# Closing Down the Shop: Optimal Health and Wealth Dynamics near the End of Life*

Julien Hugonnier[1,4], Florian Pelgrin[2,5] and Pascal St-Amour[3,4,5]

[1]École Polytechnique Fédérale de Lausanne

[2]EDHEC Business School

[3]University of Lausanne, Faculty of Business and Economics (HEC)

[4]Swiss Finance Institute

[5]CIRANO

May 27, 2016

*Financial support from the Swiss Finance Institute is gratefully acknowledged.

**Abstract**

This paper examines health and wealth distributions under which agents approaching the end of life choose to *close down the shop*, i.e. a depletion of the health stock is optimally selected (and eventually accelerated), leading to states characterized by indifference between life, and death. We rely on a life cycle model (Hugonnier et al., 2013) with health spending, portfolio, insurance and consumption decisions for which we have closed-form optimal decisions in order to characterize optimal health and wealth dynamics. This model is estimated structurally using HRS data over a population of elders. Under economically plausible, and statistically verified conditions, we find that, unless sufficiently rich and healthy, agents will optimally select expected depletion of their health capital. Moreover, there exists a threshold wealth below which all agent deplete their health, regardless of how healthy they are. Finally, we identify a wealth and health locus below which agents accelerate their health depletion. Importantly, wealth is also expected to decline for all, such that *all* surviving agents eventually enter the closing down phase.

# 1 Introduction

## 1.1 Motivation and outline

Under most circumstances, a deterioration in health (e.g. because of illness) would be met by countervailing measures. Assuming sufficient financial resources, health expenses would be increased so as to either (i) restore the former level or, if not feasible, to (ii) slow down the deterioration until a stabilized lower level of health is attained. This paper looks at situations where neither option is optimally chosen. By focusing on the joint dynamics of health and wealth in the last years of life, we look at conditions under which health depletion is *optimally selected* to fall towards a region associated with very high mortality risk, and indifference between life and death. Put differently, agents are optimally choosing to let go, i.e. to *close down the shop*. The conditions under which they do are associated with lower wealth, such that richer individuals may delay the health depletion. However, we show that under reasonable assumptions, wealth depletion is *also* optimally selected, such that agents eventually enter the closing-down phase. We further identify threshold effects whereby the depletion of the health capital is initially slowed down, before being *accelerated*.

This paper has two main contributions: theoretical and empirical. First, we build upon a rich dynamic model that we developed in a companion paper (Hugonnier, Pelgrin and St-Amour, 2013) in order to identify optimal depletion, and acceleration. This life cycle model encompasses a health investment setup with endogenous exposure to death risk, and exogenous sickness shocks that further deplete the health capital. In addition to investing in their health, agents can buy actuarially fair insurance against health shocks, and save in risky, and risk-less assets. Agents also earn income, part of which is fixed (e.g. social security), and part which is health-dependent, reflecting their physical ability to work. Finally, preferences are characterized by subsistence consumption, as well as by generalized attitudes towards the various sources of risk (mortality, morbidity, and financial), and towards inter-temporal substitution. Importantly, they also guarantee strict *ex-ante* preference for life, so that agents have no predisposition towards premature death.

We rely on the closed-form solutions of that model for health expenses to characterize the optimal dynamics for health, and wealth capitals. Our main theoretical results (i) define the conditions under health and wealth depletion arises, and (ii) partition the health and wealth state space to identify whether or not agents are in these regions. First, the assumptions necessary for depletion are suitable for agents approaching the end of life. Indeed, they require that consumption (including subsistence) propensities as well as, sickness-adjusted depreciation of health capital are high, whereas the health-adjusted ability to generate income is low.

Second, under these assumptions, we identify a U-shaped locus in the health-wealth nexus such that all agents who are insufficiently rich/healthy optimally select expected depletion of their health stock. Consequently, there exists a threshold wealth level below which all agents expect a health depletion, regardless of the health status. Importantly, wealth depletion is also optimally selected, irrespective of the health and wealth levels. Combining these element entails that health is set on a downward spiral leading to drops in available resources, further cuts in health spending, and additional depletion of the health stock. We can also identify an accelerating locus below which health spending falls faster than health, such that agents initially slow down (yet do not reverse) the depletion, before choosing to accelerate the decline in health. Health thus eventually falls towards low levels that are associated with very high mortality risks, and indifference between life and death.

Our second contribution is empirical. Using HRS cross-sectional data, we rely on a trivariate econometric system composed of optimal health spending, risky asset holdings, and health-dependent income to structurally estimate the model over a population of relatively old agents. This exercise allows us to estimate the model's deep parameters, and evaluate the induced parameters that are used to partition the state space. The first set of results helps gauge and confirm the model's realism. The second allows us to evaluate and also confirm the economic relevance of the depletion zones.

In particular, we show that all the required conditions are met for existence of optimal closing-down strategies. Moreover, we show that the bulk of the population is located in the health depletion region, a small, yet significant portion of which are in the accelerating subset. We also substitute the estimated theoretical allocations in the laws of motion for health and wealth in order to simulate the life cycles in the last period of life. The results we get are consistent with expectations for end of life dynamics. Starting from a base age of 75, all agents optimally select to be within the health depletion region after 2 years, and increasingly in the accelerating phase afterwards. The endogenous exposure to death risk increases rapidly, leading to a life expectation that is consistent with the unconditional longevity of less than 80 years found in the data. Put differently, agents act in a manner that results in a terminal horizon of 5 years, and select an optimal depletion strategy that is consistent with this horizon.

These results raise two important normative issues. First, from a distributional point of view, we show that reducing the incidence of depletion zones can be achieved through an increase in base income, (e.g. through social security, or minimal revenue programs). This feature adds supplemental resonance to the usual finding of savings inadequacy for U.S. households.[1] Yet, to the extent that health and wealth depletion stems from optimizing behavior in a complete, and frictionless market setting, whether or not the state should intervene to prevent their occurrence is open to debate. Second, and related, an unresolved ethical question is whether or not medical treatment should be imposed to agents in the closing-down phase.[2] This paper argues that exploding end-of-life spending[3] may not reflect what agents actually want. Again from the perspective that the downward spiral in health is optimally selected, a more subtle approach to end-of-life care may be required.

The rest of this paper proceeds as follows. We summarize the theoretical model in Section 2. The depletion and accelerating regions are defined, and formally characterized in Section 3. The empirical evaluation is performed in Section 4, with main results outlined in Section 5. We close the discussion with concluding remarks in Section 6.

# 2 Theoretical framework

We rely upon the theoretical framework developed by Hugonnier et al. (2013) which we develop in order to analyze the existence of depletion regions of the state space. The

[1] See Hubbard et al. (1994, 1995); Skinner (2007) among others on insufficient financial, and pension savings.

[2] The legal right to refuse treatment is protected under both common law, and the American constitution (Legal Advisors Committee of Concerns for Dying, 1983), and recognized as such by the AMA (American Medical Association, 2016).

[3] For example, Marshall et al. (2010) provide evidence that end-of-life out-of-pocket health expenses increase sharply, and are strongly related to wealth level. Philipson et al. (2010) show that aggregate spending over the last year of life corresponds to a quarter of total health expenses in the US.

main features of this model are briefly reproduced below for completeness, with more detailed treatment to be found in the original paper.

## 2.1 Economic environment

The agent's health level $H_t$ follows a generalized stochastic version of the Grossman (1972) demand-for-health model:

$$\mathrm{d}H_t = \left( (I_t/H_t)^\alpha - \delta \right) H_t \mathrm{d}t - \phi H_t \mathrm{d}Q_{st}, \quad H_0 > 0, \tag{1}$$

where $I_t \geq 0$ is health spending, and $\mathrm{d}Q_{st}$ is a stochastic health shock. The positive restriction on investment is standard implies that the agent cannot short his health. The Cobb-Douglas parameter $\alpha \in (0, 1)$ captures diminishing returns to investing in one's health, and the continuous deterministic depreciation $\delta$ is augmented by a factor $\phi$ upon occurrence of a discrete sickness shock. The latter follows a Poisson process with constant intensity:[4]

$$\lambda_s(H_t) = \lambda_{s0}. \tag{2}$$

In a parallel vein, the age of death $T_m$ also follows a Poisson process, however with health-dependent endogenous death intensity:

$$\lambda_m(H_t) = \lim_{\tau \to 0} \frac{1}{\tau} P_t \left[ t < T_m \leq t + \tau \right] = \lambda_{m0} + \lambda_{m1} H_t^{-\xi_m}. \tag{3}$$

The component $\lambda_{m0}$ captures endowed exposure to death risk, whereas the second term $\lambda_{m1} H_t^{-\xi_m}$ determines endogenous exposure in that healthier agents can expect longer time horizon. The parameter $\xi_m$ controls diminishing returns to investing in one's health to prolong life, whereas $\lambda_{m1}$ controls the degree of endogeneity.

[4] Hugonnier et al. (2013) consider a more general endogenous sickness intensity function given by:

$$\lambda_s(H_{t-}) = \eta + \frac{\lambda_{s0} - \eta}{1 + \lambda_{s1} H_{t-}^{-\xi_s}},$$

and where we have set $H_{t-} = \lim_{\tau \uparrow t} H_\tau$ as the health level prior to occurrence of health shocks. For the current study, we restrict our analysis to the case of $\lambda_{s1} = 0$ corresponding to exogenous morbidity $\lambda_s(H_{t-}) = \lambda_s(H_t) = \lambda_{s0}$. Given that our main focus is centered upon elders, this restriction does not appear excessive in assuming that elders have limited leverage in adjusting their exposure to illness risk. Importantly, it will considerably facilitate the exposition.

