (2004) and French (2005).

[38] In our estimation, we smooth the targeted average labor income profile conditional on working by the fitted third degree polynomials of age (dashed lines in the left panel of Figure 7).

[39] Specifically, net worth is derived from the value of business/farm, checking/saving accounts, real estate,

heterogeneity in family size, we adjust observed wealth by family size as follows:

$$wealth_{it} = \sum_{j=0,1} \left(d^j_{age} D^{age}_{it} + d^j_1 n_{it} + d^j_2 n^2_{it} \right) \mathbf{1}_{\{h_{it}=j\}} + \sum_{t=1994}^{2013} d_t D_t + res_{it}, \tag{20}$$

where $wealth_{it}$ is net worth, D^{age}_{it} and D_t are age and year dummy variables, and n_{it} is the number of individuals in a family unit. Given the estimated coefficients and the residuals res_{it}, we replace n_{it} in the above equation with the average family size at each age, $\overline{n}_{age}$, to get our measure of net worth.[40] Then we construct the targeted $25^{th}, 50^{th}$, and 75^{th} percentiles of wealth distribution by health status and report them as dashed lines in the left panels of Figure 8. As a comparison, we also apply the same method to net worth in the HRS (1994-2012) and plot the results as crossed marks in the left panels of Figure 8.

The wealth profiles from the two datasets are remarkably similar. Figure 8 displays the wealth-health gradient typically documented in the literature: the right-hand-side panels of the figure emphasize this gradient by plotting the gap in wealth between the healthy and unhealthy. This gap starts at relatively young ages and widens until retirement age. This feature of the data suggests that it is important to model the entire life-cycle to understand the costs of bad health.

5.2.1 The estimation algorithm

Our estimation strategy exploits the fact that in our model there are two mechanisms generating a gap in accumulated wealth between the healthy and unhealthy. First, due to the high out-of-pocket medical expenses and low earnings, unhealthy individuals have limited resources to save. Second, a larger fraction of the unhealthy might have a lower discount factor (β_{low}) and thus choose to save less. The basic idea in our estimation strategy is to sequentially isolate these two mechanisms. First, we ensure that the resources available to the unhealthy in our model are comparable to those in the data by capturing the important differences in labor market outcomes in Figure 7. Second, our algorithm adjusts the composition of individuals with β_{low} among the healthy and unhealthy until our model replicates the wealth-health gradient in Figure 8.[41] Appendix C formally explains our estimation

stock, vehicles, other assets, annuity/IRA accounts, and home equity net of the value of mortgages/debts. We convert it to 1996 dollars using the CPI.

[40] Specifically, we compute $\sum_{j=0,1} \left(\widehat{d}^j_{age} D^{age}_{it} + \widehat{d}^j_1 \overline{n}_{age} + \widehat{d}^j_2 \overline{n}^2_{age} \right) \mathbf{1}_{\{h_{it}=j\}} + \widehat{d}_{1999} + \widehat{res}_{it}$, where $\widehat{d^j}$ and $\widehat{res}_{it}$ are the estimated coefficients and the residuals from Equation (20). By construction, we remove the variation in net worth due to the variation in family size that is orthogonal to health status and age. We choose $\widehat{d}_{1999}$ because it is the closest year to our base year.

[41] Similar strategies are implemented in Erosa and Kambourov (2016), French (2005), Kaplan (2012) and Pashchenko and Porapakkarm (2013, 2017).

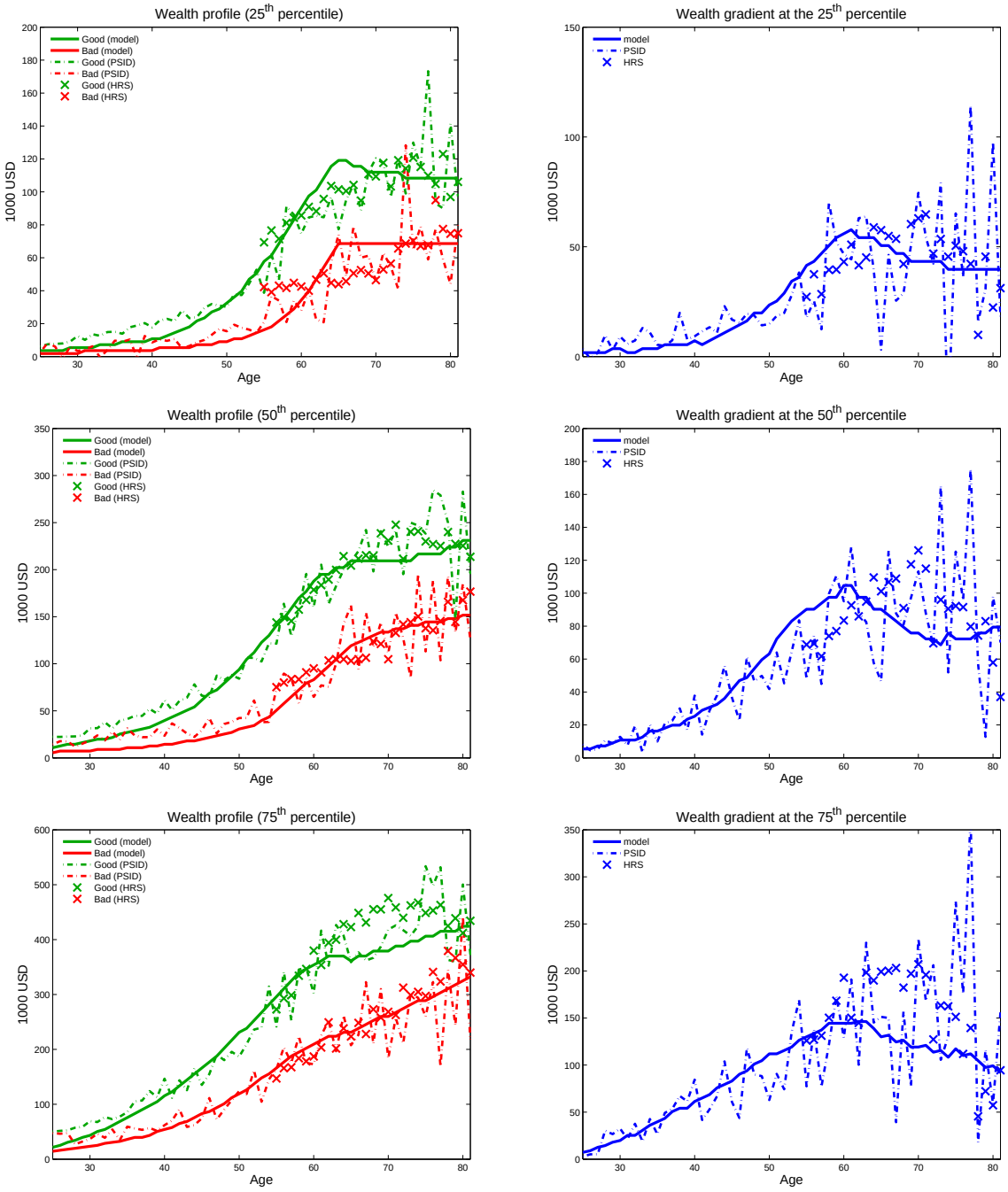


Figure 8: Wealth profiles by health status (left) and wealth gradient (right): data vs model.

algorithm.

Our estimation strategy is to highlight that a life-cycle model with endogenous labor supply can well match the income-health gradient but significantly underpredicts the wealth-health gradient. And the compositional difference between the healthy and unhealthy plays an important role in explaining the wealth-health gradient. In Section 5.3 below we illustrate this by reestimating an alternative model with no compositional difference between

the healthy and unhealthy (no correlation between η and β). This alternative model can reproduce only 40% of the difference in median wealth between health status even though it matches the income-health gradient precisely.

5.2.2 Second step estimation results

Figure 7 compares the employment rate (left panel) and the average labor income of workers (right panel) generated by our model (solid lines) with the targeted profiles from the PSID (crossed marks). Our model matches the important differences in labor market outcomes between the healthy and the unhealthy very well. Section 5.4 shows that our model also matches the observation that the lower-income group contains a much larger fraction of unhealthy people.

The left panel of Figure 8 displays the wealth profiles from our model (solid line) and the data (dashed line for the PSID and crossed marks for the HRS). Our model succeeds in matching the wealth gap between healthy and unhealthy people for the $25^{th}, 50^{th}$, and 75^{th} percentiles. Note that even though all individuals in our model start with zero assets, the simulated profiles also track very well the widening wealth-health gradient starting from younger ages, as illustrated in the right panel of the figure. Since the monetary costs of bad health (low earnings and high medical spending) among the young are relatively small comparing to older groups, the wealth gradient for the younger group is mostly explained by the larger fraction of β_{low}-individuals among the unhealthy.

