Københavns Universitet

Optimal Aging and Death

Dalgaard, Carl-Johan; Strulik, Holger

Discussion Papers
Department of Economics
University of Copenhagen

Optimal Aging and Death: Understanding the Preston Curve

Carl-Johan Dalgaard, Holger Strulik

Optimal Aging and Death:
Understanding the Preston Curve*

Carl-Johan Dalgaard[†] and Holger Strulik[**]

First Version: July 2010. This Version: March 2011.

Abstract. The present study examines whether the Preston curve reflects a causal impact of income on longevity or, for example, factors correlated with both income and life expectancy. In order to understand the Preston curve better, we develop a model of optimal intertemporal consumption in which the representative consumer is subject to physiological aging. In modeling aging we draw on recent research in the fields of biology and medicine. The speed of the aging process, and thus the time of death, are endogenously determined by optimal health investments. We calibrate the model to US data and proceed to show that the model accounts for nearly 80% of the cross-country differences in life expectancy that the Preston curve captures.

Keywords: Aging, Longevity, Health investments, Savings, Preston curve.
JEL: D91, J17, J26, I12.

1. Introduction

In a seminal contribution Samuel Preston (1975) documented a striking empirical fact: In a cross-section of countries higher levels of income per capita are associated with greater life expectancy; the curve that fits the data best – dubbed "the Preston curve" by later research – is concave. This discovery served to promote the idea that prosperity leads to greater longevity, and that income inequality works to lower average longevity; the latter being a consequence of the observed concavity of the mapping between income and life expectancy.[1]

These are powerful ideas, which continue to be influential. As Bloom and Canning (2007, p. 498) observe: "Samuel H. Prestons classic paper, 'The Changing Relation between Mortality and Level of Economic Development', published in 1975, remains a cornerstone of both global public health policy and academic discussion of public health." Despite its prominence, however, the interpretation of the Preston curve continues to be shrouded in mystery. Does the slope of the curve reflect a causal impact of income *per se*? Or is income perhaps a stand-in for other underlying correlates with longevity and income, such as efficiency differences in health care? Deaton (2003, p. 152) formulates the key challenge for research in this area succinctly: "If income is indeed directly protective, we need to know whether the effect is really nonlinear [...] because it is this [...] that determines whether and by how much income redistribution can improve population health". The objective of the present research is to offer some progress in this regard.

We employ a theory driven approach in order to elucidate the mechanisms underlying the shape and position of the Preston curve. In order to understand human longevity we need to understand the determinants of the process which *culminates* in death: the aging process (Arking, 2006). We therefore develop a life-cycle model for a representative consumer who is subject to physiological aging. In modeling the aging process we draw on recent research in the fields of biology and medicine. This approach has the advantage that estimates for key physiological parameters influencing the aging process are available to us; this is what facilitates a credible calibration of life expectancy. In our model the individual consumes, saves, and makes deliberate investments in slowing down the aging process, thereby postponing his or her "date

---

[1]A second observation made in the paper was that the curve shifts upwards over time, implying greater longevity at all levels of income per capita. Preston hypothesized that these shifts represented improvements in health technology (broadly defined), and noted that the shifts accounted for the lions share of global improvements in longevity over time.

of expiry". This model allows us to study the impact from e.g. changes in income and health care efficiency on changes in longevity.

With the model in place we subsequently calibrate it to US data, and proceed to examine the calibrated model's ability to account for the *cross-country* income gradient in the year 2000, as reflected in the Preston curve. Accordingly, using income data for a cross section of countries as "input" we ask the model to predict the associated levels of life expectancy for those countries, which we then compare to the estimated income gradient: the Preston curve. The model does a good job at mimicking the Preston curve; about 80% of the variation nested in the Preston curve can be accounted for by the model. That is, the pure income effect on longevity, mediated by health investments, goes a long way towards explaining the Preston curve.

What should one make of the left-over residual (i.e., the remaining 20% of the variation captured by the Preston curve)? When we predict longevity using our model we are keeping the relative price of health goods (to non-health goods) in efficiency units fixed at the US level. Empirically, the (efficiency adjusted) price of health undoubtedly differs across countries. Such variation can, according to our model, account for the residual. While the model does not allow us to decompose this left-over residual, it undoubtedly captures a range of factors such as cross-country variation in relative prices of health goods; climate (influences disease prevalence, and thus attainable longevity at a given heath investment rate); health care institutions (influences how many years a given dollar of health investment can buy) and more.[2] The main message from the analysis, however, is that the Preston curve indeed seems to capture a strong - nonlinear - causal effect of income on longevity. To put it in popular terms: "The Preston curve is 80% about the impact of income on longevity".

The proposed theory is centered around a novel approach to the modeling of aging. Based on research on aging from the natural sciences, we conceptualize the process of aging as one whereby the human organism gradually looses redundancy and thus becomes more *fragile*. This leads us to a law of motion for human frailty, which depends on physiological parameters and health investments. The process of increasing frailty is relentless and accelerating over time, but it may be slowed by health investments. The incentive to slow down aging is a longer life, which facilitates greater consumption and thus utility. The costs of doing so is foregone utility

---

[2]The Preston curve may also be capturing - in part at least - an impact of longevity on income. Since our model deliberately abstracts from this causal mechanism, instead distilling the power of the income-to-longevity mechanism, this sort of influence would also go into the left-over residual. We return to this issue below.

from consumption of goods that do not serve to prolong life. In a setting where individuals are maximizing lifetime utility from consumption we examine the optimal intertemporal choice with respect to savings and health investments. Subject to the physiological constraints faced by humans, and the standard budget constraint, this allows us to characterize the optimal speed of aging and thereby optimal longevity. The aging process and time of death is thus endogenously determined by optimal health investments. The physiological foundations of the aging process, as implemented in the analysis, allow us to exploit that key physiological parameters of relevance to the process all have been estimated with great precision in the medical science literature. This is largely what enables us to calibrate the model in a meaningful way, and quantitatively address the interpretation of the Preston curve.[3]

Our analysis is primarily related to the seminal work of Grossman (1972) on the demand for health and to Ehrlich and Chuma's (1990) important work on optimal longevity.[4] The key theoretical difference between our work and previous contributions lies in the law of motion which governs the aging process; the counterpart to the law of motion for health capital in the existing literature. We elaborate on the differences between the standard approach to health capital accumulation and our proposed approach below. Another difference is that we employ our model to examine the underlying forces that shape the Preston curve from a quantitative angle; this has not been attempted in the existing literature.

The paper proceeds as follows. In the next section we derive the key law of motion that characterizes the aging process, and we discuss how this law of motion differs from the standard health accumulation equation. Section 3 contains the main analysis of optimal aging and death. In Section 4 we calibrate the model to US data and examine its predictions as well as its dynamical properties in the presence of shocks to income, technology and more. Then, in

---

[3]A few remarks on the set-up are warranted. First, one could imagine that health investments also stimulate productivity, which inevitably influences the income/longevity correlation in the cross-country data. In the present context, however, we are interested in understanding the potential strength of the causal mechanism running from income to longevity; this objective is compromised if we admit the reverse chain of causality to be present in the model, for which reason it is suppressed. Second, in our model health matters for the time horizon, but does not enter the per period utility function. Abandoning this assumption would certainly add flexibility to the model, making it far easier to map it into US data on longevity etc. At the same time it would be in conflict with the "Stigler-Becker dictum" of not appealing to preferences when explaining phenomena of interest (Stigler and Becker, 1977, p. 76.). Our approach to the modeling of the direct utility value of rising longevity therefore follows Becker (2007).