Regarding the budget constraint, agents receive an income $Y_t$ at a rate that positively depends on their health:

$$Y(H_t) = y_0 + \beta H_t. \tag{4}$$

The base income $y_0$ captures health-independent elements such as Social Security revenue, whereas the health-dependent component $\beta H_t$ captures the deteriorated work ability for unhealthy agents. Furthermore, individuals can save and invest $\pi_t$ in a risky asset following a Brownian motion with market price of financial risk $\theta = \sigma_S^{-1}(\mu - r) \geq 0$, where $\mu$ is the drift, and $\sigma_S$ the diffusion of the risky asset, and $r$ is the risk-free asset rate. They can also purchase $X_{t-}$ units of an actuarially fair health insurance contracts paying one unit of the numeraire upon positive occurrence of the health shock.[5] The net return on insurance contracts $\mathrm{d}M_{st}$ is:

$$X_{t-}\mathrm{d}M_{st} = X_{t-}\mathrm{d}Q_{st} - X_{t-}\lambda_{s0}\mathrm{d}t. \tag{5}$$

Denoting $c_t$ the consumption, the budget constraint is thus:

$$\mathrm{d}W_t = (rW_{t-} + Y_t - c_t - I_t)\,\mathrm{d}t + \pi_t\sigma_S\,(\mathrm{d}Z_t + \theta\mathrm{d}t) + X_{t-}\mathrm{d}M_{st}. \tag{6}$$

To close the model, the agent's preferences are characterized by generalized recursive utility pioneered by Duffie and Epstein (1992), which are augmented by source-dependent risk aversion:

$$U_t = 1_{\{T_m > t\}} E_t \int_t^{T_m} \left( f(c_\tau, U_{\tau-}) - \frac{\gamma\sigma_\tau^2}{2U_\tau} - \sum_{k=m}^{s} F_k(U_{\tau-}, H_{\tau-}, \Delta_k U_\tau) \right) \mathrm{d}\tau, \tag{7}$$

where $U_t$ is the continuation utility. The risk aversion with respect to Brownian financial shocks is captured by $\gamma > 0$, whereas the Kreps-Porteus aggregator function is

$$f(c, v) = \frac{\rho v}{1 - 1/\varepsilon} \left( ((c-a)/v)^{1-\frac{1}{\varepsilon}} - 1 \right) \tag{8}$$

with elasticity of intertemporal substitution $\varepsilon > 0$, time preference rate $\rho > 0$ and subsistence consumption level $a \geq 0$. The aversion to Poisson sickness ($k = s$), and death

[5] We have set $X_{t-} = \lim_{\tau \uparrow t} X_\tau$ as the insurance purchased prior to occurrence of health shocks.

($k = m$) risks is captured by the functions $F_k$ which encode the utility costs associated with the discrete jumps in the continuation utility:

$$\Delta_k U_t = E_{t-}[U_t - U_{t-} | \mathrm{d}Q_{kt} \neq 0],$$

and where we have set:

$$F_k(v, h, \Delta) = v\lambda_k(h) \left[ \frac{\Delta}{v} + u(1; \gamma_k) - u\left(1 + \frac{\Delta}{v}; \gamma_k\right) \right],$$

with $u(x; \gamma_k)$ being a CRRA functional with curvature indices $0 \leq \gamma_m < 1$ for death risk aversion, and $\gamma_s \geq 0$ for sickness risk aversion. Given these elements, the agent's problem is to select optimal consumption, portfolio, insurance and investment so as to maximize utility (7):

$$V(W_t, H_t) = \sup_{(c,\pi,X,I)} U_t(c, I, H)$$

subject to the distributional assumptions, and laws of motion for health (1), and wealth (6).

## 2.2 Optimal dynamic policies

The presence of endogenous exposure to death risk implies that the previous model has no closed-form solution. However, Hugonnier et al. (2013) rely on a two-step analytical approximation. First, they rewrite the agent's incomplete market, and stochastic finite horizon problem as an iso-morphic one with complete markets, infinite horizon, and endogenous health-decreasing discounting. Second, they show that a closed-form solution exists in the restricted case of endogenous mortality corresponding to $\lambda_{m1} = 0$ (referred to as order-0 solution). They then perform an asymptotic expansion to calculate the first-order effect of endogenous mortality, and use this expansion to obtain approximate solutions (referred to as order-1 solution).

Adapting the results of Hugonnier et al. (2013) to our setting shows that the optimal investment in health can be written as:

$$I(W, H) = K_0 BH + \bar{K}_m \lambda_{m1} H^{-\xi_m} N_0(W, H) \tag{9}$$

where $K_0$ and $K_m = \bar{K}_m \lambda_{m1}$ are positive constants defined in equations (33), and (34) in Appendix A. The net total wealth is:

$$N_0(W, H) = W + BH + C \tag{10}$$

and comprises financial wealth $W$ plus the shadow value of the human capital $BH$, for which $B$ is the marginal (and average)-$Q$ of health, and $C = (y_0 - a)/r$ is the net present value (NPV) of base income $y_0$ net of subsistence consumption $a$. The first term in (9) is the order-0 investment that is proportional to the shadow value. The second term captures the additional demand for health that relates to its death risk hedging capacity; that demand is increasing in the endogenous component $\lambda_{m1} H^{-\xi_m}$ of the death intensity (3).

It can further be shown that optimal consumption, risky asset holdings, insurance and welfare are given by:

$$c(W, H) = a + AN_0(W, H) \tag{11}$$

$$\pi(W, H) = L_0 N_0(W, H) \tag{12}$$

$$X(H) = \phi BH \tag{13}$$

$$V(W, H) = \Theta N_0(W, H) \tag{14}$$

where the positive marginal propensity to consume $A > 0$, the portfolio share $L_0$ and the marginal value of net total wealth $\Theta > 0$ are also constant functions of the deep parameters that are defined in Appendix A. Observe that health investment (9) is the only optimal rule responding to endogenous death risk; all other variables in (11)–(14) have no first-order effect of $\lambda_{m1} H^{-\xi_m}$, and encompass at most only the exogenous component of mortality, $\lambda_{m0}$.

The optimal rules for investment, consumption, portfolio, and insurance are defined only over an *admissible* state space, i.e. the set of wealth and health levels such that net total wealth $N_0(W, H)$ is positive. Indeed, observe from optimal consumption (11), and welfare (14) that admissibility is required to guarantee that consumption $c_t$ is above subsistence $a$, and that the continuation utility of living $V_t$ is positive. Otherwise, negative total wealth entails negative continuation utility, and from preferences (7), a lower utility of living, than of dying ($\equiv 0$). More precisely, we can define:

**Definition 1 (admissible)** The admissible region $\mathcal{A}$ obtains when net total wealth is positive:

$$\begin{aligned}\mathcal{A} &= \{(W,H) : N_0(W,H) \geq 0\}, \\ &= \{(W,H) : W > x(H) = -C - BH\},\end{aligned}$$

with complement non-admissible set denoted $\mathcal{NA}$.

# 3 Health and wealth dynamics

The joint health and wealth system composed of the the laws of motion (1), and (6), and evaluated at the optimal rules (9), and (11)–(13) has complex nonlinear dynamics whose analysis is made even more challenging by Brownian financial, and Poisson health shocks. Indeed, the presence of the latter makes analytical solutions of the pair of stochastic differential equations $(\mathrm{d}H, \mathrm{d}W)$ intractable, and we will therefore restrict our analysis to expected local changes instead. Noting that the expected net return $\mathrm{d}M_{st}$ on actuarially fair insurance contracts (5) is zero reveals that the expected changes in health and wealth are:

$$E[\mathrm{d}H] = \left[I^h(W,H)^\alpha - \tilde{\delta}\right] H\mathrm{d}t, \tag{15}$$

$$E[\mathrm{d}W] = \left[rW + Y(H) - c(W,H) - I(W,H) + \pi(W,H)\sigma_S\theta\right]\mathrm{d}t, \tag{16}$$

where $I^h(W,H) = I(W,H)/H$ is the investment-to-health capital ratio, and $\tilde{\delta} = \delta + \phi\lambda_{s0}$ is the sickness-adjusted expected depreciation rate. Since our main focus concerns elders, the local expected changes (15), and (16) can be relied upon to define *depletion* regions of the admissible state space where health and wealth are (locally) expected to fall:

**Definition 2 (depletion)** Health, and wealth depletion regions $(\mathcal{D}_H, \mathcal{D}_W) \subseteq \mathcal{A}$ are characterized by optimal expected depletion of the health and wealth stocks:

$$\mathcal{D}_H = \{(W, H) \in \mathcal{A} : E[\mathrm{d}H] < 0\},$$
$$\mathcal{D}_W = \{(W, H) \in \mathcal{A} : E[\mathrm{d}W] < 0\}.$$

The following result relies on intuitive conditions to further characterize the depletion regions of the state space.