Parameters		Baseline	No correlation
discount factor	$\{\beta_{low}, \beta_{high}\}$	$\{0.904, 0.995\}$	$\{0.898, 0.99\}$
% β_{low} by η_i at age 20	$Pr(\beta_{low}\|\eta_1)$	89%	50%
"	$Pr(\beta_{low}\|\eta_2)$	81%	50%
"	$Pr(\beta_{low}\|\eta_3)$	66%	50%
"	$Pr(\beta_{low}\|\eta_4)$	36%	50%
"	$Pr(\beta_{low}\|\eta_5)$	12%	50%
bequest parameter	θ_{Beq}	4,464	4,370
"	κ_{Beq}	246,371	228,476
consumption floor	$\overline{c}$	\$3,593	\$3,540

Table 3: Preference parameters and the consumption floor.

The third column of Table 3 reports our estimated preference parameters and consumption floor. The discount factors play an important role in wealth accumulation before retirement and its distribution; our estimated β_{low} and β_{high} are 0.904 and 0.995, respectively. The correlation between the discount factor and health type is identified by matching the wealth levels of the healthy and unhealthy people, conditional on reproducing the observed

labor market outcomes by health status. We find a strong correlation between one's discount factor and health type: the fraction of impatient individuals (β_{low}) among those with the worst health types (η_1, η_2) is 80%, while among those with the best health type (η_5) it is only 12%. The unconditional average of the discount factor in our model $E(\beta)$ is 0.944.

The estimated bequest parameters θ_{Beq} and k_{Beq}, which mostly affect wealth decumulation after retirement, are equal to be 4,464 and 246,371, respectively. In a one-period consumption-saving model with a risk aversion of 3, these values imply that the bequest motive becomes operational at an asset level of \$15,000 and the marginal propensity to bequeath (MPB) is 0.94. In other words, individuals with assets below \$15,000 would not leave bequests, while individuals with assets above \$15,000 would leave 94 cents out of every additional dollar for bequests. These numbers are within the range of values found in other studies.[42]

The consumption floor, which mostly affects the savings of the lower-income group, is estimated to be \$3,593. This is consistent with estimates obtained within the context of other structural models featuring the full life-cycle, medical spending uncertainty, and endogenous labor supply.[43]

Finally, because the scaling constant $\overline{b}$ in Equation (4) enters the utility function additively we can calibrate it after the second step estimation by targeting the empirical statistical value of life (SVL). The measure of the SVL emerged from the literature on compensating differentials for risky occupations. It is based on the compensation individuals are willing to accept in exchange for an increase in the probability of death expressed as "dollars per death". For example, suppose people are willing to tolerate an additional fatality risk of $1/10,000$ for a compensation of \$200 per person. Among 10,000 people in this situation there will be one death and it will cost the society $10,000 \times \$200 = \$2$ million, which is thus the implied SVL.

To calibrate $\overline{b}$ we decrease the survival probability among working-age individuals by 1/10,000 and compute a compensation that makes them indifferent between this counterfactual case and the baseline. The SVL implied by our model is the average compensation among working-age individuals multiplied by 10,000. The targeted SVL in the baseline model is \$2 millions which is the lower end of the range estimated in the literature[44]; Viscusi (1993) provided an extensive review documenting that the estimates vary from \$1 million to

[42] For example, the estimation in De Nardi et al. (2010) implies a bequest threshold of about \$36,000 and a MPB of 0.88. Pashchenko (2013) provides a comparison of the MPBs and bequest thresholds across several structural life-cycle studies.

[43] Capatina's (2015) estimate of the consumption floor is \$4,114 (in 2006 USD) while Pashchenko and Porapakkarm's (2017) estimate is \$1,540 (in 2003 USD).

[44] Our calibrated $\overline{b}$ also implies that all agents prefer living to dying. Specifically, the continuation value

$16 millions (in 1990 dollars).[45]

Overall, the SVL affects the relative importance of pecuniary and non-pecuniary effects of health over the life-cycle. For example, when the SVL is small, only the monetary effects matter. In contrast, when the SVL is high, the non-pecuniary effects become more important because health affects life expectancy and, consequently, lifetime utility. In our baseline calibration we set the targeted SVL to the lower end of the empirical estimates; thus we are estimating a lower bound of the non-pecuniary effects of health. As discussed below, our results show that even under this parameterization the non-pecuniary effects of health are very important. In Appendix E we report the results from an alternative parameterization when the SVL is equal to $6 millions, which emphasizes the non-pecuniary effects of health more.

5.3 Compositional differences and the wealth-health gradient

As shown in the previous subsection, the results of our estimation imply a non-trivial compositional difference between the healthy and unhealthy due to the estimated correlation between health type and the rate of time preferences. To illustrate the importance of this compositional difference, we re-estimate an alternative model where we restrict $Pr(\beta_j|\eta_m) = 0.5$ for all m and j. We call it the "no-correlation" model. In this case, one's discount factor is orthogonal to one's health type and, consequently, to health status. The "no-correlation" model still features all of the channels through which bad health can affect individuals' savings.

Because the "no-correlation" model has no explicit parameters to separately capture the wealth profile of the healthy and unhealthy in our estimation, we replace the targeted second set of moments (described in Section 5.2) with the *unconditional* wealth quartiles; specifically, the $25^{th}, 50^{th}$, and 75^{th} percentile of wealth for each age between 30 and 85.[46]

We report the estimated parameters for the "no-correlation" model in the last column of Table 3. These parameters turn out to be similar to those in the baseline, including the rates of time preferences $(\beta_{low}, \beta_{high})$. In addition, the model-simulated data can match the

of living in Equation (7) and (14) is strictly greater than the value of dying in all possible states:

$$E_t\Big(V^i_{t+1}\left(\mathbf{s}_{t+1}\right)\Big) > \theta_{Beq}\frac{(k_{t+1}+k_{Beq})^{1-\rho}}{1-\rho} \quad \text{for } t<T,$$

where $\mathbf{s_{t+1}} \in \mathbb{S}_{t+1}$ if $t \leq R$ and $\mathbf{s_{t+1}} \in \mathbb{S}^R_{t+1}$ if $t > R$.

[45] The US government agencies (Department of Transportation, Food and Drug Administration, Environmental Protection Agency) use the SVL of $1-10 millions in their analysis involving a mortality risk (Robinson, 2007).

[46] We have tried to target the wealth moments conditional on health but since we do not have parameters to match them the results were very similar to those when we target the unconditional moments.

Wealth percentile	PSID (HRS)	Baseline	No correlation $Pr(\beta_{low}\|\eta_i) = 0.5$
25^{th} pct	\$75,997 (\$76,253)	\$83,041	\$86,652
50^{th} pct	\$169,557 (\$165,454)	\$180,525	\$187,746
75^{th} pct	\$343,298 (\$349,858)	\$339,387	\$346,608

Table 4: *Unconditional* wealth quartiles at age 60-64.

Wealth difference by health status	PSID (HRS)	Baseline	No correlation $Pr(\beta_{low}\|\eta_i) = 0.5$
25^{th} pct	\$41,225 (\$47,569)	\$54,157	\$32,497
50^{th} pct	\$97,142 (\$92,726)	\$101,094	\$39,715
75^{th} pct	\$156,824 (\$178,466)	\$146,225	\$70,404

Table 5: Wealth-health gradient at age 60-64. The table reports the wealth difference between healthy and unhealthy people for each wealth quartile.

targeted wealth quartiles unconditional on health status. The second and fourth columns of Table 4 compare the $25^{th}, 50^{th}$, and 75^{th} wealth percentiles for the 60-64 age group in the PSID (and HRS) and in the "no-correlation" model, respectively. As a reference, we also report the corresponding statistics from the baseline model. The "no-correlation" model also matches the employment rate and the average labor income by health status conditional on working (the income-health gradient).

The "no-correlation" model, however, falls short of replicating the observed large differences in wealth by health status, as shown in Table 5. For example, the difference between the median wealth of the healthy and unhealthy near retirement is only \$40k in this model compared to about \$100k in the data (and baseline model).

We conclude from these findings that the direct effect of bad health (low earnings, high out-of-pocket medical expenses and shorter life expectancy) only partially accounts for the observed difference in accumulated wealth between the healthy and unhealthy.[47] This also means that the income-health gradient does not imply the wealth-health gradient. It should be noted that the wealth gap arising from the compositional difference is large even though we focus on a relatively homogeneous group of males with the same education level. In Appendix D we discuss additional aspects of the data on wealth and health, and how the composition difference can help account for them.

[47] Using the HRS, Poterba et al. (2017) construct a continuous health index for individuals aged 51-61 in 1992 and document that there is a large difference in asset growth between those in the top and bottom one-third of the health index. They also find that only 20-40% of the difference in asset growth can be attributed to the lower earnings and annuity income of those in poor health.

5.4 Model fit with respect to aspects of the data not targeted

In this section, we evaluate how well our baseline model performs along additional dimensions related to the health gradient which are not targeted in our estimation. Specifically, we ask whether it can replicate (i) the distribution of individuals by health conditional on income and wealth, and (ii) the joint dynamics of wealth and health. Capturing these aspects of the data is important to properly evaluate the long-term effects of bad health and the contribution of health to lifetime inequality

The first three columns of Tables 6 and 7 display the percentage of unhealthy people by income and wealth terciles. In the data, conditioning by age group, there are much more unhealthy people in the lowest terciles of earnings and wealth. Our model matches these additional features of the data well.