[4]Other notable contributions in this tradition are Reid (1998), Eisering (1999) and Foster (2001). Also related are contributions investigating the evolution of life expectancy at the aggregate level; see Cervellati and Sunde (2005) and Galor and Moav (2007).

Section 5, we apply the model to the study of the Preston curve. Finally, Section 6 is reserved for concluding remarks.

2. Introducing Biological Aging

2.1. **Modeling Human Aging as a Process of Deficit Accumulation.** Aging is defined as the intrinsic, cumulative, progressive, and deleterious loss of function that eventually culminates in death (Arking, 2006). At the individual level the aging process exhibits great heterogeneity, and is only imperfectly captured by chronological age; some 60 year-olds are as fit as some 40 year-olds and vice versa. Indeed, biologists and gerontologists stress that individual aging should be viewed as an event-dependent process, rather than as a time-dependent process. There is no such thing as a "biological clock", which determines the speed of individual aging and timing of death.

At the population level, however, age is a better predictor of aging and death. The so-called Gompertz-Makeham law of mortality implies that the death rate rises exponentially with age.[5] The fact that the age-specific mortality rate increases is obviously a manifestation of the aging process. But then why is the death rate (for the average individual) increasing with age?

A reductionist approach might suggest that the organism ages because of aging organs, which in turn is caused by aging tissue, brought on by the aging of cells and so on. Such an approach, however, is a dead end. The reason is that, eventually, a level of "disaggregation" is reached, which consists of non-aging entities: atoms, for example. The aging process can not be understood in this manner. But how can a system constructed from non-aging components age in the manner suggested by the Gompertz-Makeham law of mortality?

A promising strand of literature in biology has sought an answer by drawing on reliability theory from engineering.[6] To understand the basic idea, consider the following model (Gavrilov and Gavrilova, 1991). Suppose we view the organism as a whole as consisting of a fixed number of individual parts, which we will refer to as "blocks". Each block does not age. That is, the failure rate of a block is constant. This assumption captures that the human organism is ultimately constructed from non-aging components, as noted above. Next, assume that the

---

[5]More precisely, the Gompertz-Makeham (GM) law states that the age specific mortality rate, $\pi(t)$, evolves with age, t, in accordance with the formula $\pi(t) = a + b\ e^{\alpha t}$ where a, b and α are parameters. The fit of the GM law, for the population aged 20 and above (children and teen-ages are "special") is extremely good, always featuring an R^2 in excess of 0.95. See e.g., Arking (2006).

[6]Reliability theory is used in engineering to understand the failure rate of mechanical devices. Gavrilov and Gavrilova (1991) were the first to introduce reliability theory into biology.

4

blocks are connected in parallel, and that the system as a whole is assumed to survive as long as there is one functioning block remaining. This assumption is motivated by the physiological fact that the human organism is characterized by a great deal of redundancy; as young adults the functional capacity of our organs is estimated to be tenfold higher than needed for mere survival (Fries, 1980). Though each block does not age the passing of time will reduce redundancy in the system as a whole, which leads to an increasing failure rate of the system. Hence, the simple model successfully reproduces an exponentially rising death rate with age. Many extensions of this basic model have been developed, which have lead to new insights into the aging process.[7] But for present purposes the key point of reliability theory is conveyed by the simple model: senescence can be conceptualized as the gradual loss of redundancy, ultimately leading to organism collapse.

Following the underlying reasoning of reliability theory one may therefore think of aging as being characterized by increasing *frailty*. That is, as the redundancy of the human organism shrinks we become more fragile. An empirical measure of human frailty has been developed by Mitnitski and Rockwood and various coauthors in a series of articles (e.g., Mitnitski et al, 2002a,b; 2005; Rockwood and Mitnitski, 2006).

As humans age they develop an increasing number of disorders, which Mitnitski et al. (2002a) refer to as "deficits". Some of these deficits may be viewed as rather mild nuisances (e.g., reduced vision) while others are more serious in nature (e.g., strokes). Nevertheless, the notion is that when the number of deficits rises the body becomes more frail. A frailty index can then be estimated as the proportion of the total potential deficits that an individual has, at a given age.[8]

Mitnitski et al. (2002a) show that the following equation fits data on the proportion of deficits, $D(t)$, of the representative individual at age t very well:

$$D(t) = E + Be^{\mu t}.$$

---

[7]For example, if one extends the framework we have just sketched by assuming that the body consists of a number of *essential* blocks (so that if one fails the organism fails), each of which consists of non-aging elements of which some initially are defect, the model reproduces the Gompertz-Makeham law; see e.g., Gavrilov and Gavrilova (1991).

[8]Some methodological notes. To be in the index the deficit must have been demonstrated to be a group and individual indicator of health, and an important correlate of survival (Mitnitski et al., 2002a). " The total number of potential deficits" are in practise determined by the survey at hand. This may seem arbitrary. But according to Rockwood and Mitnitski (2007) the exact choice of deficits is usually not crucial. Provided sufficiently many indicators (40 or more) are present in the index, results tend to be relatively unaffected.

This "law of increasing frailty" explains around 95% of the variation in the data, and its parameters are estimated with great precision. The parameter E turns out to be common for men and women; using a data set encompassing 66,589 Canadians, aged 15 to 79, Mitnitski et al. (2002a) estimate E to 0.02, with a standard error of 0.001. The parameters B and μ are, however, gender specific. For Canadian men (women) $\log(B)$ is -5.77 ± 0.06 (-4.63 ± 0.06), while μ is 0.043 ± 0.001 (0.031 ± 0.001). Interestingly, very similar estimates for B and μ are obtained on data for Australia, USA and Sweden (Rockwood and Mitnitski, 2007). Hence, in these four developed countries (in spite of differences in samples, the precise contents of the frailty index etc.) the average individual accumulates 3-4% more deficits from one birthday to the next.

The strong ability of the deficit model to fit the data is encouraging. But a more intriguing finding is that there appears to be a very strong link between the path of frailty and the path of mortality. If average deficits of individual age groups are plotted in a $(t, \log[D(t)])$ diagram, the resulting lines for men and women intersect. While men start out with lower deficits ($D(0) = 0.023$ compared to 0.029 for women) they tend to accumulate deficits at a higher rate. Eventually, therefore, the proportion of deficits converge. The age at which deficits coincide - based on the estimates reported above - is 94. Now, when the age-specific mortality rate (the Gompertz-Makeham equation; see footnote 5) is depicted, a similar intersection point is found. Hence, in terms of mortality there is also a trade-off between the initial death rate and the speed at which the age specific mortality rate goes up. What is remarkable is that the intersection for mortality occurs at roughly the same age as that of frailty: at age 95. In fact, even when one compares mortality patterns across different countries the same "age of convergence" is found. In biology this fact is referred to as "the law of compensating mortality"; initial high mortality appears to be "compensated" by a more slowly rising mortality rate during life. Similar intersection points are also found in other species, albeit the age of intersection naturally is different from species to species; biologists refer to the age at which an intersection is found as the "species-specific lifespan" (e.g., Gavrilov and Gavrilova, 1991, p. 148-56).

Hence, data on frailty and mortality independently predict gender convergence of aging, at roughly the same chronological age (for the average individual). This result suggests that the process of deficit accumulation indeed is capturing important aspects of the aging process. Indirectly, the results support the theoretical proposition of reliability theory. Namely, that aging involves increasing frailty, which eventually leads to organism collapse and death.