**Theorem 1 (depletion)** *Assume that the regularity, and transversality conditions* (30), (31), *and* (32) *hold. If, in addition:*

1. *The following conditions hold*

$$y_0 < a, \tag{17}$$
$$BK_0 < \tilde{\delta}^{1/\alpha}. \tag{18}$$

*Then the health depletion zone is given by:*

$$\mathcal{D}_H = \left\{(W, H) \in \mathcal{A} : W < y(H) = x(H) + DK^{1+\xi_m}\right\}, \tag{19}$$

*where,*

$$D = K_m^{-1}\left[\tilde{\delta}^{1/\alpha} - BK_0\right] > 0. \tag{20}$$

2. *If, in addition the following conditions hold:*

$$\beta < B(r + K_0), \tag{21}$$
$$\frac{\theta^2}{\gamma} + r < A, \tag{22}$$

*then the wealth depletion zone is given by:*

$$\mathcal{D}_W = \mathcal{A}. \tag{23}$$

The conditions (17), (18), (21), and (22) are economically plausible and relevant for end-of-life analysis. Condition (17), and (22) both refer to high consumption patterns, with (17) implying that base income $y_0$ in (4) is insufficient to cover subsistence consumption $a$ in (8), and condition (22) implying high marginal propensity to consume $A$ in (11). We can use the closed-form expression (37) for the latter to rewrite condition (22) as:

$$A - r - \frac{\theta^2}{\gamma} = \varepsilon(\rho - r) - (1-\varepsilon)\frac{\lambda_{m0}}{1-\gamma_m} - (1+\varepsilon)\frac{\theta^2}{2\gamma} > 0.$$

Since $\gamma_m \in (0,1)$, and assuming (as will be verified later) that the elasticity of intertemporal substitution $\varepsilon > 1$, the condition of a high marginal propensity to consume obtains when the agent is impatient, i.e. $\rho$ is high, and/or the unconditional risk of dying $\lambda_{m0}$ is high, and/or the aversion to death risk $\gamma_m$ is high. Condition (18) states that expected health depreciation $\tilde{\delta}$ is high, while condition (21) requires a low ability $\beta$ of healthier agents to generate labor revenues. Intuitively, the expression $(\tilde{\delta}^{1/\alpha} - BK_0)$ in (20) captures the order-0 expected depletion, i.e. in the absence of endogenous mortality. When the latter is reintroduced, optimal investment in (9) is larger, reflecting the additional demand for death risk hedging provided by health capital. If condition (18) is violated, then health grows in expectation absent mortality control value; positive growth is even larger when endogenous mortality is re-introduced and no relevant health depletion region exists in the admissible range.

The corresponding health and wealth dynamics in the admissible state space are plotted in Figure 1. First, the admissible region $\mathcal{A}$ is bounded below by the $W = x(H)$ locus in red, with complementary non-admissible area $\mathcal{NA}$ in shaded red region. The $W-$intercept is given by the NPV of base income deficit $-C$ which is positive under assumption (17). The $H-$intercept is given by $\bar{H}_1 = -C/B > 0$. Second, equation (19) in Theorem 1 states that the health depletion region $\mathcal{D}_H$ is bounded above by the green $W = y(H)$ locus, with health depletion region located below the locus in shaded green. Both $x(H), y(H)$ loci intersect at the same $-C$ intercept. A sufficiently high depreciation $\tilde{\delta}$ in (18) entails a positive constant $D > 0$ in (19). Consequently the $y(H)$ locus is U-shaped, and attains a unique minima at $\bar{H}_3$ given by:

$$\bar{H}_3 = \left(\frac{B}{D(1+\xi_m)}\right)^{\frac{1}{\xi_m}} > 0, \tag{24}$$

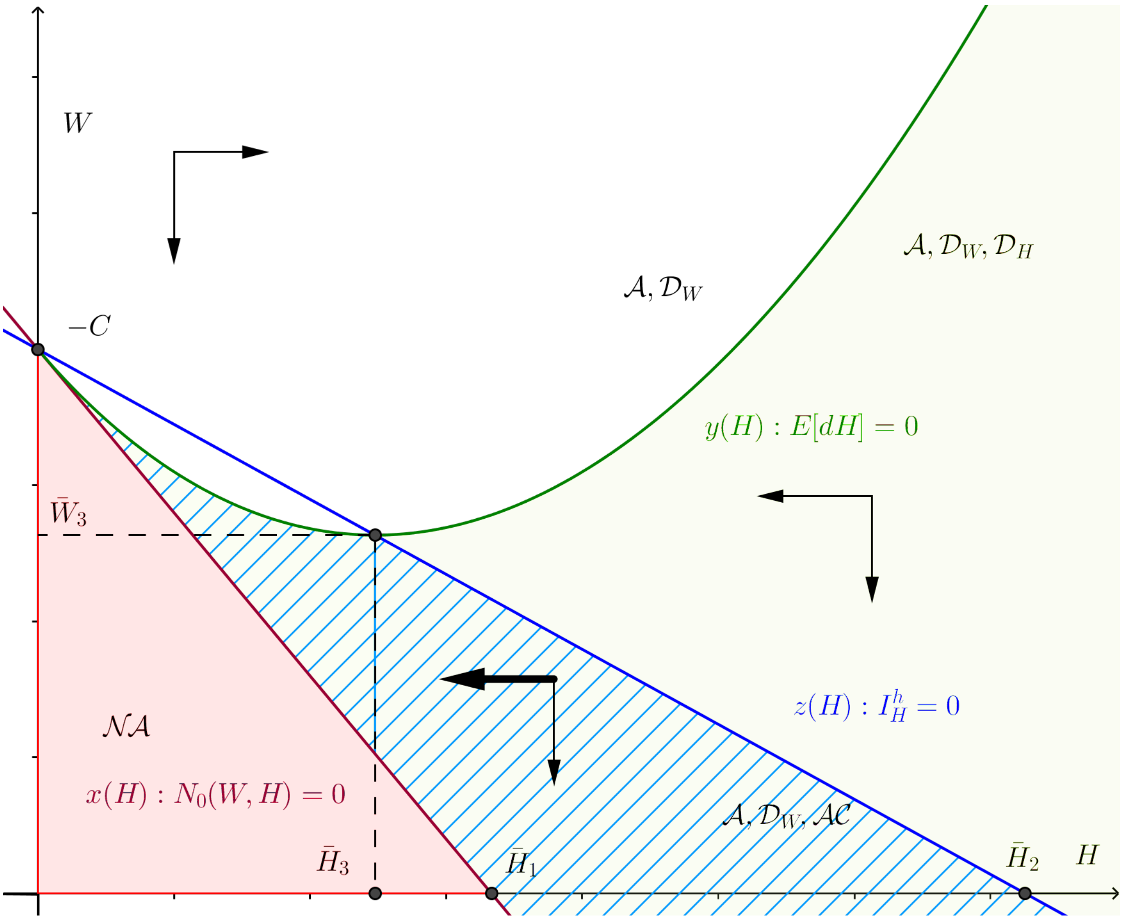


**Figure 1:** Health and wealth dynamics

*Notes:* The shaded area in red is the non-admissible set $\mathcal{NA}$ (Definition 1). The depletion area $\mathcal{D}$ (Theorem 1) is the shaded green area under the green curve. The accelerating region $\mathcal{AC}$ (Theorem 2) corresponds to the shaded green area hatched with blue lines.

with corresponding wealth level $\bar{W}_3 = y(\bar{H}_3)$. Finally, equation (23) in Theorem 1 states that the wealth depletion region $\mathcal{D}_W$ boils down to the entire admissible set $\mathcal{A}$. We will return to the interpretation of the third locus $W = z(H)$ in Theorem 2.

To see why the $W = y(H)$ locus is non-monotone, observe from (15) that expected change in health $E[\mathrm{d}H]$ increases in the investment-to-health ratio, where optimal investment (9) reveals that the latter is:

$$I^h(W, H) = BK_0 + K_m H^{-\xi_m - 1} N_0(W, H). \tag{25}$$

This ratio is monotone increasing in wealth, but not in health due to the opposing forces of total wealth, and mortality effects. First, an increase in $H$ raises net total wealth $N_0(W, H)$, and therefore raises $I^h$; constant (and zero) expected growth obtains by reducing $W$. Second, an increase in $H$ also reduces endogenous mortality risk $K_m H^{-\xi_m - 1} = \bar{K}_m \lambda_{m1} H^{-\xi_m - 1}$, and therefore also reduces $I^h$; constant zero growth obtains by increasing $W$. The analysis of the $W = y(H)$ locus in (38) thus reveals that the net total wealth effect is dominant at low health ($H < \bar{H}_3$), whereas the mortality risk effect dominates for healthier agents ($H > \bar{H}_3$).

The local expected dynamics are represented by the directional arrows in Figure 1. First, equation (19) implies that only agents who are sufficiently rich (i.e. $W > y(H)$) can expect a growth in health; all others are located in the $\mathcal{D}_H$ region, with health stock expected to fall. In particular, there exists a threshold wealth level $\bar{W}_3$ below which *all* agents, regardless of their health status, expect a health decline. Second, equation (23) stating the wealth depletion is the admissible set entails that *all* agents, regardless of their health or wealth levels expect wealth to fall. Taken together, these results suggest an optimal depletion of both human and financial capital with wealth eventually falling into the $\mathcal{D}_H$ region, and ensuing health depletion. From endogenous death intensity (3), falling health is invariably accompanied by an increase in mortality, and a decline towards the non-admissible locus $W = x(H)$ characterized by zero net total wealth, and indifference between life and death.

It is worth noting that the optimal risky asset holdings in (12) are positive, for positive net total wealth, and positive risk premia. Moreover, the investment in (9) is monotone increasing in wealth, such that a sufficiently long sequence of high positive returns on financial wealth could be sufficient to pull the agents away from the depletion region $\mathcal{D}_H$. Put differently, falling health, and higher mortality is locally expected, yet is not absolute for agents in the depletion region. We will return to this issue in the simulation exercise discussed below.