Next we check whether our baseline model can capture the joint dynamics of health and wealth for several years before retirement. To do this, we first document the relationship between wealth *changes* and the number of periods that individuals spend being unhealthy between ages 55-56 and 65-66 from the HRS. We use individuals who are healthy at age 55-56 and are observed every two years until they are 65-66 (this is the same balanced sample that we used to construct Figure 5 and Table 1). For each individual we compute the change in his wealth between ages 55-56 and 65-66. The crossed marks in Figure 9 plot the median wealth change conditional on the number of periods spent unhealthy between ages 57-58 and 65-66.

The data show that, among people who do not report being unhealthy in any period, median wealth increases by about \$50k over the ten-year period, while among individuals with the highest number of unhealthy periods (5) median wealth declines by about \$10k. The dotted marks in the figure show the corresponding statistics from our model and demonstrate that our model is successful at replicating the almost linear negative relationship between median wealth change and number of unhealthy periods during the 10 years before the retirement. It is worth noting that our estimation only targets *cross-sectional* moments of the wealth-health gradient; yet, our model can closely match the long-run *dynamic* aspect of this gradient. This gives us additional confidence that the estimated model is a good framework to evaluate the long-run effects of bad health.

6 Results

In our framework, bad health hurts individuals in a number of ways: it affects their earnings, medical spending, and life expectancy. The first two channels entail pecuniary losses while the third one lowers lifetime utility when individuals value their life. In this

	PSID (HRS)			Model		
	bottom 1/3	middle 1/3	top 1/3	bottom 1/3	middle 1/3	top 1/3
25-34	12%	5%	2%	16%	2%	0%
35-44	21%	8%	4%	22%	4%	2%
45-54	22%	12%	8%	28%	9%	5%
55-64	30% (36%)	15% (20%)	14% (13%)	33%	24%	11%

Table 6: Percentage of unhealthy individuals in each earnings tercile: data vs model.

	PSID (HRS)			Model		
	bottom 1/3	middle 1/3	top 1/3	bottom 1/3	middle 1/3	top 1/3
25-34	10%	10%	5%	8%	5%	3%
35-44	17%	10%	5%	14%	7%	5%
45-54	23%	13%	9%	24%	10%	8%
55-64	33% (36%)	17% (21%)	12% (14%)	34%	17%	13%
65-74	36% (38%)	26% (24%)	17% (16%)	41%	27%	19%
75+	46% (41%)	37% (29%)	24% (25%)	47%	38%	29%

Table 7: Percentage of unhealthy individuals in each wealth tercile: data vs model.

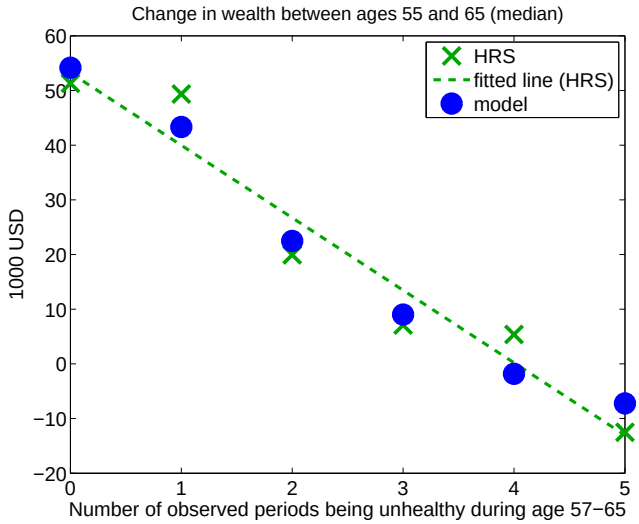


Figure 9: Median *wealth change* between age 55-56 and 65-66.

section, we evaluate both the pecuniary and non-pecuniary effects of bad health over the life-cycle.

To measure the loss in available resources, we start by constructing a comprehensive measure of the lifetime costs of bad health which includes both direct (medical spending) and indirect (labor earnings) losses. Next, we ask two questions to gauge the non-pecuniary effects of health. First, how much are individuals willing to pay to improve the probability of being in good health, and why? Second, how much does health uncertainty contribute to lifetime inequality?

6.1 Measuring the costs of bad health

In our model, the monetary costs of bad health arise from exogenous health-dependent stochastic processes (productivity, medical shock, and ESHI offer probability) and from behavioral responses (labor supply, insurance purchase, and savings). By summing these costs, we can get the measure of accumulated losses due to bad health an individual experiences over the working stage of the life-cycle.

Conceptually, to understand how costly it is for an individual to be unhealthy over his life course, we want to measure how much better off he would be in a counterfactual situation where his health is always good but everything else is the same. Our structural framework allows us to construct this counterfactual situation. In particular, we consider the following experiment. We place each individual in an environment where he unexpectedly draws good health realizations every period (becomes exceptionally lucky) while the realizations of other shocks over his life course are the same as in the baseline. Note that this approach has an advantage over estimating the costs of bad health directly from the data. In the latter case, we can only compare *different* individuals, healthy and unhealthy ones, after controlling for observables. As we showed earlier, healthy and unhealthy individuals differ in unobservables (health type and preferences) and this can bias the empirical estimates of the costs of bad health.

To formalize our approach, denote the earnings, net of total medical spending, of individual i at time t in the baseline and counterfactual cases as y_{it} and y_{it}^H, respectively. The difference between y_{it} and y_{it}^H represents the pecuniary costs of bad health in period t. Our measure of the lifetime costs of bad health averages these costs over the working stage of the life-cycle for those that survive to age 64 in the baseline economy:

$$loss_i = \frac{1}{45} \sum_{t=20}^{64} \left(y_{it}^H - y_{it} \right).$$

By focusing on individuals who survive to age 64 we ensure that our measure of the costs of bad health is not affected by mortality bias. For example, some unhealthy individuals may have low accumulated losses due to bad health because they die early.

Table 8 displays the lifetime costs of bad health by the number of years spent being unhealthy between ages 20 and 64. This table conveys several interesting findings. First, the costs of bad health quickly increase with the number of years spent being unhealthy. For example, an individual who spends between 1 and 5 years being unhealthy loses, on average, \$652 per year between ages 20 and 64. In contrast, an individual who spends between 16 and 20 years being unhealthy loses nearly \$5,000 per year, or 13.5% of the average annual

	Average	1-5 yrs	6-10 yrs	11-15yrs	16-20 yrs	> 20 yrs
Earning losses + total medical costs	\$1,243	\$652	\$2,195	\$3,608	\$4,865	\$6,788
(% of average earnings) [a]	*(3.4%)*	*(1.8%)*	*(6.1%)*	*(10%)*	*(13.5%)*	*(18.8%)*
Composition (%)						
- Medical costs paid by insurance	20%	12%	16%	23%	27%	30%
- Out-of-pocket medical costs	23%	27%	24%	21%	20%	19%
- Earnings losses	57%	61%	60%	56%	53%	51%

[a] Average earnings in our model is \$36,105.

Table 8: Annual loss due to bad health over ages 20-64, conditional on surviving to age 64, by the number of years spent being unhealthy.

	η_1	η_2	η_3	η_4	η_5
Earning losses + total medical costs	\$2,447	\$1,633	\$1,072	\$733	\$522
(% of average earnings) [a]	*(6.8%)*	*(4.5%)*	*(3.0%)*	*(2.0%)*	*(1.4%)*

[a] Average earnings in our model is \$36,105.

Table 9: Annual loss due to bad health over ages 20-64, conditional on surviving to age 64, by health type.

	top 5%	top 10%	top 20%
Losses due to bad health			
- *Incl.* costs paid by insurance	28%	46 %	71%
- *Excl.* costs paid by insurance	27%	45 %	72%

Table 10: Concentration of the lifetime costs of bad health by top quintiles of the distribution as percentage of aggregate costs.

earnings in our model.

Second, the largest component of these costs is the loss in earnings: on average, they account for 57% of the lifetime costs of bad health. For relatively healthy individuals (1 to 5 unhealthy years) it constitutes around 60% of total costs, while for those with the longest periods of sickness (more than 20 years) it is 51%.

Third, about 43% of the lifetime costs of bad health come from medical spending and approximately half of these are covered by insurance. Health insurance coverage turns out to increase with the number of unhealthy years: health insurance covers only 12% of the total lifetime costs of bad health for people who experience less than 5 unhealthy years, while this number is close to 30% for individuals with more than 16 unhealthy years.

Because individuals who experience long spells of bad health are much more likely to be of bad health types, the costs of bad health are unequally distributed across types. Table 9 shows that the average losses for η_5-individuals (the best health type) are about \$500 per year between ages 20 and 64, while the average losses for η_1-individuals (the worst type) are

about $2,500 per year, which represents 6.8% of average labor income in our model.