We can restate the law of increasing frailty in flow form by differentiating with respect to age:

$$\dot{D}(t) = \mu\left(D(t) - E\right), \tag{1}$$

where E works to slow down the speed of deficit accumulation. In order to see that the influence of E in (1) is consistent with Mitinitski and Rockwood's equation for the *level* of deficits, integrate (1) and insert the initial condition $D(0) = D_0$ to get the solution $D(t) = (D_0 - E)\,\mathrm{e}^{\mu t} + E = D_0\mathrm{e}^{\mu t} - E(\mathrm{e}^{\mu t} - 1)$. Since $\mathrm{e}^{\mu t} > 1$ for all $t > 0$, a larger autonomous component E implies less deficits at any given age t. Note also that the compound parameter $(D_0 - E)$ corresponds to Mitnitski et al.'s estimate of B. In the natural science literature the parameter E is interpreted as capturing the impact of non-biological factors on deficit accumulation (Mitnitski et al., 2002a). Accordingly, we will assume that E is amendable to change by way of deliberate investment. This is the way in which the individual may attempt to slow down aging in the model below.

Specifically, we propose the following parsimonious refinement of the process of deficit accumulation:

$$\dot{D}(t) = \mu\left(D(t) - a - Ah(t)^{\gamma}\right), \tag{2}$$

where $D(0)$ is given. The parameter a captures environmental influence on aging beyond the control of the individual (less pollution, say, implying a higher value for a), the parameters $A > 0$ and $0 < \gamma < 1$ reflect the state of the health technology, and h is health investment. While A refers to the general power of health expenditure in maintenance and repair of the human body, the parameter γ specifies the degree of decreasing returns of health expenditure. The larger γ the larger the relative productivity of cost-intensive high-technology medicine in maintaining and repairing highly deteriorated human bodies.

By way of contrast to E, the parameter μ – impressed by its empirical constancy across developed countries – is considered to be a physiological parameter. In the remaining we will refer to this physiological parameter as *the force of aging*, as it drives the inherent and inevitable process of human aging. While the force of aging is exogenous, it is unlikely to be universally constant. In fact, it varies across countries in different parts of the world, much like the force of mortality, according to the Gompertz-Makeham law, varies across countries (see Strehler and Mildvan, 1960).

In order to capture death, we need to invoke an upper boundary to deficit accumulation, $\bar{D}$. In the analysis below the representative individual remains alive as long as $D(t) < \bar{D}$. Direct

evidence on the existence of an upper boundary for D is found in Rockwood and Mitnitski (2006). Observe that equation (2), along with the restriction that $D(t) < \bar{D}$, provides a complete description of aging until death. In this process, chronological age does *not* play a role in itself. While the model developed below concerns optimal aging and death of a representative agent of a cohort, it is nevertheless worth noting that this formulation is in concordance with a central point made by biologists and gerontologists: individual aging is *not* time-dependent. This follows since $\dot{D}(t)$, by (2), is only influenced by current investments and accumulated deficits; chronological age t plays no independent role.

This completes the development of the central equation which governs aging. But before we turn to the analysis of optimal aging and death, we briefly compare our law of motion for deficits with the familiar law of motion from health, which dates back to the work of Grossman (1972).

2.2. **Deficit Accumulation vs Health Accumulation.** Usually, health is introduced as a state variable similar to human capital. In its most basic form, this would involve an equation such as $\dot{H}(t) = I(t) - \delta H(t)$, where $I(t)$ is investment in health and $H(t)$ is the stock of health capital (see e.g. Ehrlich and Chuma, 1990).

If this equation is employed as a description of the aging process it contains a rather unfortunate implication: It predicts that health depreciation (δH) is greater when the stock of health is large, which, of course, usually would mean when individuals are relatively young. In reality, the process of aging is a process where the rate of decline in health status accelerates during life; as explained above, both health deficits and the mortality rate rises exponentially with age, implying slow aging early in life and rapid deterioration in latter stages. In practise, therefore, health losses are greater in states where the health index is *low*, which usually means when one reaches a more advanced age.[9]

Naturally, this problem has not gone unnoticed. Since the work of Grossman (1972), the standard "fix" has been to introduce an age-dependent rate of health depreciation, $\delta(t)$, with $\dot{\delta}(t) > 0$. While the rate of change in the rate of depreciation may seem intuitively appealing the question remains how *exactly* the function is to be specified. Insofar as one does not get this

---

[9]Of course, in other applications beyond aging the law of motion may be perfectly reasonable. As a macro representation, for instance, the depreciation term may reflect lost health capital from diseased individuals. Our critique of the equation is confined to its relevance in the context of modeling the aging process

choice "right" the resulting model will represent an inaccurate description of the aging process, and, by extension, its predictive power will be diminished.[10]

However, even with an age-varying depreciation rate the law of motion for health still holds a counterfactual implication: an improvement in the agent's health status (H) accelerates health depreciation ($\delta(t)H$). The model implies that investments that improve health accelerate aging. This effect tends to make health investments unattractive when the stock of health is high, which in practise means that early-in-life health investments become unappealing.

The approach proposed above, by way of contrast, provides a physiological foundation for health depreciation. To convert our equation for deficit accumulation into one for health accumulation, assume health is defined as best attainable health minus accumulated frailties: $H = \bar{H} - D$ where $\bar{H}$ is "maximum health"; the state of health of a normal 15 year old, say. This implies $\dot{H} = -\dot{D}$. Inserting equation (1) into into $\dot{H}$ and substituting D for $\bar{H} - H$ provides $\dot{H} = -\mu\left(\bar{H} - H - E\right)$, or,

$$\dot{H}(t) = \mu E - \mu(\bar{H} - H(t)). \tag{3}$$

Following the approach above, we may associate E with investments. Finally, using the terminology of Grossman (1972) and Ehrlich and Chuma (1990) we may define the minimum level of health below which life is infeasible as $H_{min} \equiv \bar{H} - \bar{D}$, where $\bar{D}$ is maximum deficits. We are thus left with a simple linear differential equation for health, which differs in one crucial respect from the one adopted in the economics literature: Consistent with the facts, the equation predicts that health loss is small at a good state of health and increasing losses are predicted when the health stock deteriorates. Hence, a meaningful description of how health deteriorates with age (as the health index erodes) is implicit in our frailty equation. Notice that according to equation (3) an improvement in health status works to slow down aging; i.e., it stimulates health accumulation.

The key advantage of our modeling approach is that various parameters, μ for example, have clear empirical counterparts. This is also the case for the frailty index itself, which has already

---

[10]Making $\delta(t)$ an *explicit* function of time also implies that chronological age is supposed to matter to aging *per se*. This is a problematic assumption from a biological perspective, as explained above. This position of biologists is supported by work of health economists. For example, Zweifel et al. (1999) demonstrate that among the elderly health expenditure is not predicted by chronological age once "time remaining until death" is controlled for. This suggest that health status (e.g., frailty), and not the year on the birth certificate, is what matters to health investments.

been developed and tested in the natural sciences. These aspects are very useful in the context of calibration and will be valuable when the model is taken to the data.

3. A Theory of Optimal Aging and Death

3.1. **The optimization problem.** Consider an adult maximizing utility from consumption $c(t)$ over his or her life. We are considering a representative member of a cohort, for which reason the maximization problem can be viewed as deterministic.[11] The initial age is for convenience normalized to zero. Longevity T is finite and endogenous. Let $\rho \geq 0$ denote the rate of pure time preference; the rate of time preference need not be strictly positive in order for the problem to be well-defined. Summarizing, intertemporal utility is given by (4)

$$\int_0^T e^{-\rho t} u(c(t)) \ dt \tag{4}$$

with $u(c) = (c^{1-\sigma} - 1)/(1-\sigma) + b$ for $\sigma \neq 1$ and $u(c) = \log(c) + b$ for $\sigma = 1$. The case of $\sigma > 1$ might lead to perverse results. If, additionally, consumption is below unity such that $c^{1-\sigma} > 1$, flow utility $u(t)$ becomes negative and the individual may prefer a shorter life. In order to exclude this degenerate case we introduce, inspired by Hall and Jones (2007), a constant $b \geq 0$ which is assumed to be large enough to guarantee positive utility. For our numerical results discussed below, it turns out that c is never below unity which allows us to assume $b = 0$. In this case flow utility converges smoothly to $\log(c)$ for $\sigma \to 1$.