Interestingly, its is also possible to characterize differences in how fast the health capital is allowed to deplete. To do so, we can define an *acceleration* subset in the health depletion region whereby the investment-to-health ratio is an increasing function of health. Consequently, a depletion of the health capital leads to a decrease in $I^h$, and thus accelerating health depletion in (15). More precisely,

**Definition 3 (acceleration)** An accelerating zone $\mathcal{AC} \subset \mathcal{D}_H$ is a health depletion subset where the investment to health ratio $I^h(W,H)$ increases in health:

$$\mathcal{AC} = \left\{(W,H) \in \mathcal{D}_H : I^h_H(W,H) > 0\right\}$$

Relying on the optimal investment-to-health ratio (25) allows us to obtain the following result:

**Theorem 2 (acceleration)** *Assume that the conditions of Theorem 1 hold. Then the accelerating region is given by:*

$$\mathcal{AC} = \begin{cases} \mathcal{D}_H, & \textit{if, } H < \bar{H}_3 \\ \left\{(W,H) \in \mathcal{D}_H : W < z(H) = x(H) + \frac{BH}{1+\xi_m}\right\}, & \textit{otherwise} \end{cases} \tag{26}$$

The accelerating locus $W = z(H)$ is plotted as the blue line in Figure 1; the accelerating region is the dashed blue subset of $\mathcal{D}_H$. It is straightforward to show that this locus intersects the $x(H), y(H)$ loci at the same $-C$ intercept, that it intersects the $H-$axis at $\bar{H}_2 = \bar{H}_1(1+\xi_m)/\xi_m$, and finally and that it also intersects the health depletion locus $y(H)$ at its unique minimal value $\bar{H}_3$ in (24).

It follows from expected growth (15), and the characterization of $\mathcal{AC}$ in equation (26) that agents in the health depletion region $\mathcal{D}_H$ optimally slow down (but do not reverse) the depreciation of their health capital only if sufficiently rich and healthy ($W > z(H)$). Otherwise, for $(W,H) \in \mathcal{AC}$, the health depletion accelerates (illustrated by the thick directional vector) as falling health is accompanied by further cuts in the investment-to-health ratio. This dynamics suggests optimal *closing down the shop* behavior whereby falling health is initially optimally fought back, before eventually being accelerated.[6] Note finally that regardless of whether it is accelerating or not, the optimal descent of health and increased exposure to death risk for those agents in the health depletion region obtains even when life is strictly preferred. Indeed, as shown in Hugonnier et al. (2013), the non-separable preferences (7) ensure strictly positive continuation utility under life (versus zero under death), under admissible health and wealth statuses. The agents we are considering therefore have no predisposition in favor of premature death.

[6] See also Galama and Kapteyn (2011) for discussion of threshold effects in the demand for health in a different perspective.

Such a closing-down strategy of optimal wealth depletion, and eventual health depletion is arguably more appropriate for agents nearing death, than for younger ones. Indeed, a base income deficit relative to subsistence consumption (condition (17)), and a high marginal propensity to consume (condition (22)) are suitable for elders nearing end of life, with high maintenance expenses and no deliberate bequest motives.[7] Moreover, a high sickness-augmented depreciation rate for the health capital (condition (18)), and a low ability to generate labor revenues (condition (21)) both seem legitimate for old agents in the last period of life, yet less so for younger ones. The next section verifies empirically whether or not these conditions are valid.

# 4 Empirical evaluation

The structural econometric model that we rely upon to (i) estimate the deep parameters and (ii) evaluate the induced parameters ($B$, $C$, $D$, $\bar{H}_i$ for $i = 1, 2, 3$, and $\bar{W}_3$) that are relevant for the various regions of the state space is based on a subset of the optimal rules in Section 2.2.

## 4.1 Econometric model

The tri-variate nonlinear structural econometric model that we estimate over a cross-section of agents $j = 1, 2, \ldots, n$ is the optimal investment (9), and the risky asset holdings (12), to which we append the income equation (4) :

$$I_j = K_0 B H_j + K_m H_j^{-\xi_m} N_0(W_j, H_j) + u_{Ij}, \tag{27}$$

$$\pi_j = L_0 N_0(W_j, H_j) + u_{\pi j}, \tag{28}$$

$$Y_j = y_0 + \beta H_j + u_{Yj}, \tag{29}$$

where the $u_j$ are (potentially correlated) error terms. Data limitations discussed below explain why optimal consumption (11), and insurance (13) are omitted from the econometric model. The latter thus assumes that agents are heterogeneous only with respect

[7] The no-bequest assumption in preferences (7) is made necessary for tractability, and is justified by unclear empirical relevance. Hurd (2002) could find no evidence of a bequest motive in savings decisions, while Hurd (1987) finds no differences between the saving behavior of the elders with and without children. Both elements suggest accidental, rather than deliberate bequest decisions.

to their health, and wealth statuses; the deep parameters are considered to be the same across individuals. This assumption does not appear unreasonable to the extent that we are considering a relatively homogeneous subset of old individuals, thereby ruling out potent cohort effects. The joint estimation of (27), (28) and (29) is undertaken with respect to the deep parameters, under the theoretical restrictions governing $K_0, K_m, L_0$, as well as $B, C$, and also subject to the regularity conditions (30), (31), and (32).

The identification of the deep parameters is complicated by the significant non-linearities that are involved. Consequently, not all the parameters can be estimated, and a subset was therefore calibrated. Of those, certain parameters could be set at standard values from the literature. For others however, scant information was available, and we relied on thorough robustness analysis, especially with respect to $\gamma_m$, and $\phi$. These alternative estimates, which are available upon request, are reasonably robust.

The estimation approach is an iterative two-step procedure. In a first step, the convexity parameters $(\xi_m, \xi_s)$ are fixed and a maximum likelihood approach is conducted on the remaining structural parameters. In a second step, the structural parameters are fixed and the maximum likelihood function is maximized with respect to $\xi_m$ and $\xi_s$. The procedure is iterated until a fixed point is reached for both the structural parameters and the convexity parameters.

The likelihood function is written by assuming that there exist some cross-correlation between the three equations (investment, portfolio, and income). For the first two equations, the cross-correlation can be justified by the fact that we use an approximation of the exact solution (see Hugonnier et al., 2013, for details). Moreover, our benchmark case assumes that the three dependent variables are continuous. However, the risky holdings $\pi_j$ contain a significant share of zero observations. For that reason, we also experiment a mixture model specification in which the asset holdings variable is censored (Tobit) and the other two dependent variables (investment and income) are continuous, resulting in qualitatively similar results.[8]

[8]Note however that our structural model neither rules out zero holdings, nor does predict a Tobit-based specification for the portfolio equation.

## 4.2 Data

The data base used for estimation is the 2002 wave of the Health and Retirement Study (HRS, Rand data files). A main reason of using this HRS wave is that it is the last one with detailed information on total health spending; subsequent waves only report out-of-pocket expenses. Under OOP ceilings, total health expenses $I$ are not uniquely identified for insured agents, and we therefore resort to the 2002 HRS wave. Also, even though the HRS contains individuals aged 51 and over, we restrict our analysis to elders (i.e. agents aged 65 and more). In doing so, we avoid endogenizing the insurance choice $X_t$ in (5) which, under Medicare coverage, can be considered as exogenous. Unfortunately, this data set does not include a consumption variable, so that we omit equation (11) from the econometric model.

We construct financial wealth $W_j$ as the sum of safe assets (checking and saving accounts, money market funds, CD's, government savings bonds and T-bills), bonds (corporate, municipal and foreign bonds and bond funds), retirement accounts (IRAs and Keoghs), and risky assets (stock and equity mutual funds) $\pi_j$. Health status $H_j$ is evaluated using the self-reported general health status, where we express the polytomous self-reported health variable in real values with increments of 0.75 corresponding to: 0.5 (poor), 1.25 (fair), 2.00 (good), 2.75 (very good), and 3.50 (excellent).[9]

Health investments $I_j$ are obtained as the sum of medical expenditures (doctor visits, outpatient surgery, hospital and nursing home, home health care, prescription drugs and special facilities), and out-of-pocket medical expenses (uninsured cost over the two previous years). Finally, we resort to wage/salary income $Y_j$. The estimates presented below are obtained for a scaling of \$100'000 applied to all nominal variables $(I_j, W_j, \pi_j, Y_j)$.

Table 1 reports the median values for wealth, investment and risky asset holdings, for wealth quintiles, and self-reported health. Overall, these statistics confirm earlier findings. A first observation concerns the relative insensitivity of financial wealth to the health status.[10] Second, we find that health investment increases moderately in wealth,

[9]Self-reported health has been shown to be a valid predictor of the objective health status (Benítez-Silva and Ni, 2008; Crossley and Kennedy, 2002; Hurd and McGarry, 1995).

[10]See Hugonnier et al. (2013); Michaud and van Soest (2008); Meer et al. (2003); Adams et al. (2003) for additional evidence.

and falls sharply in health.[11] Conversely, risky holdings increase sharply in wealth, and are also higher for healthier agents.[12]

# 5 Results

Table 2 reports the calibrated, and estimated deep (panels a–d), the induced parameters that are relevant for the various subsets (panel e), as well as the hypothesis testing for the assumptions relevant to Theorems 1, and 2. The standard errors in parentheses indicate that all the estimates are significant at the 5% level.

## 5.1 Deep parameters

First, the law of motion parameters in panel a are indicative of significant diminishing returns to the health production function ($\alpha = 0.66$). Moreover, depreciation is important ($\delta = 8.8\%$), and sickness is rather consequential, with additional depreciation ($\phi = 1.1\%$) suffered upon realization of the health shock.