Another important finding is that the distribution of lifetime costs of bad health is highly concentrated (see Table 10): 20% of people with the highest lifetime costs due to bad health account for 71% of aggregate lifetime costs, while the top 5% account for 28%. Interestingly, even when the costs covered by health insurance are excluded from this measure, the concentration of costs remains high. Thus, even though health insurance plays an important role for people with very high lifetime costs, it does little to reduce the concentration of the total costs.

Overall, two important conclusions can be drawn from the results presented in this section. First, the lifetime costs of bad health can be very high and individuals are ex-ante different in their probability of ending up in the top end of the distribution; those born with bad health type have significantly higher costs over their life-cycle. Second, studies that confine the effects of bad health only to medical expenses can significantly underestimate the total losses that unhealthy people experience over their working lives.

6.2 Measuring the value of good health

We now turn to a more comprehensive measure of the costs of bad health, which takes into account that bad health also affects one's welfare by lowering his life expectancy, thus incorporating its non-pecuniary effect. Specifically, we consider the effects of health using individuals' valuation.

We start by computing working-age individuals' willingness to pay to increase their probability of being healthy next period by one percentage point. We assume that this change in health dynamics is temporary: it applies only to the probability of being in good health next period and after that the health transition probability returns to the status quo.

Table 11 displays the results of this experiment using the baseline economy. On average, working-age individuals are willing to pay $1,903 to increase their probability of being healthy next period, or 5.3% of average income in our model. There is considerable variation in the willingness to pay by health type: individuals with the worst health type (η_1) are willing to pay more than double of those with the best health type (η_5); $2,933 for the former vs $1,200 for the latter. Because the unhealthy state is much more persistent among individuals with bad health types, they are willing to pay more to decrease the probability of becoming sick.

Another important observation from Table 11 is that, within a particular health type, asset-rich individuals are willing to pay much more than asset-poor ones. The ratio of the willingness to pay between the top and bottom asset terciles is ranging from 2.1 for η_1-individuals to 2.8 for η_5-individuals. The higher willingness to pay of the rich can be explained by two forces. First, given their lower marginal utility of current consumption

	$\eta_1 - \eta_5$	η_1	η_2	η_3	η_4	η_5
Willingness to pay	\$1,903	\$2,933	\$2,352	\$1,718	\$1,359	\$1,200
(% of average earnings)[a]	*(5.3%)*	*(8.1%)*	*(6.5%)*	*(4.8%)*	*(3.8%)*	*(3.3%)*
By asset terciles[b]						
- 1^{st} Tercile	\$1,333	\$1,958	\$1,474	\$1,037	\$760	\$560
- 2^{nd} Tercile	\$1,770	\$3,175	\$2,354	\$1,528	\$963	\$710
- 3^{rd} Tercile	\$2,453	\$4,199	\$3,552	\$2,606	\$1922	\$1,592

[a] Average earnings in our model is \$36,105.

[b] Asset terciles are calculated from the wealth distribution of individuals aged 20-64.

Table 11: Willingness to pay to increase the probability of being healthy next period by one percentage point (among age 20-64) in the baseline economy.

	$\eta_1 - \eta_5$	η_1	η_2	η_3	η_4	η_5
Baseline economy	\$1,903	\$2,933	\$2,352	\$1,718	\$1,359	\$1,200
Dollar value when only one channel exists						
- Survival channel	60%	52%	56%	61%	69%	74%
- Labor market channel	36%	45%	40%	34%	26%	22%
- Medical expenses channel	5%	5%	5%	4%	4%	4%

Table 12: Decomposition of the willingness to pay for an increase in the probability of being healthy next period. All dollar values when only one channel exists are the averages over the distribution of individuals aged 20-64 in the baseline. We report the resulting values as a percentage of the dollar value when all channels operate (baseline).

	1^{st} Tercile	2^{nd} Tercile	3^{rd} Tercile
Baseline economy	\$1,333	\$1,770	\$2,453
Dollar value when only one channel exists			
- Survival channel	35%	47%	78%
- Labor market channel	58%	45%	21%
- Medical expenses channel	7%	5%	3%

Table 13: Decomposition of the willingness to pay for an increase in the probability of being healthy next period by asset tercile. The reported values are a percentage of the dollar value when all channels operate (baseline).

they are willing to give up more of their resources today to improve health in the future. Second, better health lengthens one's lifespan and the rich can enjoy the additional years of life more since they can secure a higher consumption flow.[48]

Next, to separate the non-pecuniary effects from the monetary ones, we measure how

[48] In general, asset-rich and asset-poor individuals differ in terms of average age. However, the described pattern still holds even when we control for age.

much individuals are willing to pay for the improved health dynamics when the effects of health are limited to only one channel; namely, life expectancy, labor market outcomes, or medical spending. To do this we consider three counterfactual experiments. In the first experiment, we assume that bad health *only* affects one's survival probability, i.e., individuals who become sick experience a decline in their life expectancy but no change in their productivity, disutility from work or medical spending. In the second experiment, we assume that bad health *only* affects one's productivity and disutility from work, i.e., there is no effect on life expectancy or medical spending. In the third experiment, we assume that bad health *only* affects one's medical spending. Note that in the last experiment, health affects not only the distribution of total medical costs but also insurance premiums in the individual market and the probability of getting ESHI. In each of these three experiments, we reevaluate how much individuals in the baseline economy are willing to pay to increase the probability of being healthy next period by one percentage point.[49]

Table 12 displays the results of these experiments expressed as a percentage of the willingness to pay in the baseline economy.[50] The results imply that the most valuable aspect of being healthy is having a longer life expectancy: the second row of Table 12 shows that when bad health only affects longevity the willingness to pay to improve health dynamics constitutes 60% of the baseline level. As for the monetary effects of health, consistent with the results in the previous subsection, the labor market channel is significantly more important than the medical spending channel. The willingness to pay when only the former channel exists amounts to 36% of the baseline level, compared to just 5% when only the latter channel exists.

There is variation in the relative importance of pecuniary vs non-pecuniary effects by health types (columns 3 to 7 of Table 12). For better health types, the life expectancy channel is far more important than the monetary channel. This variation is driven by the fact that η_5- and η_4-individuals tend to have more assets. Table 13 shows that individuals with high assets mostly care about the longevity aspect of good health, while those with low assets put a large weight on the earning aspect. This is because rich individuals, on average, are better insured against the monetary consequences of bad health; they also enjoy longer lives more since they can afford to consume more per period.[51]

[49] In these three experiments, bad health affects individuals through only one channel from the next period onward while the effect of bad health in the current period is the same as in the baseline. This ensures that each individual has the same amount of resources and faces the same immediate survival probability (from the current period to the next) as in the baseline, so that the dollar value of the willingness to pay in each counterfactual experiment is comparable to this value in the baseline economy.

[50] Our decomposition exercise is not supposed to sum to 100% by construction. The purpose of this exercise is to rank the importance of each channel through which health affects individuals.

[51] This pattern holds even when we look into the same age group.

6.3 Health and lifetime inequality

Our previous results show that the lifetime pecuniary costs of bad health can be very high but what dominates people's valuations of good health are its non-pecuniary benefits. In this section we ask how much bad health affects lifetime inequality.

To capture both the pecuniary and non-pecuniary aspects of good health, we evaluate lifetime inequality by comparing the variation in lifetime utility in our baseline economy and in the counterfactual economy where everyone always (unexpectedly) draws good health realization while still receives all other shocks as in the baseline.[52]

Let denote the lifetime utility of an individual i in the baseline economy as:

$$U_i = \sum_{t=20}^{\text{age of death}+1} \beta_i^{t-20} \left(u(c_t, l_t, h_t) \times \mathbf{1}_{alive_t} + \theta_{Beq} \left(\frac{k_t + k_{Beq}}{1-\rho} \right)^{1-\rho} \times (1 - \mathbf{1}_{alive_t}) \right).$$

Denote $Var(U_i)$ as the variance of lifetime utility in the baseline. Denote the lifetime utility in the counterfactual case as U_i^H and its variance as $Var(U_i^H)$. The share of lifetime inequality due to health uncertainty is the percentage reduction in $Var(U_i)$ compared to $Var(U_i^H)$:

$$\left(1 - \frac{Var\left(U_i^H\right)}{Var\left(U_i\right)} \right) \times 100\%. \tag{21}$$

The results of this exercise are shown in the second and third columns of Table 14. Since our inequality measure depends on the rate of time preferences, we report the calculation separately for individuals with β_{low} and β_{high}.

Once removing bad health realizations, lifetime inequality is significantly lower: the percentage decline ranges from 14% of the baseline case among the β_{high}-types to 47% among the β_{low}-types. In addition, for both groups, the effect of health uncertainty on inequality is larger among individuals with worse health types (η_1, η_2) compared to the effect on better health types (η_3, η_4, η_5). As shown in Figure 6, a non-trivial fraction of η_1- and η_2-individuals endures multiple unhealthy years; consequently, there is larger variation in monetary losses and lifespans within this group. Once bad health realizations are removed, the variation in their lifetime utility decreases more.