It is perhaps tempting to allow frailty to enter the per period utility function as well. We will nevertheless refrain from doing so. In this manner we are much more constrained when trying to match the cross-country data on longevity and income in Section 5 below. Nevertheless, as will become clear, the simple model matches the cross-country data remarkably well. In only allowing health investments to affect utility via longevity, our approach is similar to that of Becker (2007).

The individual receives a wage income w. We assume the wage rate is constant during life; a simplifying assumption which serves to highlight the central workings of the model. Income can be spend on consumption goods c or on health goods h. The relative price of health goods is p. While consumption goods are directly utility enhancing, health goods (e.g. a hip replacement,

---

[11]The life-cycle problem for an individual member of a cohort could be solved by assuming a stochastic process for deficit accumulation and solving the problem – as, for example, in the business cycle literature – by imposing a certainty equivalence assumption. By focusing on a deterministic framework we are following Grossman (1972) and Ehrlich and Chuma (1990).

a weekend at the spa) are instrumental in repairing or delaying bodily decay and, ultimately, in prolonging the period during which consumption goods can be enjoyed. Accordingly, the only value of "good health" in this model is via its impact on longevity. Besides spending income on final goods, the individual may invest in capital k and receive a net interest rate r. The individual takes all prices as given. The law of motion for individual wealth is thus given by (5).

$$\dot{k}(t) = w + rk(t) - c(t) - ph(t). \tag{5}$$

The individual is assumed to inherit wealth $k(0) = k_0$, and to leave a bequest $k(T) = \bar{k}$ (which both could be zero).

The problem is to maximize (4) subject to the accumulation equations (2) and (5), the initial conditions $D(0) = D_0$, $k(0) = k_0$, and the terminal conditions $k(T) = \bar{k}$, $D(T) = \bar{D}$. The problem can be solved by employing optimal control theory; the state variables are $k(t)$ and $D(t)$ and the control variables are consumption $c(t)$ and health investments $h(t)$. From now on time indices are suppressed in the interest of brevity.

3.2. **Optimal Aging.** From the first order conditions we obtain the Euler equations

$$g_c \equiv \frac{\dot{c}}{c} = \frac{r - \rho}{\sigma} \tag{6}$$

$$g_h \equiv \frac{\dot{h}}{h} = \frac{r - \mu}{1 - \gamma}. \tag{7}$$

While equation (6) is the standard Consumption Euler equation, equation (7) provides a novel Euler equation for health expenditures: the "Health Euler equation". As in the context of non-health expenditure a higher intertemporal marginal rate of transformation (i.e., a higher real rate of interest) calls for growing health expenditures. At the same time, growth in health expenditures is tempered by the force of aging, μ. The intuition is that if μ is high, deficits will accumulate very fast at the end of life, making late-in-life health investments a relatively ineffective way of prolonging life. Instead, the optimal strategy is to invest more heavily early in life. Conversely, if the force of aging is low (i.e., μ is "small") late-in-life health expenses are more effective in prolonging life, for which reason it can be optimal to allow h to grow over time. Finally, growth of health expenditure is also influenced by γ, which captures the curvature of the health investment function: a larger γ increases growth in health expenditures, *ceteris paribus*. Intuitively, if γ is "small", diminishing returns set in rapidly, which makes it optimal to

11

smooth health expenditure to make the deficit-reducing effect as large as possible. Thus, in this setting where health does not enter the per period utility function it is the state of technology that determines the extent of expenditure smoothing across life.

It is interesting to note that it is not *necessarily* optimal to plan for increasing health expenditures during life. In societies where the force of mortality is sufficiently strong it may be the optimal strategy to prioritize early-in-life health spending, and thus allow health spending to *decline* over time. One way to think about this result is to associate early-in-life investments in health with "preventive measures" and late-in-life health expenses with "treatment measures". If so, then the Health Euler simply says that in societies where individuals are aging at a rapid pace ($r < \mu$) it is optimal to focus resources on prevention, rather than on treatment.

Overall, the model predicts that countries that are technologically similar, and are inhabited by genetically similar populations should exhibit similar investment patterns across the life cycle. Is there any evidence to support this prediction? Following e.g. Hall and Jones (2007) and Fonseca et al. (2009), we may try to gauge life-cycle developments by examining per capita spending on health across age groups. Figure 1 illustrates such data for four "Western Offshoots": Australia, Canada, New Zealand and United States.

The immediately visually arresting theme is that the four trajectories appear to be more or less parallel, suggesting similar per capita spending growth. Indeed, if one calculates the slope of the trend, the average annual growth rates across the life cycle are nearly identical: 2.0% (Australia), 2.1% (Canada), 1.9% (New Zealand) and 2.0 % (USA). Hence, while the four countries differ in terms of the level of health spending per capita, the rise in health spending per capita across life is nevertheless strikingly similar.[12] These patterns are consistent with the predictions of the health euler derived above.

Even if it is optimal to allow health expenditure to rise over time, the expenditure *share* for health, $\epsilon_h \equiv h/(h+c)$, may nevertheless fall, if pure consumption is growing sufficiently rapidly. Hence, the expenditure share for health is increasing over time if $g_h > g_c$, or (using (6) and (7)) if

$$r - \frac{1-\gamma}{\sigma} \cdot (r - \rho) > \mu, \tag{8}$$

---

[12]That the countries otherwise seem to behave differently in the context of health is nicely illustrated by their aggregate health shares. In 2006 the share of total health spending in GDP was 8.7% in Australia, 10.0% in Canada, 9.4% in New Zealand and 15.3 % in the US.

Figure 1: Health expenditures per capita across age groups: Selected countries 1999-2001
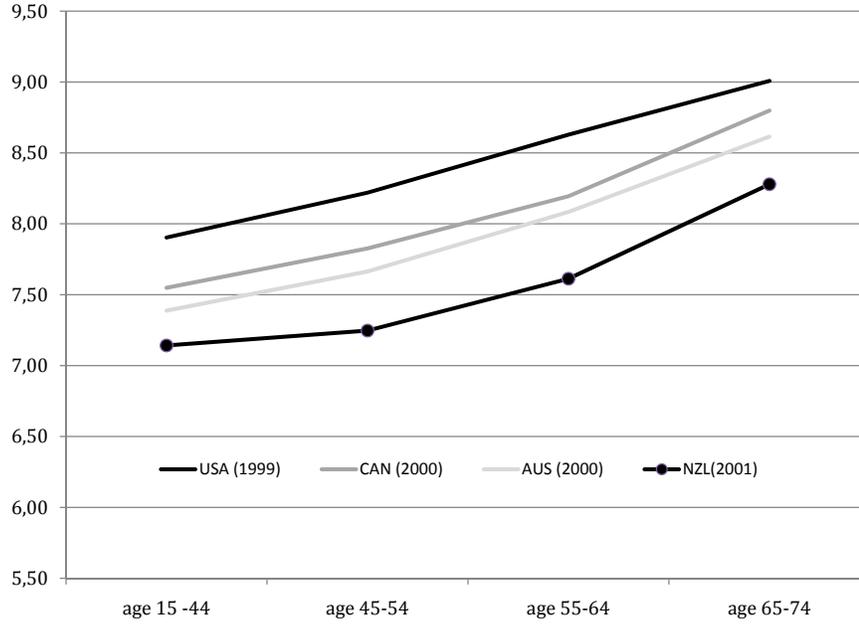
The figure shows log health expenditures per capita by age-group in Australia,
Canada, New Zealand and the USA. *Sources*: US from Keehan, Lazenby, Zezza,
and Catlin (2004); Canada from Health Canada (2001); Australia from Australian
Institute of Health and Welfare (2004); for New Zealand from New Zealand Ministry
of Health (2004). *Notes*: (i) In order to consolidate the age intervals for the purpose
of illustration simple averages across age intervals and gender have been invoked. (ii)
For the USA the first age group concerns individuals aged 19-44. (iii) The year of
data collection varies slightly, as indicated by the legend to the figure: The data for
the US is from 1999; data for Australia and Canada is from 2000 and data for New
Zealand concerns 2001.