Second, in panel b the intensity parameters indicate a high, and significant incidence of health shocks ($1 - \exp(-\lambda_{s0}) = 49\%$). The death intensity (3) parameters are realistic, with an expected lifetime of 79.5 years for an individual with a an average (i.e. good) health.[13] Importantly, the null of exogenous exposure to death risk is rejected ($\lambda_{m1}, \xi_m \neq 0$), indicating that agent's health decisions are consequential for their expected life horizon. Taken together, these law of motion and risk exposure parameters compare well to estimates in Hugonnier et al. (2013), and are consistent with expectations regarding an elders' population.

[11]Similar findings with respect to wealth (e.g. Hugonnier et al., 2013; Meer et al., 2003; DiMatteo, 2003; Gilleskie and Mroz, 2004; Acemoglu et al., 2013) and health (e.g. Hugonnier et al., 2013; Smith, 1999; Gilleskie and Mroz, 2004; Yogo, 2009) have been discussed elsewhere.

[12]Similar positive effects of wealth on risky holdings have been identified in the literature (e.g. Hugonnier et al., 2013; Wachter and Yogo, 2010; Guiso et al., 1996; Carroll, 2002) whereas positive effects of health have also been highlighted (e.g. Hugonnier et al., 2013; Guiso et al., 1996; Rosen and Wu, 2004; Coile and Milligan, 2009; Berkowitz and Qiu, 2006; Goldman and Maestas, 2013; Fan and Zhao, 2009; Yogo, 2009).

[13]In particular, Hugonnier et al. (2013) show that an age-$t$ person's remaining life expectancy can be computed using:

$$\ell(W_t, H_t) = (1/\lambda_{m0})(1 - \lambda_{m1}\kappa_0 H_t^{-\xi_m}), \quad \text{where } \kappa_0 = [\lambda_{m0} - F(-\xi_m)]^{-1} > 0.$$

The average age in our HRS sample is 75.3 years, and the expected remaining life horizon is 4.2 years. The unconditional expected lifetime was 77.3 years in 2002, with 74.5 for males, and 79.9 for females (Arias, 2004).

Third, the income parameters of equation (4) are reported in panel c, and are both significant, and indicative of a positive health effects on income ($\beta \neq 0$). The base income $y_0$ is realistic, with a value of 820$ (corresponding to 1'082$ in 2016). The returns parameters $(\mu, r, \sigma_S)$ are calibrated at standard values. Fourth, the preference parameters in panel d suggest a realistic, and significant subsistence consumption $a$ of 2'000$ (2'640$ in 2016), which is larger than base income $y_0$. Our estimate of the intertemporal elasticity $\varepsilon$ is larger than one, as identified by others using micro data. For example, Gruber (2013) finds estimates centered around 2.0, relying on CEX data. In our case, the recourse to elders' data, and the assumption of no bequest function could explain a relatively strong consumption reaction to interest rates movements. Aversion to financial risk is realistic ($\gamma = 4.18$), whereas aversion to mortality risk is calibrated in the admissible range ($0 < \gamma_m < 1$), and close to the value set by Hugonnier et al. (2013) ($\gamma_m = 0.75$). The aversion to morbidity risk $\gamma_s$ is unidentifiable/irrelevant under the exogenous morbidity risk assumption. Finally, the subjective discount rate is set at usual values ($\rho = 2.5\%$). Overall, we conclude that the estimated structural parameters are economically realistic.

## 5.2 Induced parameters

Panel e of Table 2 reports the induced parameters that are relevant for the admissible, depletion and accelerating subsets; panel f shows that all the corresponding conditions in Theorem 1 are verified. These composite parameters therefore allow us to evaluate the position of the loci $x(H), y(H), z(H)$, and thus of the various subsets in Figure 2. The $H$ axis also records the positions associated with Poor ($H = 0.5$), and Fair ($H = 1.25$) reported health statuses, where the scaling is the one used in the estimation. The $W$ axis is reported in 100'000$, again using the same scaling as for the estimation. The health and wealth joint distribution for the HRS data is indicated by plotting the median for wealth associated with each quintiles, as blue points for $P_{20}, \ldots, P_{60}$, ($P_{80}, P_{100}$ omitted) for each health statuses.

First, in panel e, we identify a relatively large shadow value of health $B = 0.25$, indicating that a drop from one level of health to another is equivalent to a net total wealth change of $\Delta N_0 = B \Delta H = 0.75 * 0.25 * 10^5 = 18'750\$$ (24'750$ in 2016). Second, the large negative value for $C$ corresponds to a capitalised base income deficit of $0.2442 * 10^5 =$

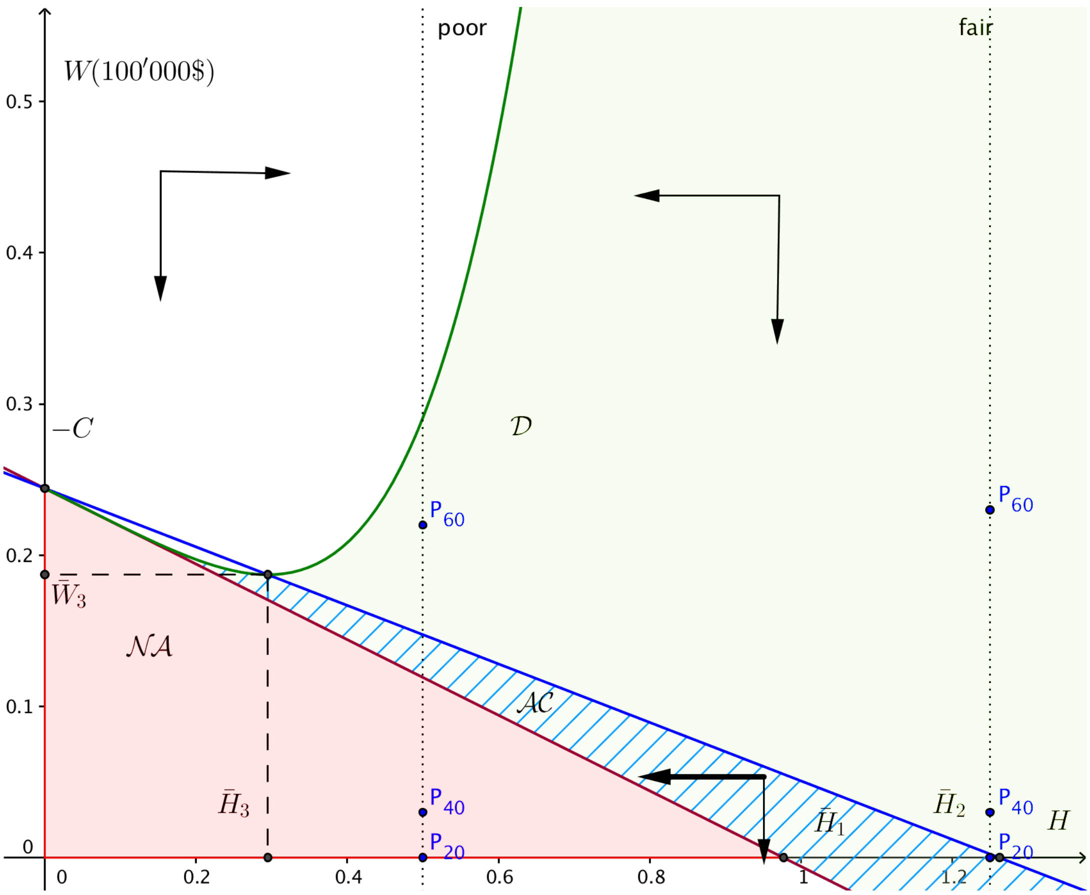


**Figure 2:** Estimated depletion, accelerating, and non-admissible regions

*Notes:* The shaded area in red is the non-admissible set $\mathcal{NA}$ (Definition 1). The depletion area $\mathcal{D}$ (Theorem 1) is the shaded green area under the green curve. The accelerating region $\mathcal{AC}$ (Theorem 2) corresponds to the shaded green area hatched with blue lines. Position of loci, and areas evaluated at estimated parameters in Table 2. Median levels for wealth quintiles $P_{20}, P_{40}, P_{60}, P_{80}$ ($P_{100}$ not reported) are taken from Table 1, and are reported as blue points for health levels poor, fair and good.

24'420\$ (32'234\$ in 2016), and confirms that condition (17) in Theorem 1 is verified. Third, the value for $D$ is significant which confirms the verification of condition (18). From the definition of $y(H)$ in (19), a large value of $D$ also entails a very steep health depletion locus. It follows that the bulk of the sample will be located in the depletion zone $\mathcal{D}_H$ below the green curve in Figure 2. Given that our sample is composed of elders, declining health is as expected. Fourth, our estimates are consistent with a narrow accelerating region $\mathcal{AC}$. Indeed, the values for $B, C, \xi_m$ are such that intercepts $\bar{H}_1, \bar{H}_2$ are relatively low (i.e. between Fair, and Poor self-reported health), and close to one another (less than one discrete increment of 0.75). Put differently, accelerating regions of the state space where agents are cutting down expenses in the face of falling health are relatively thin, yet significant, and associated with poor, unhealthy individuals.