Health uncertainty affects lifetime inequality through two channels. First, it increases the variation in lifetime resources, which affects consumption and leisure. Second, it raises the variation in lifespans, which directly affects utility. Figure 10 shows the distribution of the age of death in the baseline and the counterfactual economy where everyone is always

[52] In the counterfactual economy, some people will live longer once we remove bad health realizations. During their extended life, they are always healthy and draw their labor productivity, medical expense, ESHI offer, and survival probability from the corresponding shock processes.

	Incl. survival channel		*Excl.* survival channel	
	β_{low}	β_{high}	β_{low}	β_{high}
All η_i	47%	14%	24%	4%
$\Rightarrow \{\eta_1, \eta_2\}$	54%	25%	28%	11%
$\Rightarrow \{\eta_3, \eta_4, \eta_5\}$	30%	10%	11%	1%

Table 14: Variation of lifetime utility due to health uncertainties. The columns under "*Incl.* survival channel" are from the experiment where everyone is always healthy. The columns under "*Excl.* survival channel" are from the experiment where individuals are always healthy but they die at the same age as in the baseline.

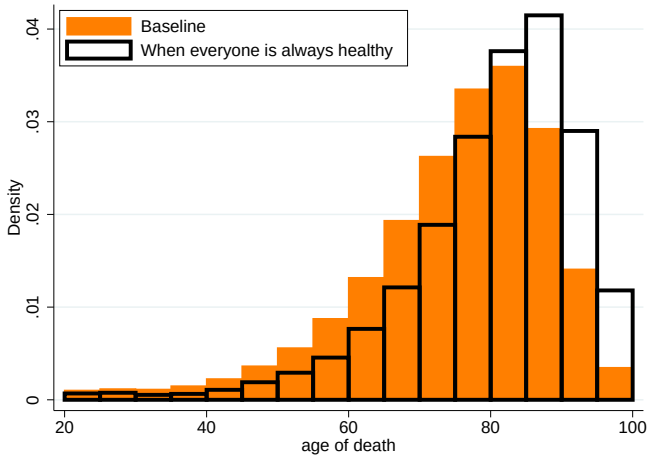


Figure 10: Distribution of age of death. The shaded bars are from the baseline. The white bars are from the counterfactual experiment where all individuals are always healthy and their age of death is allowed to adjust.

healthy. The distribution in the counterfactual case shifts toward the maximum lifespan, 99 years old, and becomes less dispersed. The median age of death shifts up from age 77 in the baseline to age 82 in the counterfactual economy.

To separate the two channels, we simulate an additional counterfactual experiment in which everyone is always healthy but individuals' age of death is fixed as in the baseline. The percentage reduction in the baseline variance of lifetime utility is reported in the last two columns of Table 14. A striking difference from the previous experiment is that the variance decreases much less: only by 4% of the baseline for the β_{high} group and 24% for the β_{low} group. Thus, the large portion of the decline in lifetime inequality reported in the second and third columns of the same table comes from the lower variation in the age of death.

Overall, our results in this section show that the contribution of bad health to lifetime inequality is substantial. Consistent with our findings in Section 6.2 an important mechanism through which health uncertainty maps into lifetime inequality is by changing the length of

life (survival channel). Thus, focusing only on the monetary losses will underestimate the lifetime inequality due to bad health by half or more.

7 Conclusion

In this paper we develop a structural framework for measuring the lifetime consequences of bad health. We first estimate a model of health dynamics using a rich set of data moments and allowing for history-dependence and fixed heterogeneity. We find that fixed heterogeneity is important to account for the persistence of bad health. We then incorporate this estimated health process into a life-cycle model with incomplete markets. The estimated model can replicate important facts related to health and economic outcomes, including the income-health and wealth-health gradients. We show that the wealth-health gradient does not naturally follow from the income-health gradient and the fact that the unhealthy have higher medical spending and lower life expectancy. The large gap in wealth by health is also driven by the compositional difference between the healthy and unhealthy; specifically, on average, the latter group includes more people with a lower propensity to save.

We use our estimated model to quantify the pecuniary and non-pecuniary effects of bad health over the life-cycle. We find that the monetary costs of bad health quickly accumulate as working-age individuals spend more years being unhealthy and the largest component of these costs is the loss in labor earnings, while the contribution of out-of-pocket medical spending is much smaller. However, when taking into account the non-monetary aspect of health, the most detrimental consequence of being unhealthy is lower life expectancy as evidenced by two findings. First, individuals are willing to pay a substantial amount to access technology that increases their chance to be healthy; 60% of the willingness to pay is attributed to the fact that good health can extend life expectancy. Second, we find that health uncertainty can account for a substantial share of lifetime inequality even among individuals within the same educational group, largely by affecting the variation in individuals' lifespans.

Overall, we show that a life-cycle model can account for many cross-sectional and dynamic aspects of socioeconomic status-health gradient when taking into account the correlation between predetermined factors affecting health outcomes and preferences underlying economic decisions. Our structural framework can serve as an important building block for further examining issues in health-related inequality. The growing body of studies about early childhood interventions and the recent advancements in the use of genetic information in economic research allow us to better understand the importance of childhood circumstances and genetic endowments in shaping health and economic outcomes in adulthood. Incor-

porating these mechanisms into a structural framework will be a valuable tool to address many important policy questions. For example, to what extent early childhood interventions can reduce health-related inequality or, more broadly, the intergenerational persistence of lifetime inequality.

References

[1] Ameriks, J., Briggs, J., Caplin, A., Shapiro, M., Tonetti, C., 2017. Long-Term Care Utility and Late in Life Saving. NBER working paper.

[2] Attanasio, O., Kitao, S., Violante G., 2011. Financing Medicare: A General Equilibrium Analysis. In Shoven, J.B. (Ed.), Demography and the Economy, University of Chicago Press..

[3] Barth, D., Papageorge, N., Thom, K. 2017. Genetic Ability, Wealth, and Financial Decision-Making. IZA Discussion Paper.

[4] Blundell, R., Britton, J., Dias, M., French, E., 2016. The Dynamic Effects of Health on the Employment of Older Workers. Michigan Retirement Research Center Working Paper.

[5] Capatina, E., 2015. Life-cycle Effects of Health Risk. Journal of Monetary Economics, 74, pp.67-88.

[6] Capatina, E., Kene, M., Maruyama, S., 2017. Health and Earning Dynamics over the Life-Cycle. Mimeo, University of New South Wales.

[7] Carroll, C., Slacalek, J., Tokuoka, K., White, M., 2017. The Distribution of Wealth and the Marginal Propensity to Consume, Quantitative Economics, 8(3), pp.977-1020.

[8] Case, A., Lubotsky, D., Paxson, C., 2002. Economic Status and Health in Childhood: The Origins of the Gradient. American Economic Review, 92, pp.1308-1664.

[9] Case, A., Fertig, A., Paxson, C., 2005. The Lasting Impact of Childhood Health and Circumstance. Journal of Health Economics, 24(2), pp.365-389.

[10] Case, A., Paxson, C., 2010. Causes and Consequences of Early-Life Health. Demography, 47, pp.S65-S85.

[11] Cole, H., Kim, S., Krueger, D., 2018. Analyzing the Effects of Insuring Health Risks: On the Trade-off between Short-Run Insurance Benefits vs. Long-Run Incentive Costs. Review of Economic Studies (forthcoming).

[12] Conesa, J., Costa, D., Kamali, P., Kehoe, T., Nygard, V., Raveendranathan, G., Saxena, A., 2018. Macroeconomic Effects of Medicare, Journal of the Economics of Ageing, 11, pp.27-40.

[13] Cronqvist, H., Siegel, S., 2015. The Origins of Savings Behavior, Journal of Political Economy, 123(1), pp.123-169.

[14] Cutler, D., Lleras-Muney, A., and Vogl, T., 2011. Socioeconomic Status and Health: Dimensions and Mechanisms. In Glied, S. and Smith, P. (Eds), Oxford Handbook of Health Economics, Oxford University Press.

[15] De Nardi, M., 2004. Wealth Inequality and Intergenerational Links. Review of Economic Studies, 71, pp.743-768.

[16] De Nardi, M., French, E., Jones, J., 2010. Why Do the Elderly Save? Journal of Political Economy, 118(1), pp.39-75.

[17] De Nardi, M., French, E., Jones, J., 2016. Medicaid Insurance in Old Age. American Economic Review, 106(11), pp.3480-3520

[18] Dobkin, C., Finkelstein, A., Kluender, R., Notowidigdo, M., 2016. The Economic Consequences of Hospital Admissions, NBER Working Papers.

[19] Erosa, A., Fuster, L., Kambourov, G., 2016. Towards a Micro-founded Theory of Aggregate Labour Supply. Review of Economic Studies, 83(6), pp.1001-1039.