If the ratio $\frac{1-\gamma}{\sigma}$ increases it becomes less likely that the condition is met. The interpretation is
that a small σ leaves little incentive to smooth consumption, which implies (*ceteris paribus*) faster
growth in consumption. Meanwhile, γ parameterizes the incentive to smooth health investments;
the smaller γ is the greater the incentive to smooth health investments. Consequently, when $\frac{1-\gamma}{\sigma}$
increases it implies a greater desire, on the part of the consumer, to smooth health investments
relative to consumption, which suggests a declining share of health during the life cycle as long
as $r > \rho$.

A higher rate of interest will increase the rate of growth of both health investments and
consumption. But whether a higher r makes it more likely that the condition is fulfilled depends
on the incentive to smooth consumption and health investments, respectively. If consumers

have a greater desire to smooth health investments ($\frac{1-\gamma}{\sigma} > 1$) a higher r reduces the health expenditure share.

This leaves the impact of ρ and μ. Health spending is independent of ρ since h prolongs life, but does not affect per period utility, while a higher ρ (as usual) works to slow down consumption growth.[13] Hence, it becomes more likely that the health share is rising if individuals are highly impatient. At the same time, in the presence of a greater force of aging people will act more impatiently vis-a-vis health investments, inducing them to invest early in life. If the force of aging is sufficiently strong, the expenditure share for health will therefore be declining over the life cycle.

The bottom line is that the life cycle path for the health share is ambiguous. In particular, it is not a given that the expenditure share should be rising with age, and in theory one may expect differences in the health share across countries that differ in terms of technology and physiology.

3.3. **Optimal Death.** In contrast to the pioneering contributions on the topic of optimal health investments (Grossman, 1972; Ehrlich and Chuma, 1990), the present model is sufficiently simple to allow for an explicit solution of the involved differential equations, from which we can then infer optimal longevity T. In order to see this, first note that the boundary value problem with variable terminal value T requires that the boundary conditions $D(0) = D_0$, $k(0) = k_0$, $k(T) = \bar{k}$, $D(T) = \bar{D}$ and $h(T) = 0$ are fulfilled. Then integrate (5) in order to solve for $k(T)$ and integrate (2) in order to solve for $D(T)$. Finally solve the associated Hamiltonian for $H(T) = 0$. This provides the following set of equations (see Appendix A for details):

$$\bar{D} = D_0 \exp(\mu T) - \frac{\mu A h(0)^\gamma \exp(\mu T)}{g_D} \left[\exp(g_D T) - 1\right] - a \left[\exp(\mu T) - 1\right], \tag{9a}$$

$$\bar{k} \exp(-rT) = k_0 - \frac{w}{r} \left[\exp((r)T) - 1\right] - \frac{c(0)}{g_c - r} \left[\exp((g_c - r)T) - 1\right] - \frac{ph(0)}{g_D} \left[\exp(g_D T) - 1\right] \tag{9b}$$

$$0 = u_T - \frac{\exp(-\sigma g_c T)}{c(0)^\sigma} \times \tag{9c}$$

$$\left\{ \frac{(\bar{D} - a)}{\gamma A} ph(0)^{1-\gamma} \exp((1-\gamma)g_h T) - \frac{1-\gamma}{\gamma} ph(0) \exp(g_h T) - w - r\bar{k} + c(0) \exp(g_c T) \right\}$$

---

[13]Appendix A offers a more detailed discussion of why g_h is independent of ρ.

where $g_D \equiv (\gamma r - \mu)/(1 - \gamma)$ and $u_T \equiv \log(c(0)) + g_c T + b$ in the case of log-utility and $u_T \equiv [c(0)\exp(g_c T) - 1]^{1-\sigma}/(1 - \sigma) + b$ otherwise. These three equations can be solved for the three unknowns: $c(0)$, $h(0)$, and T. Having found the optimal initial values and the optimal terminal time, the four-dimensional dynamic system (2) and (5) – (7) is fully specified and it can be solved for the optimal life-cycle trajectories of c, h, k and D.

4. Comparative Dynamics

4.1. **Calibration and Model Predictions for the US.** Before we turn to the experiments, we calibrate the model to USA data. In this endeavor we begin by employing the Health Euler (7) in order to calibrate the curvature of the health production function, i.e. γ. Using data from Figure 1 we put $g_h = 0.021$, to capture the growth of health spending across the life cycle. From Mitnitski et al. (2002a) we take the estimate of $\mu = 0.043$ for (Canadian) men, and finally we put $r = 0.06$ (e.g., Barro et al., 1995).[14] This produces $\gamma = 1 - (r - \mu)/g_h = 0.19$, which squares well with the independent estimates obtained by Hall and Jones (2007).[15]

Mitnitski et al's regressions do not involve children. Much like the Gompertz-Makeham law for mortality, individuals below roughly the age of 20 are presumably not well described by the law of frailty (in stark contrast to the group above 20). Hence, when calibrating the model we assume individuals are "born" at the age of 20.

With this in mind we do the following in order to parameterize the deficit accumulation equation. From Mitnitski et al.'s (2002a) regression analysis we can back out $D(0) = D(20) = 0.0274$ as the relevant initial value for a 20 year old and $\bar{D} = 0.1005$ 55.6 years later; the life-expectancy of a male US American in the year 2000 was 75.6. In 1900 the life expectancy of a 20 year old U.S. American was 42 years (Fries, 1980). We will think of 1900 as characterized by virtually no health investments. Hence, we therefore set $a = 0.013$ so that the model predicts a life-expectancy at 20 of 42 years, assuming *no* health investment.

Since consumption tends to be essentially constant across the life cycle, once family size has been taken into account (Browning and Ejrnæs, 2009), we put $\rho = r$. The most natural

---

[14]As explained in Section 2: the force of aging within the US and Canada are similar (Rockwood and Mitnitski, 2007). Thus, using the estimate from the Canadian sample should be a good approximation. While Rockwood and Mitnitski (2007) stress the similarity of their results for US and Canadian populations they do not report the detailed results for their US analysis, for which reason we are forced to rely on the results from the Canadian sample.

[15]Hall and Jones allow the curvature of the health production function to be age dependent. The average value is close to 0.2.

specification of the intertemporal elasticity of substitution is probably unity (Chetty, 2006). So we set utility to be logarithmic for the benchmark case. In order to focus on health expenditure as a motive for savings we assume $k(0) = k(T) = 0$ for the baseline simulation. Hall and Jones (2007) emphasize that the presence of autonomous utility $b > 0$ is essential in their theory to produce the result that the health expenditure share is increasing with age. By way of contrast, the present theory does not rely on $b > 0$, the model predicts an empirically plausible association of health expenditure and age, through a different, novel channel: the law of increasing frailty. To emphasize this point we set $b = 0$ in our benchmark calibration.[16]

Finally, we set w equal to GDP per worker in the US in the year 2000 (PPP$ 77,003) and adjusted A such that the individual dies with deficits $\bar{D}$ at age 75.6, which was the life-expectancy of a 20 year-old US American male in the year 2000. This provides the estimate $A = 0.0013$.