Our results indicate that unless wealthy, and very unhealthy, the bulk of the population would be located in the depletion regions for health and wealth capitals. Besides being consistent with expectations regarding agents near the end of life, the results rationalize other aspects of behavior, such as unmet medical needs.[14] Moreover, our results are consistent with better longevity for the rich (Bosworth et al., 2016; Bosworth and Zhang, 2015). Indeed, agents who are not in in the health depletion region will select expected increase in the health capital as long as their wealth maintains them out of the $\mathcal{D}_H$.

## 5.3 Simulation analysis

The dynamic analysis presented thus far has focused upon local expected changes for health and wealth $E[\mathrm{d}H], E[\mathrm{d}W]$. At this stage there is no clear indication that such small anticipated depletions will translate into *bona fide* life cycle declining paths for health and wealth. To verify whether they do, we conduct a Monte-Carlo simulation exercise as follows.

1. We initialize the health and wealth distribution at base age $t = 75$ using a common uniform distribution for health, $H_0 \sim U[0.5, 3.5]$, and two distinct distributions for wealth:

    (a) Rich: $W_0 \sim U[2.5, 5.0]$;

    (b) Poor: $W_0 \sim U[0.1, 1.0]$;

    over a population of $n = 1'000$ individuals.

[14] Indeed, in 2010, among adults below 100% of the poverty level, 23.4% did not get or delayed medical care due to cost, 21.5% did not get prescription drugs due to cost, and 30.4% did not get dental care due to cost; these numbers fell to 6.8%, 3.9%, and 7.0% for richer households 400% above poverty (Tab. 79, National Center for Health Statistics, 2012, pp. 272–75). High prevalence of unmet medical needs due to financial reasons for uninsured Americans is also identified by Ayanian et al. (2000), especially in the case of unhealthy individuals. Park et al. (2016) find similar high incidence due to financial limitations in the case of Korean elders.

2. We simulate individual-specific Poisson health shocks $\mathrm{d}Q_s \sim P(\lambda_{s0})$, as well as a population-specific sequence of Brownian financial shocks $\mathrm{d}Z \sim N(0, \sigma_s^2)$ over a 10-year period $t = 75, \ldots, 85$.

3. At each time period $t = 75, \ldots, 85$, and using our estimated and calibrated parameters:

   (a) For each agent with health $H_t$, we generate the Poisson death shocks with endogenous intensities $\mathrm{d}Q_m \sim P[\lambda_m(H_t)]$, and keep only the surviving agents for the computation of the statistics.

   (b) We verify admissiblity, for each agent with health and wealth $(H_t, W_t)$ and keep only agents in the admissible region.

   (c) We use the optimal rules $I(W_t, H_t), c(W_t, H_t), \pi(W_t, H_t), X(H_{t-})$, as well as income function $Y(H_t)$, and the sickness and financial shocks $\mathrm{d}Q_{st}, \mathrm{d}Z_t$ in the stochastic laws of motion $\mathrm{d}H_t, \mathrm{d}W_t$.

   (d) We update the health and wealth variables using the Euler approximation:

$$
\begin{aligned}
H_{t+1} &= H_t + \mathrm{d}H_t(H_t, I_t, \mathrm{d}Q_{st}) \\
W_{t+1} &= W_t + \mathrm{d}W_t\left[W_t, c(W_t, H_t), I(W_t, H_t), \pi(W_t, H_t), X(W_t, H_t), \mathrm{d}Q_{s,t}, \mathrm{d}Z_t\right]
\end{aligned}
$$

4. We replicate the simulation 1–3 for $k = 1, \ldots, 1'000$ times.

Figure 3 plots the resulting mean values for the optimal life cycles for financial wealth $W_t$ (panel a), net total wealth $N_0(W_t, H_t)$ (panel b), health level $H_t$ (panel c), using only the alive, and admissible agents, as well as the share of the surviving, admissible population in the health depletion, and accelerating regions (panel d). The figures distinguish between the initial draw of rich (dashed line), and poor (dotted line) populations. Overall, these results provide strong evidence in favor of our previous findings, and confirm that agents optimally select a rapid depletion of both health and wealth capitals as they enter the end of life period. Indeed, recalling that expected longevity is 79.5 years, the optimal strategy is to bring down net total wealth to zero at terminal age (panel b), an objective obtained by running down wealth very rapidly, consistent with our finding that $\mathcal{D}_W = \mathcal{A}$, and a somewhat slower decline for health. The resulting subset shares in panel d confirm that virtually all the population is in the health depletion region at an early stage, and a sizable share enters the accelerating region halfway through. This accelerating share where agents cut down health expenses, as well as attrition effects whereby only the fittest agents survive explains the short-lived and moderate increase in $W_t$ after 80.

Contrasting rich versus poor cohorts reveals that, as expected, wealth, and health depletion is faster for poor agents, such that low-wealth individuals enter the depletion, and accelerating regions more rapidly. Put differently, our simulations indicate that agents entering the last period of life optimally select a short expected lifetime, and allocations that are consistent with optimal closing down, i.e. depletion of the health and wealth capitals during the end of life. High initial wealth thus has a moderating effect on the speed of the depletion, but not on its ultimate outcome.

# 6 Conclusion

This paper identifies conditions under which agents approaching the end of life optimally select to close down the shop, i.e. run down their health, and wealth capitals, bringing them to a state where they are indifferent between life and death. We relied on closed-form solutions to a life cycle model of optimal health spending and insurance, portfolio, and consumption to characterize the end of life dynamics for health, and wealth. Our findings can be summarized as follows. First, under certain plausible, end empirically verified conditions, agents optimally choose an expected depletion of their capital, unless they are sufficiently healthy and wealthy. We also identify a threshold wealth level below which health decline is independent on how healthy or not the agent is. Moreover, this depletion is accelerated below certain levels of health and wealth. Importantly, wealth is expected to fall regardless of the health status, such that all agents eventually close down the shop.

The previous analysis combined suggest a policy role in reducing the incidence of depletion regions of the state space. In particular, such a reduction is readily achieved

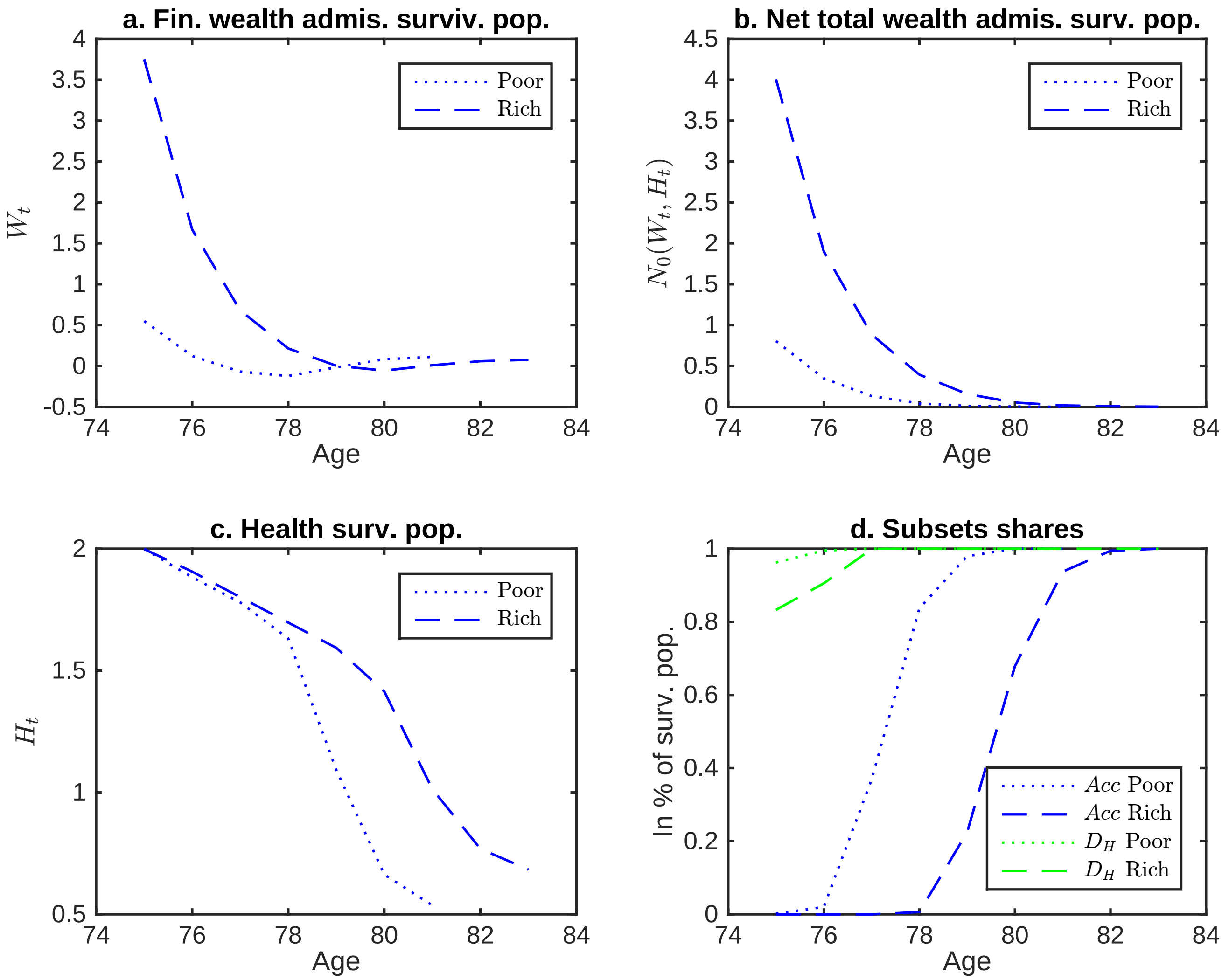


**Figure 3:** Simulated optimal paths

*Notes:* Mean values for simulated optimal life cycles taken over an initial population of 1'000 agents with 1'000 replications. Initial draw from rich (dashed lines), and poor (dotted lines) populations. Mean values in panels a–c are taken with respect to surviving, admissible agents. Subset shares taken as percentage of admissible surviving population in each subset. Death rate in panel e taken as percentage of initial population.

by increasing base income (e.g. through enhanced Social Security, Medicaid, or minimal revenue programs),[15] or via subsidized improvements in medical technology.[16]

[15]To see this, observe from equation (24) that the health threshold $\bar{H}_3$ is unaffected by the intercept $-C$, whereas the wealth threshold $\bar{W}_3 = z(\bar{H}_3)$ increases in the latter. Consequently, increasing base income $y_0$ directly lowers the wealth threshold through its effect on the base income deficit $-C$, and therefore the prevalence of health depletion. For example, sharp increases in the use of formal home care by elders is observed under more generous Social Security is identified by Tsai (2015).