[20] Epper, T., Fehr, E., Fehr-Duda, H., Kreiner, C., Lassen, D., Leth-Petersen, S., Rasmussen, G., 2018. Time Discounting and Wealth Inequality, Working paper.

[21] Finkelstein, A., Luttmer, E., Notowidigdo, M., 2009. Approaches to Estimating the Health State Dependence of the Utility Function, American Economic Review: Papers & Proceedings, 99(2), pp.116-121.

[22] Floden, M., 2008. A Note on the Accuracy of Markov-chain Approximations to Highly Persistent AR(1) Processes. Economic Letters, 99(3), pp.516-520.

[23] French, E., 2005. The Effects of Health, Wealth, and Wages on Labor Supply and Retirement Behavior. Review of Economic Studies, 72(2), pp.395-427.

[24] French, E., Jones, J., 2011. The Effects of Health Insurance and Self-Insurance on Retirement Behavior. Econometrica, 79(3), pp.693-732.

[25] Gouveia, M. and Strauss, R. P., 1994. Effective Federal Individual Tax Functions: An Exploratory Empirical Analysis, National Tax Journal, 47(2), pp.317-39.

[26] Hall, R., Jones, C., 2007. The Value of Life and the Rise in Health Spending. Quarterly Journal of Economics, 122(1), pp.39-72.

[27] Halliday, T., 2008. Heterogeneity, State Dependence and Health. Econometrics Journal, 11(3), pp.499-516.

[28] Hendricks, L., 2007a. Retirement Wealth and Lifetime Earnings, International Economic Review, 48(2), pp.421-456.

[29] Hendricks, L., 2007b. How Important is Discount Rate Heterogeneity for Wealth Inequality?, Journal of Economic Dynamics & Control 31, pp.3042-3068.

[30] Huggett, M., Ventura, G., Yaron, A., 2010. Sources of Lifetime Inequality. American Economic Review, 101(7), pp.2923-2954.

[31] Idler, E., and Benyamini, Y., 1997. Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies. Journal of Health and Social Behavior, Vol 38, pp.21-37.

[32] Jeske, K., Kitao, S., 2009. U.S. Tax Policy and Health Insurance Demand: Can a Regressive Policy Improve Welfare? Journal of Monetary Economics, 56(2), pp.210-221.

[33] Kaiser Family Foundation, 2004., The Cost of Care for the Uninsured. What Do We Spend, Who Pays, and What Would Full Coverage Add to Medical Spending?. Available at http://www.kff.org/uninsured/upload/The-Cost-of-Care-for-the-Uninsured-What-Do-We-Spend-Who-Pays-and-What-Would-Full-Coverage-Add-to-Medical-Spending.pdf.

[34] Kaplan, G., 2012. Inequality and the Life Cycle. Quantitative Economics, 3, pp.471525.

[35] Keane, M., Wolpin, K., 1997. The Career Decisions of Young Men. Journal of Polictical Economy, 105(3), pp.473-522.

[36] Krueger, D., Mitman, K., Perri, F., 2016. Macroeconomics and Household Heterogeneity. In Taylor, J., Uhlig, H., (Eds.), Handbook of Macroeconomics Volume 2, Elsevier.

[37] Krusell, P., Smith, A., 1998. Income and Wealth Heterogeneity in the Macroeconomy, Journal of Political Economy, 106(5), pp.867-896.

[38] Lange, F., McKee, D., 2012. The Evolution of Latent Health over the Life Course. Mimeo, Yale University.

[39] Lawrance, E., 1991. Poverty and the Rate of Time Preference: Evidence from Panel Data, Journal of Political Economy, 99, pp.54-77.

[40] Lockwood, L., 2014. Incidental Bequests: Bequest Motives and the Choice to Self-Insure Late-Life Risks, NBER working paper.

[41] Lundborg, P., Nilsson, M., Vikstrom, J., 2015. Heterogeneity in the Impact of Health Shocks on Labour Outcomes: evidence from Swedish workers. Oxford Economic Papers, 67(3), pp.715-739.

[42] Murphy, K., Topel, R., 2006. The Value of health and Longevity, Journal of Political Economy, 114(5), pp.871-904.

[43] Nakajima, M., Telyukova, I., 2011. Home Equity in Retirement. Working paper, Federal Reserve Bank of Philladephia,

[44] Ozkan, S., 2014. Preventive vs. Curative Medicine: A Macroeconomic Analysis of Health Care over the Life Cycle. Mimeo, University of Toronto.

[45] Papageorge, N., Thom, K., 2016. Genes, Education, and Labor Market Outcomes : Evidence from the Health and Retirement Study., IZA Discussion Paper.

[46] Pashchenko, S., 2013. Accounting for Non-annuitization. Journal of Public Economics 98, pp.53-67.

[47] Pashchenko, S., Porapakkarm, P., 2013. Quantitative Analysis of Health Insurance Reform: Separating Regulation from Redistribution. Review of Economic Dynamics 16, pp.383-404.

[48] Pashchenko, S., Porapakkarm, P., 2016a. Cross-subsidization in Employer-based Health Insurance and the Effects of Tax Subsidy Reform. National Tax Journal, 69(3), pp.583-612

[49] Pashchenko, S., Porapakkarm, P., 2016b. Medical Spending in the U.S.: Facts from the Medical Expenditure Panel Survey Dataset. Fiscal Studies, 37(3-4), pp.689-716.

[50] Pashchenko, S., Porapakkarm, P., 2017. Work Incentives of Medicaid Beneficiaries and the Role of Asset Testing. International Economic Review, 58(4), pp.1117-1154.

[51] Pijoan-Mas, J. and Ríos Rull, V., 2014. Heterogeneity in Expected Longevities, Demography, 51(6), pp.2075-2102.

[52] Parro, F., and Pohl, V., 2017. Health Shocks, Human Capital, and Labor Market Outcomes. Mimeo, University of Georgia.

[53] Poterba, J., Venti, S., Wise, D., 2017. The Asset Cost of Poor Health, The Journal of the Economics of Ageing, 9, pp.172-184.

[54] Robinson, L., 2007. How US Government Agencies Value Mortality Risk Reductions, Review of Environmental Economics and Policy, 1(2), pp.283-299.

[55] Rust, J., Phelan, C, 1997. How Social Security and Medicare Affect Retirement Behavior In a World of Incomplete Markets, Econometrica, 65(4), pp.781-831.

[56] Samwick, A., 1998. Discount Rate Heterogeneity and Social Security Reform, Journal of Development Economics, 57(1), pp.117-146.

[57] Smith, J., 1999. Healthy Bodies and Thick Wallets: The Dual Relation Between Health and Economic Status. Journal of Economic Perspectives, 13(2), pp.144-166.

[58] Storesletten, K., Telmer, C., Yaron, Y., 2004. Consumption and Risk Sharing Over the Life Cycle. Journal of Monetary Economics, 51(3), pp.609-633.

[59] Van Doorsaler, E., and Gerdtham, U., 2002. Does Inequality in Self-Assessed Health Predict Inequality in Survival by Income? - Evidence from Swedish Data. Social Science and Medicine, 57, pp.1621-1629.

[60] Viscusi, K., 1993. The Value of Risks to Life and Health. Journal of Economic Literature, 31(4), pp.1912-1946.

[61] Warner, J., Pleeter, S., 2001. The Personal Discount Rate: Evidence from Military Downsizing Programs, American Economic Review, 91, pp.33-53.

Appendix

A Polygenic scores and health outcomes in HRS

Here we use the same HRS sample as in Table 2 and report the $25^{th}, 50^{th}$, and 75^{th} percentiles of the three Polygenic scores among individuals reporting different unhealthy periods. Consistent to the average scores in Table 2, among individuals reporting 4-5 unhealthy periods all percentiles of the scores for educational attainment are noticeable lower while all percentiles of the scores for smoking and BMI are higher than the corresponding percentiles of those reporting at most one unhealthy period.

# unhealthy periods	Polygenic scores (HRS)		
	educational attainment	smoking	BMI
	25^{th} *percentile*		
0-1	-0.719	-0.692	-0.649
4-5	-1.657	-0.431	-0.572
	50^{th} *percentile*		
0-1	-0.120	0.020	-0.023
4-5	-0.579	0.324	0.435
	75^{th} *percentile*		
0-1	0.483	0.672	0.618
4-5	0.276	0.842	1.122

Table 15: The $25^{th}, 50^{th}$, and 75^{th} percentiles of Polygenic scores by the number of unhealthy periods between ages 57-58 and 65-66. All individuals are healthy at age 55-56.