In sum, the model is calibrated to match initial deficits, end-of-life deficits, and longevity in 2000; $D(0)$, $\bar{D}$ and T, respectively. It also matches the growth of health spending across age groups exactly. Yet the *path* of deficits (between age 20 and death), as well as the evolution of expenditure *shares* across the life cycle, are left unrestricted. This provides the opportunity for an informative consistency check of our approach: does the path of the health expenditure share and cohort frailty match the data?

Figure 2 shows the basic run of the model; including the association between model predictions and actual data. In the lower right panel of Figure 2 stars indicate actual health expenditure shares by age-group inferred from Mazzocco and Szemely (2010), which can be compared to the solid line representing the prediction from the calibrated model. Admittedly, our calibration fails to match the actual expenditure for the oldest age group (0.28 for ages 65+). We conjecture this failure is mainly caused by the fact that our model does not allow wages to decline with age. Nevertheless, the model does well on average. If we calculate the age-structure weighted health expenditure share from the data it comes to 10.8%. Taking the mean age of the male US citizen in 2000 (35.4) and plugging it into the model we match this number exactly, as indicated by the red dot in the figure.

In the upper left window the accumulation of deficits for the average US (male) citizen is depicted. The dotted line reflects the scenario where there are no health investments occurring; the solid line involves optimal health investments as predicted by the model. In the figure we

---

[16]In the Appendix we discuss alternative specifications of σ and b.

FIGURE 2: OPTIMAL AGING: BASIC RUN
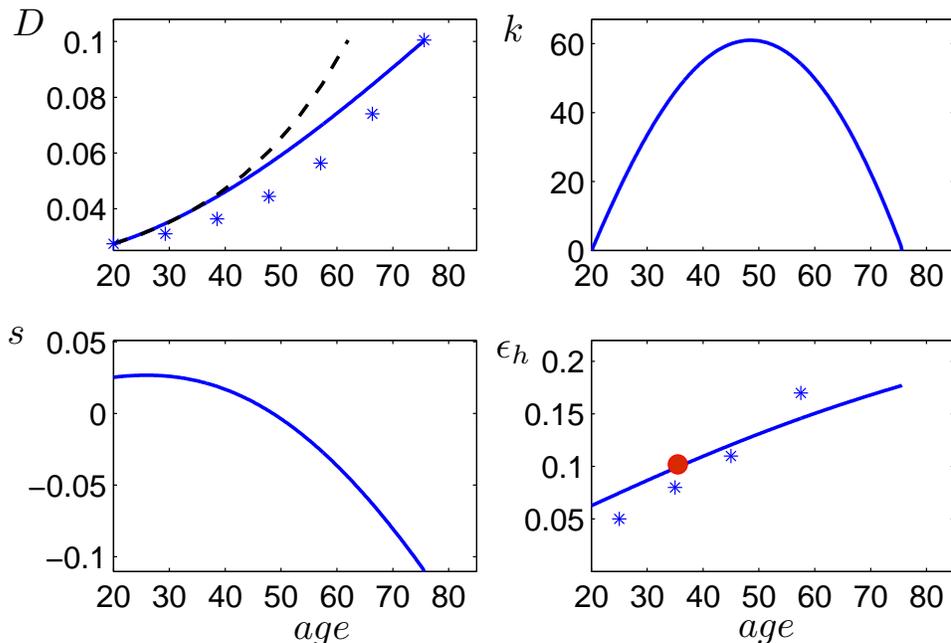
Solid lines: basic run. $T = 55.6$. Parameters: $a = 0.013$, $\gamma = 0.19$, $A = 0.0013$, $\mu = 0.043$, $\rho = r = 0.06$, $D_0 = 0.0274$, $\bar{D} = 0.10$, $p = 1$. Dashed line: no health expenditure ($A = 0$) stars: data, red dot: US mean male age and age-structure-weighted expenditure share.

also illustrate the law of frailty, as estimated by Mitnitski et al. (2002a) (represented by stars). As can be seen, the model's fit is rather good in that the path of deficits is fairly close to the one found in the data, though we are slightly overestimating the level of deficits during life. But on the whole the model seems to do well in predicting the evolution of spending shares on health across the life cycle, and the evolution of deficits (i.e., the aging process) for the average US citizen.

As another check, we can employ the model to calculate the value of life (VOL) of a US citizen at different age intervals, so as to compare the present frameworks predictions with previous estimates. Using the transformation $\tilde{V}(t) = \int_t^T e^{-\rho(\tau - t)} u(c)/u'(c)\mathrm{d}\tau$ we can compute the VOL at age t:

$$\tilde{V}(t) = \frac{\exp(\rho t)\left[\log(c_0) + b\right]c_0}{g_c - \rho}\left(\exp(g_c - \rho)T - \exp(g_c - \rho)t\right)$$
$$+ \frac{g_c c_0}{(g_c - \rho)^2}\left\{\exp(g_c - \rho)\left[1 - (g_c - \rho)\right]t - \exp(g_c - \rho)\left[1 - (g_c - \rho)\right]T\right\}.$$

The benchmark calibration predicts a VOL of about \$ 11.5 million for a 40 year-old and of \$ 10.3 million for a 50 year-old. These values are larger than, but still comparable in order of

magnitude to, those obtained by Aldy and Viscusi (2008), who estimate a value of life of $ 9.7 million for the age group 35-44 and of $ 8.1 million for the age group 45-54.

Finally, the model also holds predictions about the path of the savings rate and wealth across the life cycle. Contingent on the imposed parameter values health expenditures are rising during life, as we just saw, which requires early-in-life savings so as to finance late-in-life investments. The wealth of the representative individual therefore follows a "hump shaped" trajectory; as seen from the upper right hand side window, supported by positive savings early in life and dissaving late in life (lower left hand side window). These patterns are qualitatively consistent with the standard life cycle theory of consumption. We have not examined whether the model matches wealth and savings quantitatively, as it seems unlikely that we can predict wealth during working years without a careful discussion of e.g. bequest.[17]

FIGURE 3: HEALTH AND WEALTH OVER THE LIFE CYCLE

A: Variation of Labor Income

B: Variation of Wealth
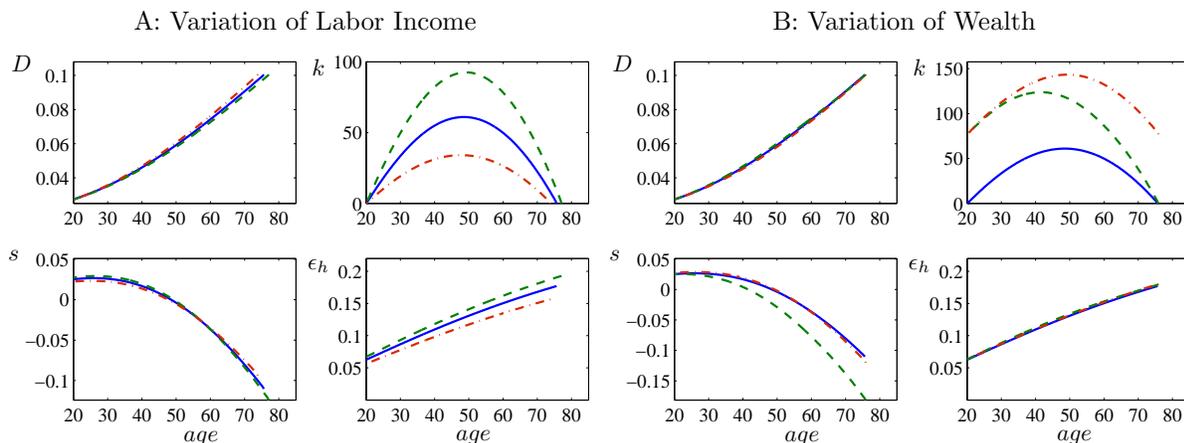
Green (dashed): wage income increases by 1/3. $\Delta T = 1.47$, implied elasticity 0.07). Red (dotted): wage income decreases by 1/3 ($\Delta T = -1.78$, implied elasticity -0.09).