[16]Hence, improvements that result in less sickness-adjusted depreciation $\tilde{\delta} = \delta + \lambda_{s0}\phi$ have a direct effect in lowering $D$, and therefore how steep the $y(H)$ locus is evaluated. Again the prevalence of $\mathcal{D}_H$ would be reduced.

However, whereas the positive arguments are readily obtained, the normative reasons for intervening are less clear. Indeed, continuous depletion of the health stock leading to very high death risks, and indifference to death is optimally selected, even in the case of agents with no predisposition for early death. Moreover, this downward spiral is obtained in a complete markets setting, such that no market failure argument for intervention can be invoked. Finally, assuming away policy changes in base income, state intervention on the public health domain in order to minimize unmet medical needs may also be questioned if, as the theory, and empirical evaluation suggest, failure to seek treatment is the result of an optimal dynamic decision by individuals.

# References

Acemoglu, Daron, Amy Finkelstein, and Matthew J. Notowidigdo (2013) 'Income and health spending: Evidence from oil price shocks.' *Review of Economics and Statistics* 95(4), 1079 – 1095

Adams, Peter, Michael D. Hurd, Daniel McFadden, Angela Merrill, and Tiago Ribeiro (2003) 'Healthy, wealthy, and wise? tests for direct causal paths between health and socioeconomic status.' *Journal of Econometrics* 112(1), 3–56

American Medical Association (2016) 'AMA policy on provision of life-sustaining medical treatment.' AMA website

Arias, Elizabeth (2004) 'United States life tables, 2004.' *National Vital Statistics Report* 53(6), 1–40

Ayanian, John Z., Joel S. Weissmann, Eric C. Schneider, Jack A. Ginsburg, and Alan M. Zaslasky (2000) 'Unmet health needs of uninsured adults in the United States.' *Journal of the American Medical Association* 284(16), 2061–2069

Benítez-Silva, Hugo, and Huan Ni (2008) 'Health status and health dynamics in an empirical model of expected longevity.' *Journal of Health Economics* 27(3), 564–584

Berkowitz, Michael K., and Jiaping Qiu (2006) 'A further look at household portfolio choice and health status.' *Journal of Banking and Finance* 30(4), 1201–1217

Bosworth, Barry P., and Kan Zhang (2015) 'Evidence of increasing differential mortality: A comparison of the HRS and SIPP.' Working papers, Center for Retirement Research at Boston College

Bosworth, Barry P., Gary Burtless, and Kan Zhang (2016) 'Later retirement, inequality in old age, and the growing gap in longevity between rich and poor.' Economic Studies at Brookings, Brookings Institute

Carroll, Christopher D. (2002) 'Portfolios of the rich.' In *Household Portfolios,* ed. Luigi Guiso, Michael Haliassos, and Tullio Jappelli (Cambridge MA: MIT Press) chapter 10, pp. 389–429

Coile, Courtney, and Kevin Milligan (2009) 'How household portfolios evolve after retirement: The effect of aging and health shocks.' *Review of Income and Wealth* 55(2), 226 – 248

Crossley, Thomas F., and Steven Kennedy (2002) 'The reliability of self-assessed health status.' *Journal of Health Economics* 21(3), 643–658

DiMatteo, Livio (2003) 'The income elasticity of health care spending: A comparison of parametric and nonparametric approaches.' *European Journal of Health Economics* 4(1), 20–29

Duffie, Darrell, and Larry G. Epstein (1992) 'Stochastic differential utility.' *Econometrica* 60(2), 353–394

Fan, Elliott, and Ruoyun Zhao (2009) 'Health status and portfolio choice: Causality or heterogeneity?.' *Journal of Banking and Finance* 33(6), 1079 – 1088

Galama, Titus, and Arie Kapteyn (2011) 'Grossmans missing health threshold.' *Journal of Health Economics* 30(5), 1044 – 1056

Gilleskie, Donna B., and Thomas A. Mroz (2004) 'A flexible approach for estimating the effects of covariates on health expenditures.' *Journal of Health Economics* 23(2), 391 – 418

Goldman, Dana, and Nicole Maestas (2013) 'Medical expenditure risk and household portfolio choice.' *Journal of Applied Econometrics* 28(4), 527 – 550

Grossman, Michael (1972) 'On the concept of health capital and the demand for health.' *Journal of Political Economy* 80(2), 223–255

Gruber, Jonathan (2013) 'A tax-based estimate of the elasticity of intertemporal substitution.' *Quarterly Journal of Finance* 3(1), 1 – 20

Guiso, Luigi, Tullio Jappelli, and Daniele Terlizzese (1996) 'Income risks, borrowing constraints, and portfolio choice.' *American Economic Review* 86(1), 158–172

Hubbard, R. Glenn, Jonathan Skinner, and Stephen P. Zeldes (1994) 'Expanding the life-cycle model: Precautionary saving and public policy.' *American Economic Review* 84(2), 174 – 179

— (1995) ‘Precautionary saving and social insurance.’ *Journal of Political Economy* 103(2), 360 – 399

Hugonnier, Julien, Florian Pelgrin, and Pascal St-Amour (2013) ‘Health and (other) asset holdings.’ *The Review of Economic Studies* 80(2), 663–710

Hurd, Michael D. (1987) ‘Savings of the elderly and desired bequests.’ *American Economic Review* 77(3), 298–312

— (2002) ‘Are bequests desired or accidental.’ Working Paper Series DRU-3010, RAND, January

Hurd, Michael D., and Kathleen McGarry (1995) ‘Evaluation of the subjective probabilities of survival in the Health and Retirement Study.’ *Journal of Human Resources* 30(0), s268–s292. Supplement

Legal Advisors Committee of Concerns for Dying (1983) ‘The right to refuse treatment: A model act.’ *American Journal of Public Health* 73(3), 918–921

Marshall, Samuel, Kathleen M. McGarry, and Jonathan S. Skinner (2010) ‘The risk of out-of-pocket health care expenditure at end of life.’ Working paper 16170, National Bureau of Economic Research

Meer, Jonathan, Douglas L. Miller, and Harvey S. Rosen (2003) ‘Exploring the health-wealth nexus.’ *Journal of Health Economics* 22(5), 713–730

Michaud, Pierre-Carl, and Arthur van Soest (2008) ‘Health and wealth of elderly couples: Causality tests using dynamic panel data models.’ *Journal of Health Economics* 27(5), 1312–1325

National Center for Health Statistics (2012) *Health, United States, 2011: With Special Feature on Socioeconomic Status and Health* (Hyattsville, MD: United States Department of Health and Human Services. Centers for Disease Controls and Prevention)

Park, Sojung, BoRin Kim, and Soojung Kim (2016) ‘Poverty and working status in changes of unmet health care need in old age.’ *Health Policy* pp. –. forthcoming

Philipson, Tomas J., Gary Becker, Dana Goldman, and Kevin M. Murphy (2010) 'Terminal care and the value of life near its end.' Working paper 15649, National Bureau of Economic Research

Rosen, Harvey S., and Stephen Wu (2004) 'Portfolio choice and health status.' *Journal of Financial Economics* 72(3), 457–484

Skinner, Jonathan (2007) 'Are you sure you're saving enough for retirement?' *Journal of Economic Perspectives* 21(3), 59 – 80

Smith, James P. (1999) 'Healthy bodies and thick wallets: The dual relation between health and economic status.' *Journal of Economic Perspective* 13(2), 145–166

Tsai, Yuping (2015) 'Social security income and the utilization of home care: Evidence from the social security notch.' *Journal of Health Economics* 43, 45 – 55

Wachter, Jessica A., and Motohiro Yogo (2010) 'Why do household portfolio shares rise in wealth?.' *Review of Financial Studies* 23(11), 3929 – 3965

Yogo, Motohiro (2009) 'Portfolio choice in retirement: Health risk and the demand for annuities, housing, and risky assets.' NBER Working Paper 15307, National Bureau Of Economic Research, September