B Medical shocks and insurance coverage

To estimate medical expenses, we follow Pashchenko and Porapkkarm (2017). First the medical expenses in the MEPS are converted into 1996 price using the CPI, then, we separate our sample into 12 age groups (20-24, 25-29, 30-34, ..., 75+). We assign the age of each group to the mid-point of the corresponding age interval. For example, 22 for 20-24, 27 for 25-29, 32 for 30-34, etc. For each year, we divide medical expenditures into 3 bins corresponding to the bottom 50th, 50-95th, and top 5th percentiles for each health status and age group. To obtain a value of medical expenses in each bin, we run a regression of medical expenses on a set of age group and year dummies. The coefficients on age dummies in this regression are the average medical expenses for the corresponding age and health in a particular bin. The resulting numbers are multiplied by 1.60 for people younger than 65 years old and by 1.90 for people who are 65 or older to make medical spending in our model consistent with

the aggregate medical spending in the NHEA as explained in Section 5.1.3. Then, we fit our estimated coefficients with a quadratic function of age. Figure (11) shows the medical costs for each grid separately for healthy and unhealthy individuals.

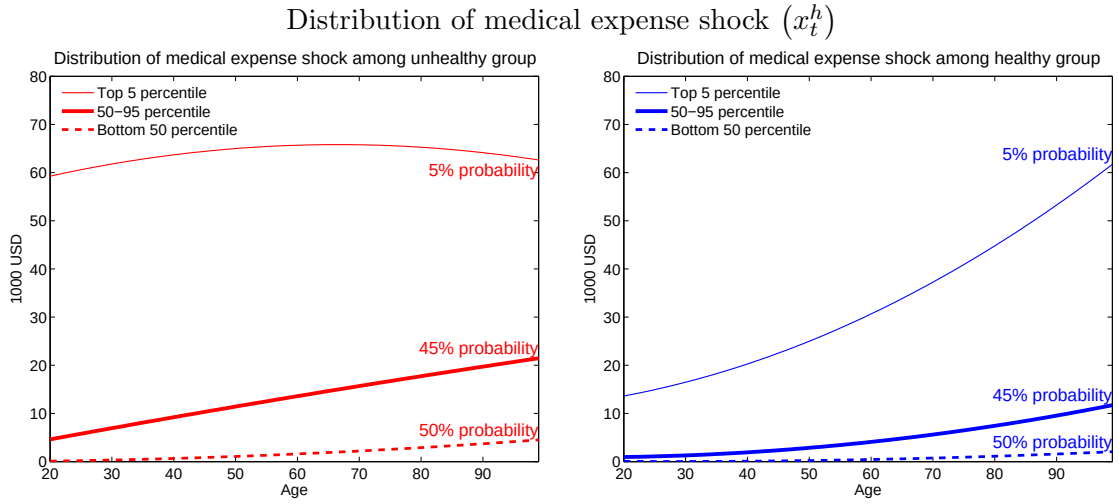


Figure 11: Medical expense grids by health status, x_t^h.

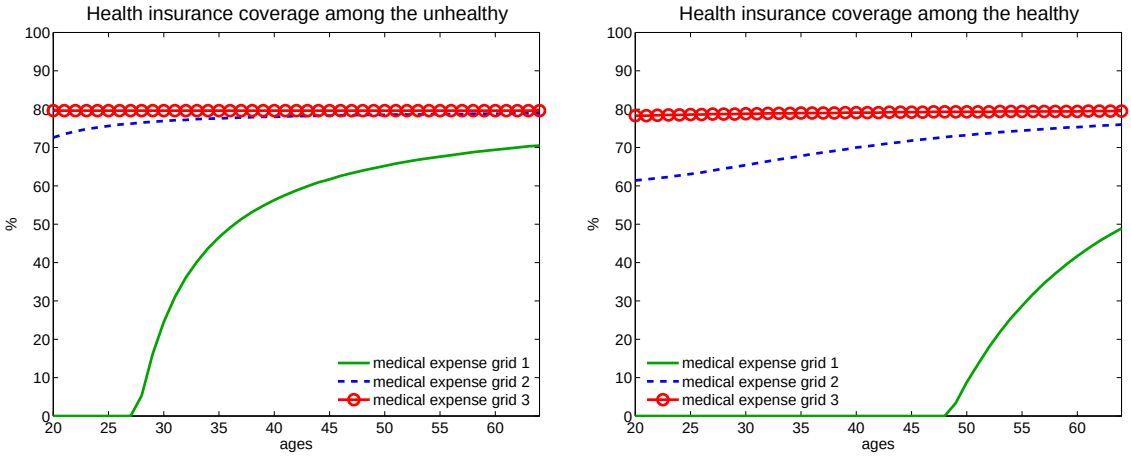


Figure 12: Private health insurance coverages: $cvg\left(x_t^h, i_H\right), i_H \in \{1, 2\}$.

To determine the fraction of medical expenses covered by private insurance $cvg(x_t^h, i_H)$ where $i_H \in \{1, 2\}$, we do the following. We estimate medical expenditures paid by private insurers as a function of total medical expenditures and year dummy variables using only individuals who are categorized as individually insured or group-insured. Then, we convert our estimates into the fraction of expenditures covered by insurers. Figure 12 shows the estimated coverages by medical expense grids.

C Algorithm for the second step estimation

To implement our sequential estimation strategy in Section 5.2.1, we divide the parameters to be estimated into two sets: $\boldsymbol{\theta_1} = \{\phi^W, \phi^B, \lambda_t^h\}$ and $\boldsymbol{\theta_2} = \{\beta_{low/high}, Pr\left(\beta_{low}|\eta_i\right), \theta_{Beq}, k_{Beq}, \overline{c}\}$. $\boldsymbol{\theta_1}$ is the set of parameters directly affecting the labor market outcomes (first set of moments) while $\boldsymbol{\theta_2}$ is the set of parameters closely linked to the wealth profiles (second set of moments). Our search algorithm is separated into two nested loops. In the outer loop, we use the simplex method to find $\boldsymbol{\theta_2}$ that minimizes the difference between the vector of the targeted and simulated second set of moments in Figure 8 ($\mathcal{M}_2^D$ and $\mathcal{M}_2^S$, respectively). In the inner loop, for a given set of $\boldsymbol{\theta_2}$, we search for $\boldsymbol{\theta_1}$ that minimizes the distance between the targeted and simulated first set of moments in Figure 7 ($\mathcal{M}_1^D$ and $\mathcal{M}_1^S$, respectively).

Formally, our algorithm solves the following problem:

$$\min_{\boldsymbol{\theta_2}} \ \Big(\mathcal{M}_2^D - \mathcal{M}_2^S\left(\boldsymbol{\theta_1^*}, \boldsymbol{\theta_2}\right)\Big)'\Big(\mathcal{M}_2^D - \mathcal{M}_2^S\left(\boldsymbol{\theta_1^*}, \boldsymbol{\theta_2}\right)\Big) \tag{22}$$

subject to

$$\boldsymbol{\theta_1^*} = \underset{\boldsymbol{\theta_1}}{\operatorname{argmin}} \ \big(\mathcal{M}_1^D - \mathcal{M}_1^S\left(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}\right)\big)'\big(\mathcal{M}_1^D - \mathcal{M}_1^S\left(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}\right)\big), \tag{23}$$

$$\text{and} \qquad \left\|\mathcal{M}_1^D - \mathcal{M}_1^S\left(\boldsymbol{\theta_1^*}, \boldsymbol{\theta_2}\right)\right\|_\infty \ \leq \ tol. \tag{24}$$

The additional constraint in Equation (24) is to ensure that the difference between the targeted and simulated moments from the first set (labor market outcomes) is not larger than a certain tolerance level *tol*.[53,54]

D Wealth level and number of years being unhealthy

In Section 5.4, we document the negative relationship between wealth *change* and the number of periods being unhealthy over a ten-year interval between the age of 55-56 and 65-66. In this section, we consider an additional dimension by looking at the wealth *level* at the age of 55-56 and 65-66. Similar to Section 5.4, we use the balanced panel of individuals from the HRS whom we observe between ages 55-56 and 65-66 and who are healthy at age 55-56. These additional moments serve as external validation and supporting evidence for the existence of the compositional effect in the data.

[53] The infinite norm $\|\cdot\|_\infty$ measures the largest gap between the targeted and simulated moments. Specifically, $\|\boldsymbol{x}\|_\infty = \max_j |x_j|$, where x_j is the element of vector $\boldsymbol{x}$.

[54] More specifically, we have two tolerance levels: one for the average labor income of workers and another for the employment profiles. We set the tolerance level of the former to \$72 (or 0.2% of average income in the model), while the latter is set to 8 percentage points. Different tolerance levels that we tried did not affect our estimation results in any meaningful way.