Green (dashed lines): $k(0) = w$, $\bar{k} = 0$ ($\Delta T = 0.28$). Red (dotted): $k(0) = w$ and $\bar{k} = w$ ($\Delta T = 0.45$).

4.2. **Experiment 1: Income.** Panel A of Figure 3 shows how the agent reacts if his income, w, is perturbed. The green (dashed) line is associated with an increase in w of 1/3, the red (dashed-dotted) line depicts the reaction to a reduction of w by 1/3 (in all the experiments below, "green" is associated with increases, and "red" with reductions in the parameter of interest). As can be seen from the figure, the consequence of higher income is an increase in longevity, peak wealth and the share of health spending. In regard to the latter, note that we are keeping

---

[17]The relative importance of life cycle motive vs. the bequest motives for saving is still in debate (e.g., Dynan et al., 2002).

r, ρ, μ, γ constant. Hence the increase in the health share solely reflects a "level effect" through $h(0)/c(0)$. The intuition for why income changes entail a larger change in health spending than regular consumption is that the incentive to smooth the latter is stronger due to diminishing per period marginal utility. Higher income therefore leads to a larger adjustment in $h(0)$ than in $c(0)$.

The issue of main interest, however, is the quantitative impact on longevity. As seen from the top left hand side corner of panel A, the impact is modest though not inconsequential. An increase of income of 1/3 (achievable in a generation with a growth rate of about 1% per year) translates into an increase in longevity of 1.7 years; the reduction involves a fall in longevity of 1.6 years.[18] If we convert the impact into an elasticity - the elasticity of longevity with respect to income - we find it to be around 0.08. These effects are not outlandish in comparison to econometric estimates (e.g., Pritchett and Summers, 1996).

4.3. **Experiment 2: Wealth and Bequest.** In the baseline model we did not allow for wealth transfers. Hence, an interesting question is how longevity changes if the individual receives an inheritance, and is forced to pass on a bequest. The dashed line in Figure 3, panel B, examines the impact from offering the individual a transfer comparable to her annual wage, while she is not obliged to pass on any bequest. The dotted line then examines the effect from forcing the individual to pass on a bequest equal to the initial transfer.

When agents receive an inheritance they are naturally able to invest more in health than otherwise, as it increases their life time income. It is interesting to observe, however, that what really matters is the initial inheritance; if bequests are passed on the simulated increase in longevity is about a quarter of a year (0.27 years), whereas the increase is only marginally higher without the bequest requirement. Why is the bequest requirement so relatively inconsequential? The intuition is that what dominates over a lifetime is the accumulated interest on the initial bequest. A 6% annual interest will double the value of the initial inheritance almost every decade. In this light it is perhaps unsurprising that the paying of a one year's wage in bequest has a relatively minor impact on the results.

4.4. **Experiment 3: Health costs.** Our next experiment concerns health costs; the relative price of h. Here we consider a doubling of the relative price of health goods. Rising relative

---

[18]The system is very non-linear, for which reason the results from our experiments are unlikely to be symmetrical.

health prices is a realistic scenario; the CPI of medical care has risen faster than the GDP deflator in the US (Cutler et al., 1998).[19]

As is clear from Panel A of Figure 4, when the relative price of health increases, individuals substitute towards regular consumption. As a result, the health share declines. With less health investments, savings decline as well. The end result of a doubling of the relative health price for longevity is a reduction by 2.5 years. This amounts to a longevity-price elasticity of 0.09.

FIGURE 4: HEALTH AND WEALTH OVER THE LIFE CYCLE

A: Variation of Health Costs

B: Variation of General Medical Effectiveness (A)
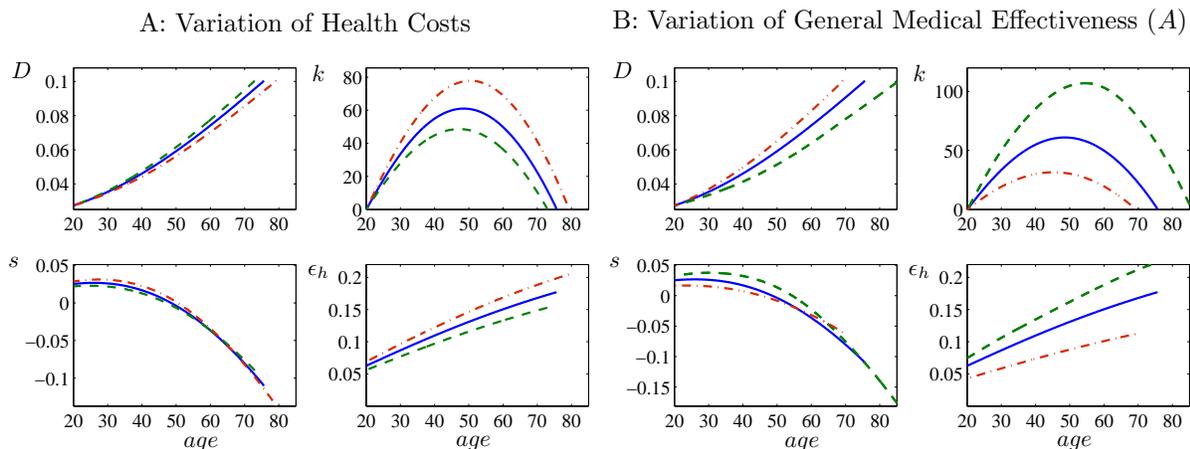
Green (dashed):doubling p, $\Delta T = -2.7$, implied elasticity -0.09). Red (dotted): halving p ($\Delta T = 3.6$, implied elasticity 0.13).

Green (dashed): 33 percent increase of A ($\Delta T = 9.8$, implied elasticity 0.53). Red (dotted): 33 percent decrease of A ($\Delta T = -6.4$, implied elasticity -0.35).

4.5. **Experiment 4: Health Technology.** In Figure 4, panel B, we depict the impact of health technologies on longevity. In the experiments above, the impact from the parameter of interest were indirect. For instance, an increase in income translates into both higher health spending and higher consumption. Medical technologies (or the productivity of health investments more broadly), however, have a direct impact on the evolution of deficits and therefore on longevity. A larger impact is therefore to be expected.

In Panel B we examine the impact from increasing A by $1/3$; an increase of a similar magnitude to that which we analyzed in Section 4.2, in terms of income. If health productivity rises by 33% the consequence is an increase in longevity of nearly a decade. The implied elasticity is about $1/2$. This is a very large effect, suggesting that the impact from improvements in health productivity easily may have towered that of rising income.

---

[19]However, according to Cutler et al., productivity of health care has also risen. Indeed, the authors argue that relative medical prices, appropriately quality adjusted, might have been declining over time. Below we examine the influence from changes in health productivity.