# A Theoretical restrictions

Define the following functions:

$$\begin{aligned}
\chi(x) &= 1 - (1-\phi)^{-x}, \\
F(x) &= x(\alpha B)^{\frac{\alpha}{1-\alpha}} - x\delta - \lambda_{s0}\chi(-x), \\
L_m(H) &= \left((1-\gamma_m)(A - F(-\xi_m))\right)^{-1} H^{-\xi_m},
\end{aligned}$$

and assume that the following regularity and transversality conditions hold:

$$\beta < (r + \delta + \phi\lambda_{s0})^{\frac{1}{\alpha}}, \tag{30}$$

$$\max\left(0; r - \nu_{m0} + \theta^2/\gamma\right) < A, \tag{31}$$

$$0 < A - \max\left(0, r - \nu_{m0} + \theta^2/\gamma\right) - F(-\xi_m), \tag{32}$$

where the shadow price of health $B$, and the marginal propensity to consume $A$ are defined below. Then, the closed-form expression for the parameters in the optimal rules are given as follows. The positive parameters of the optimal investment in (9) are:

$$K_0 = \alpha^{\frac{1}{1-\alpha}} B^{\frac{\alpha}{1-\alpha}}, \tag{33}$$

$$K_m = \lambda_{m1} \frac{\xi_m K_0 L_m(1)}{1-\alpha} \tag{34}$$

where the shadow price $B$ of health solves:

$$g(B) = \beta - (r + \delta + \phi\lambda_{s0})B - (1 - 1/\alpha)\,(\alpha B)^{\frac{1}{1-\alpha}} = 0 \tag{35}$$

subject to $g'(B) < 0$, and the NPV of excess base income in (10) is:

$$C = \frac{y_0 - a}{r}. \tag{36}$$

The other parameters include the marginal propensity to consume in (11):

$$A = \varepsilon\rho + (1-\varepsilon)\left(r - \frac{\lambda_{m0}}{1-\gamma_m} + \frac{\theta^2}{2\gamma}\right), \tag{37}$$

the risky portfolio share in (12):

$$L_0 = \frac{\theta}{\gamma \sigma_S},$$

and the positive marginal value of total wealth in (14) :

$$\Theta = \rho(A/\rho)^{1/(1-\varepsilon)}.$$

# B Proof Theorem 1

## B.1 Health depletion $\mathcal{D}_H$

First, substituting the optimal investment (9) in the expected local change for health (15), and using the definition of net total wealth (10) shows that:

$$E[\mathrm{d}H] = 0 \iff W = x(H) + DH^{1+\xi_m} = y(H).$$

with expected depletion for $W < y(H)$. Second, observe that condition (17) implies that $-C > 0$ in (36), whereas condition (18) implies that $D > 0$ in (19). It follows directly that $y(H) \geq x(H), \forall H$, i.e. the locus $y(H)$ lies everywhere in the admissible zone, and is characterized by:

$$y_H(H) = -B + (1+\xi_m)DH^{\xi_m} \begin{cases} < 0, & \text{if } H < \bar{H}_3, \\ = 0, & \text{if } H = \bar{H}_3, \\ > 0, & \text{if } H > \bar{H}_3, \end{cases} \quad \text{and} \tag{38}$$

$$y_{HH}(H) = \xi_m(1+\xi_m)DH^{\xi_m - 1} > 0.$$

The locus $y(H)$ is therefore convex, and U-shaped and attains a unique minima at $\bar{H}_3$ in the $(H, W)$ space, where $\bar{H}_3$ is given in (24), with corresponding wealth level $\bar{W}_3 = y(\bar{H}_3)$.

## B.2 Wealth depletion $\mathcal{D}_W$

Substituting the optimal investment (9), consumption (11), risky portfolio (12), and insurance (13) in the expected local change for wealth (16), and using the definition of net total wealth (10) reveals that

$$E[\mathrm{d}W] = 0 \iff Wl(H) = x(h)[l(H) + r] + k(H),$$

where

$$\begin{aligned} l(H) &= \left[A + K_m H^{-\xi_m} - \sigma_S \theta L_0 - r\right], \\ k(H) &= (y_0 - a) + H(\beta - BK_0). \end{aligned}$$

Observe that since $K_m > 0$, condition (22) is sufficient to guarantee that $l(H) > 0, \forall H$. Consequently, the wealth depletion zone $\mathcal{D}_W$ is delimited by:

$$W > \frac{x(H)[l(H) + r]}{l(H)} + \frac{k(H)}{l(H)} = w(H).$$

We now have to show that this locus lies everywhere in the $\mathcal{NA}$ region:

$$w(H) < x(H) \iff x(H)r + k(H) < 0 \iff \beta < B(r + K_0)$$

as indicated by condition (21). Consequently, the wealth depletion $\mathcal{D}_W \subseteq \mathcal{A}$ coincides with the entire admissible set, i.e. $\mathcal{D}_W = \mathcal{A}$.

# C Proof Theorem 2

By a similar reasoning, we can observe from optimal investment (9) that the investment-to-health ratio is given by (25) . Taking the derivative with respect to $H$ and setting to zero shows that the accelerating region obtains as:

$$I_H^h(W_H) > 0 \iff W > -C - \frac{BH\xi_m}{1 + \xi_m} = z(H) = x(H) + \frac{BH}{1 + \xi_m}.$$

Since $B > 0$, this locus lies everywhere above the $\mathcal{A}$ locus, and is therefore admissible.

# D Tables

**Table 1:** HRS data statistics

| Variable | Wealth quintile 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| | a. Poor health ($H = 0.5$) | | | | |
| Financial wealth | 0.000 | 0.030 | 0.220 | 0.814 | 2.930 |
| Investment | 0.379 | 0.417 | 0.469 | 0.427 | 0.615 |
| Risky holdings | 0.005 | 0.079 | 0.216 | 0.485 | 0.800 |
| | b. Fair health ($H = 1.25$) | | | | |
| Financial wealth | 0.000 | 0.030 | 0.230 | 0.760 | 3.400 |
| Investment | 0.255 | 0.254 | 0.233 | 0.252 | 0.266 |
| Risky holdings | 0.000 | 0.046 | 0.253 | 0.514 | 0.782 |
| | c. Good health ($H = 2.0$) | | | | |
| Financial wealth | 0.000 | 0.040 | 0.220 | 0.770 | 3.300 |
| Investment | 0.157 | 0.149 | 0.156 | 0.129 | 0.168 |
| Risky holdings | 0.002 | 0.082 | 0.299 | 0.510 | 0.824 |
| | d. Very good health ($H = 2.75$) | | | | |
| Financial wealth | 0.000 | 0.040 | 0.230 | 0.840 | 3.500 |
| Investment | 0.100 | 0.112 | 0.106 | 0.105 | 0.107 |
| Risky holdings | 0.011 | 0.107 | 0.368 | 0.604 | 0.854 |
| | e. Excellent health ($H = 3.5$) | | | | |
| Financial wealth | 0.000 | 0.050 | 0.210 | 0.800 | 3.820 |
| Investment | 0.137 | 0.065 | 0.063 | 0.105 | 0.091 |
| Risky holdings | 0.010 | 0.131 | 0.350 | 0.520 | 0.861 |

*Notes:* Median (wealth), and mean values (investment, risky holdings), measured in 100'000$ (year 2002) per health status, and wealth quintiles for HRS data used in estimation.

**Table 2:** Estimated and calibrated parameter values

| Parameter | Value | Parameter | Value | Parameter | Value |
|---|---|---|---|---|---|
| | | a. Law of motion health (1) | | | |
| $\alpha$ | $0.6591^{*}$ | $\delta$ | $0.0882^{*}$ | $\phi$ | $0.011^{c}$ |
| | (0.1699) | | (0.0384) | | |
| | | b. Sickness and death intensities (2), (3) | | | |
| $\lambda_{s0}$ | $0.6746^{*}$ | $\lambda_{m0}$ | $0.2306^{*}$ | | |
| | (0.2105) | | (0.0911) | | |
| $\lambda_{m1}$ | $0.0025^{*}$ | $\xi_m$ | $3.4219^{*}$ | | |
| | (0.0012) | | (0.9033) | | |
| | | c. Income and wealth (4), (6) | | | |
| $y_0$ | $0.0082^{*}$ | $\beta$ | $0.0138^{*}$ | | |
| | (0.0018) | | (0.0030) | | |
| $\mu$ | $0.108^{c}$ | $r$ | $0.048^{c}$ | $\sigma_S$ | $0.20^{c}$ |
| | | d. Preferences (7), (8) | | | |
| $a$ | $0.0200^{*\$}$ | $\varepsilon$ | $1.6302^{*}$ | $\gamma$ | $4.1815^{*}$ |
| | (0.0052) | | (0.7822) | | (1.5811) |
| $\rho$ | $0.025^{c}$ | $\gamma_m$ | $0.75^{c}$ | $\gamma_s$ | N.I. |
| | | e. State space subsets (35), (36), (20), (24), (9) | | | |
| $B$ | $0.2500^{*}$ | $C$ | $-0.2442^{*\$}$ | $D$ | $3.6825^{*}$ |
| $\bar{H}_1$ | $0.9769^{*}$ | $\bar{H}_2$ | $1.2624^{*}$ | $\bar{H}_3$ | $0.2951^{*}$ |
| $K_0$ | $0.0202^{*}$ | $K_m$ | $0.0055^{*}$ | $\bar{W}_3$ | $0.1871^{*\$}$ |
| | | f. Conditions (17), (18), (21), (22) in Theorem 1 | | | |
| $y_0 - a$ | $-0.0118^{*}$ | $BK_0 - \tilde{\delta}^{1/\alpha}$ | $-0.0233^{*}$ | | |
| $\beta - B(r + K_0)$ | $-0.0032^{*}$ | $\theta^2/\gamma + r - A$ | $-0.5155^{*}$ | | |

*Notes:* *: Estimated structural and induced parameters (standard errors), significant at 5% level; $c$: calibrated parameters; \$: In 100'000\$; N.I.: non-identifiable/irrelevant under the exogenous morbidity restriction.