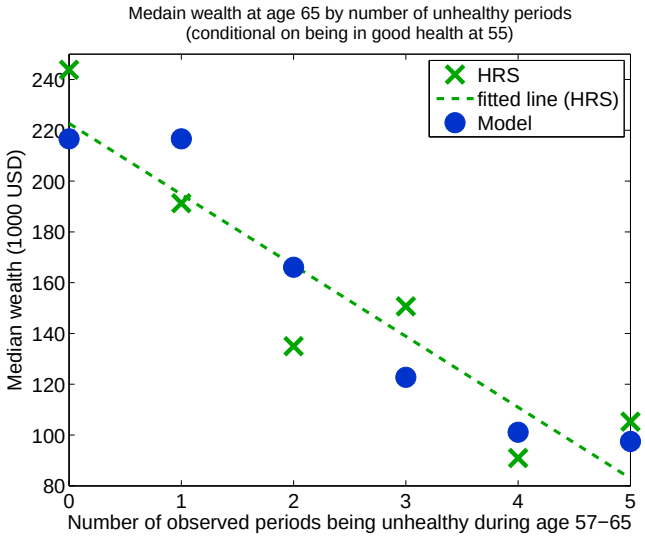
The crosses in Figure 13 (and 14) plot the median wealth *level* at age 65-66 (and 55-56) by number of unhealthy periods reported by individuals between ages 57-58 and 65-66. The dots in the figures show that the corresponding median wealth level from our simulated data can replicate these dynamic aspects well. The tables on the right of the figures report the slope coefficients from the median regressions. For each additional period an individual reports being unhealthy between 57-58 and 65-66, median wealth at age 65-66 (and 55-56) declines by \$34,473 (and \$11,749).[55] The corresponding coefficients from the model imply a decline of \$27,981 and \$10,831, respectively.

In Figure 13, the negative relationship between wealth level at 65-66 and number of unhealthy periods is generated by two mechanisms: direct causality and compositional difference. For the former mechanism, the multiple periods of bad health affect the accumulated wealth directly by lowering earnings, increasing medical expenses, and lowering survival probability. For the latter mechanism, people with multiple unhealthy periods are more likely to be β_{low}-type. The last two rows in the table on the right of Figure 13 show that once we run the median regressions separately for the β_{low} and β_{high} groups, the direct causality accounts for around 54-65% of the negative relationship between wealth level at 65-66 and number of unhealthy periods between ages 57-58 and 65-66.

In contrast, the negative relationship between wealth level at 55-56 and number of unhealthy periods in Figure 14 cannot be explained by causality; instead, it is driven only by the compositional difference. Since health shock is exogenous, the realization of future health status is orthogonal to current wealth. As shown in the table on the right, the slope coefficients from the separate median regressions over the β_{low} and β_{high} groups are zero.[56] This reiterates our results from Section 5.3: without the compositional difference, the model will underpredict the difference in wealth level among people experiencing different numbers of years in bad health.

[55] The negative relationship at age 55-56 is robust even when we restrict our analysis to a sample with a more homogeneous self-reported health measure. Specifically, we run the same median regression using a subset of our sample whose self-reported health at age 55-56 is either *excellent* or *very good*, thus excluding those with self-reported health as *good*. The slope coefficient of the median wealth at age 55-56 shows a decline of \$12,531 for one additional period of being unhealthy.

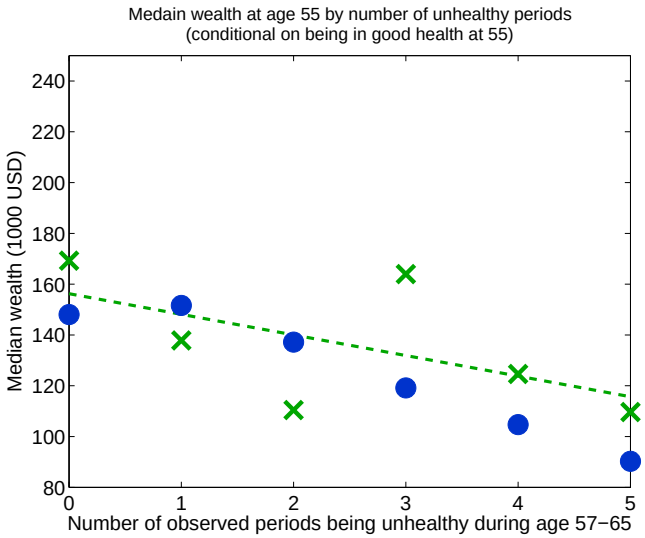
[56] It is possible that wealthy individuals invest more in their health than less wealthy ones; consequently they are less likely to experience multiple periods of bad health later on. To separate this channel from the compositional difference we need a model with health investment, which is beyond the scope of our current study. However, the dynamic moments in Figure 13 and 14 are useful for a structural model with endogenous health investment; thus far these moments have not been exploited in the existing literature.



Slope coeff. from median regression	
- HRS	$-34,473^{**}$
- Baseline model	
all	$-27,981$
β_{low}	$-15,164$
β_{high}	$-18,052$

** 5% significant level

Figure 13: Median *wealth level* at *65-66*.



Slope coeff. from median regression	
- HRS	$-11,749^{**}$
- Baseline model	
all	$-10,831$
β_{low}	0
β_{high}	0

** 5% significant level

Figure 14: Median *wealth level* at *55-56*.

E Alternative calibration with SVL at $6 millions

In our baseline parameterization, we adjust the scaling parameter $\bar{b}$ so that the statistical value of life (SVL) implied by our model is equal to $2 millions. In this section, we report the results from an alternative parameterization when $\bar{b}$ is set to match the SVL of $6 millions. Note that all other parameters in the model are the same as in the baseline.

Table 16 displays the willingness to pay to increase the probability of being healthy next period by one percentage point, and Table 17 and 18 decompose this willingness to pay by different channels through which health affects individuals over the life course. Overall, two important differences can be noticed comparing to our baseline results (Table 11-13). First, individuals are willing to pay more to increase their probability to be healthy: the willingness to pay is now 10.6% of the average income, compared to 5.3% under the baseline parameterization. Second, the relative importance of the survival channel increases: this channel now accounts for 86% of the total willingness to pay, compared to 60% under the baseline parameterization. Thus, when life is more valuable people are willing to pay more to improve their health dynamics because the increase in life expectancy due to good health is now more important.

Table 19 reports the corresponding results in Section 6.3 when the SVL is set to $6 millions. Once removing the bad health realization, the variation in lifetime utility is lowered by 42.5% for β_{low}- and 12.8% and β_{high}-individuals. And when excluding the survival channel the decreases is only 7.4% and 0.2%, respectively. This confirms the finding in Section 6.3 that a larger share of lifetime inequality due to health uncertainty comes from the survival channel.

	$\eta_1 - \eta_5$	η_1	η_2	η_3	η_4	η_5
Willingness to pay	\$3,828	\$5,113	\$4,395	\$3,506	\$3,157	\$3,026
(% of average earnings)[a]	*(10.6%)*	*(14.1%)*	*(12.2%)*	*(9.7%)*	*(8.7%)*	*(8.4%)*
By asset terciles[b]						
- 1^{st} Tercile	\$1,833	\$2,636	\$2,021	\$1,445	\$1,087	\$857
- 2^{nd} Tercile	\$2,982	\$5,059	\$3,852	\$2,611	\$1,783	\$1,428
- 3^{rd} Tercile	\$6,107	\$9,369	\$8,377	\$6,555	\$5,110	\$4,317

[a] Average earnings in our model is \$36,105.

[b] Asset terciles are calculated from the wealth distribution of individuals aged 20-64.

Table 16: Willingness to pay to increase the probability of being healthy next period by one percentage point (among age 20-64) *when the statistical value of life is set at \$6 millions.*

	$\eta_1 - \eta_5$	η_1	η_2	η_3	η_4	η_5
Reference economy	\$3,828	\$5,113	\$4,395	\$3,506	\$3,157	\$3,026
Dollar value when only one channel exists						
- Survival channel	86%	81%	83%	86%	91%	93%
- Labor market channel	18%	26%	21%	16%	11%	9%
- Medical expenses channel	2%	3%	3%	2%	2%	1%

Table 17: Decomposition of the willingness to pay for an increase in the probability of being healthy next period *when the statistical value of life is set at \$6 millions.* All dollar values when only one channel exists are the averages over the distribution of individuals aged 20-64 in the reference economy. We report the resulting values as a percentage of the dollar value when all channels operate (the first row).

	1^{st} Tercile	2^{nd} Tercile	3^{rd} Tercile
Reference economy	\$1,833	\$2,982	\$6,107
Dollar value when only one channel exists			
- Survival channel	61%	74%	91%
- Labor market channel	42%	26%	9%
- Medical expenses channel	5%	3%	1%

Table 18: Decomposition of the willingness to pay for an increase in the probability of being healthy next period by asset tercile *when the statistical value of life is set at \$6 millions.* The reported values are a percentage of the dollar value when all channels operate (the first row).

	Incl. survival channel		*Excl.* survival channel	
	β_{low}	β_{high}	β_{low}	β_{high}
All η_i	42.5%	12.8%	7.4%	0.2%
$\Rightarrow \{\eta_1, \eta_2\}$	47.5%	20.2%	9.5%	0.7%
$\Rightarrow \{\eta_3, \eta_4, \eta_5\}$	33.3%	9.9%	2.6%	0.0%

Table 19: Variation of lifetime utility due to health uncertainties *when the statistical value of life is set at \$6 millions.* The columns under "*Incl.* survival channel" are from the experiment where everyone is always healthy. The columns under "*Excl.* survival channel" are from the experiment where individuals are always healthy but they die at the same age as in the reference economy.