5. The Preston Curve

In his classic paper Preston (1975) documented that there is a positive and concave cross-country relationship between GDP per capita and life expectancy at birth. Subsequent research has found that the Preston curve still holds (e.g., Deaton, 2003), and that the relationship between income and longevity seems to have emerged sometime late in the 19th century (Bloom and Canning, 2001).

While the Preston curve thus continue to hold, questions remain as for its interpretation. Preston (1975) himself argued that an impact from income on longevity could occur in various ways, including via diet, access to clean water and sanitation, as well as medical treatment. At the same time, it is hard to exclude that the same link could be attained with causality running from life expectancy to income. Or, perhaps income is simply correlated with something else that matters to longevity, like human capital (Bloom and Canning, 2007).

Our model establishes a causal link between income (in the form of wages) and life expectancy (in the form of longevity, T). The mediating factor is health investments, h, which would include basic investments like access to clean water and sanitation, as well as more sophisticated investments associated with medical treatment. But is this mechanism sufficient to create the empirical income gradient?

In order to confront this question we begin by estimating the Preston curve for the year 2000. The fitted values from this regression constitute the income gradient to be explained. Naturally, the results from the regression most likely confounds three separate channels: The impact of income on longevity; the impact of longevity on income (reverse causality), and the influence from factors correlated with both income and longevity (omitted variables bias). Our model, however, allows us to isolate the pure effect of income on longevity. Hence, as a second step, we feed the income levels of the countries in our sample through our model, *keeping the parameters we calibrated for the US fixed*, and record the resulting levels of life expectancy. As a third step we then compare the longevity predictions of our model to those associated with the observed income gradient; this comparison allows us to quantify how important the income-to-longevity channel is in accounting for the Preston curve.[20]

---

[20]Naturally, the actions of private agents and governments may not be as optimal as our model suggests. Deviations from optimality will be reflected in deviations between the model and the observed Preston curve. That is, sub optimal behavior will appear in the left-over residual.

The prediction of the model should not be expected to coincide with the Preston curve since it is hard to believe that the income-to-longevity channel is the *only* channel captured by the Preston curve. But the deviation between points on the Preston curve and the prediction of our model will have a clear interpretation: they reflect either the impact from factors that shift the Preston curve and are correlated with income, or, they reflect the opposite chain of causality (i.e., longevity influencing income). In theory, candidate omitted factors would be captured by the relative price of health goods (p) and health efficiency (A); both p and A will shift the Preston curve if perturbed.[21] Our analysis does not allow us to pin down the source of the deviation between our model's prediction and points on the Preston curve (omitted variables vs. reverse causality). But it does allow us to gauge how important they are in accounting for the Preston curve in its entirety: Their influence can be quantified as the fraction of the variation nested in the observed income gradient that our model *cannot* account for.

5.1. **Estimating the Preston Curve.** While the original Preston curve concerns GDP per capita and life expectancy at birth, we obviously need to modify the "input data" slightly. Our model does not involve children; life expectancy at birth is therefore not the optimal empirical counterpart to T. Instead, we collected data on male life expectancy at the age of 20, to retain consistency with our calibration for the US.[22] Unfortunately, life expectancy at age 20 is not recorded consistently each year in all countries; so our actual data involves observations circa 2000. Our income measure is only slightly different from the one employed by Preston (1975) in that we opt for GDP per worker rather than GDP per capita. GDP per worker is arguably a more appropriate empirical proxy for wages w, the natural income measure in our model.

The country sample is restricted in two ways (beyond via availability of the basic data): following Preston (1975) we ignore countries with a population below 2 million citizens, and in addition we omit the OPEC countries; GDP per worker is probably a poor guide to average wages in these countries. This leaves us with a sample of 65 countries in 2000.

With this data in place we proceed by estimating the income gradient. Specifically, we estimate the equation $y_i = f(z_i) + x_i\beta + \epsilon_i$, where the dependent variable is life expectancy at the age of 20, z_i is GDP per worker, x_i is the observation year for life expectancy which adjusts for the

[21]The list of factors that may influence "A" is long and probably includes, among others, technology (if health technology does not diffuse immediately), health care institutions (how efficient is the health care system in translating spending into results?), climate (since it influences disease gradients and thereby longevity for income given) and human capital (if education means that a dollar's worth of health spending is used more effectively).
[22]All data sources are found in Appendix C.

fact that life expectancy is not measured in *exactly* the same year, and ϵ_i is a noise term. The function f is assumed to be a smooth, single-valued function with a bounded first derivative, but data decides the exact relation between GDP per worker and life expectancy.[23] The result is depicted in Figure 5.

Figure 5: The Modified Preston Curve
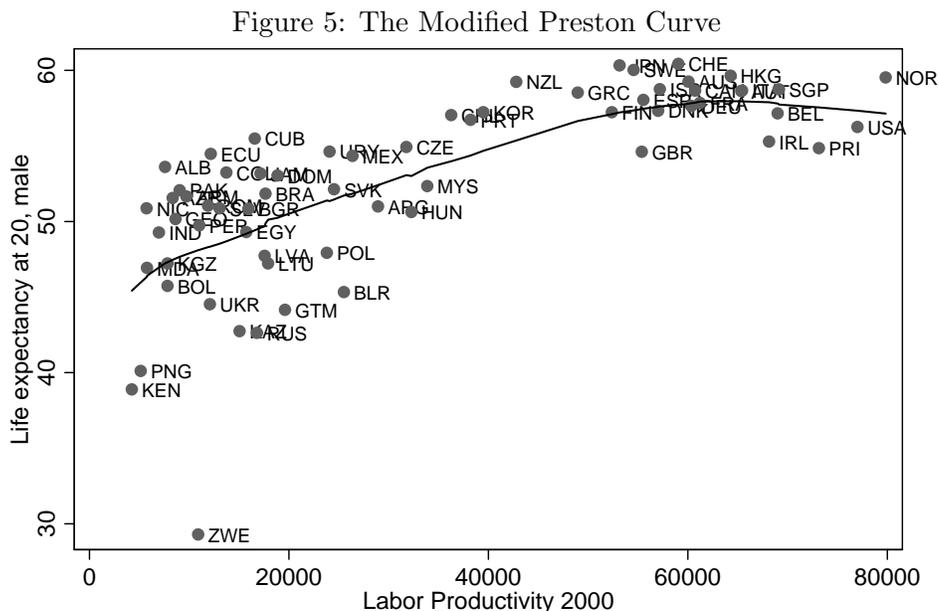
The Figure shows the cross-country link between life expectancy at age 20 and GDP per worker in 2000 (65 countries). *Notes:* The line is estimated semi-parametrically, with year of data collection for life expectancy being the linear control. Labor productivity is significant (p-value of 0.000). See Lokshin (2006) for details on the estimation algorithm. See Appendix C for data sources.

In comparison to the original Preston curve the estimated function is somewhat more linear to behold. This result is likely caused in part by the fact that we are ignoring child mortality, and in part by the fact that data is missing in many of the poorest countries. Nevertheless, we do see a flattening of the income gradient at high levels of GDP per worker; the income gradient is nonlinear and concave. The estimated income gradient, depicted in Figure 6, is the target we compare our model's predictions to.

5.2. **Understanding the Preston Curve.** In order to examine the importance of the pure income channel, in accounting for the Preston curve, we feed the income levels (PPP GDP per worker in 2000) for the 65 countries through the model, keeping the parameters we calibrated to US data fixed. Figure 6 allows for a visual comparison between the model's predictions regarding

---

[23]See Lokshin (2006) for details on the estimation algorithm that we employ.

life expectancy in 2000, and the empirically estimated income gradient from the cross section in 2000, whereas Table 1 provides some summary statistics.

Figure 6: The Model vs. the Preston Curve
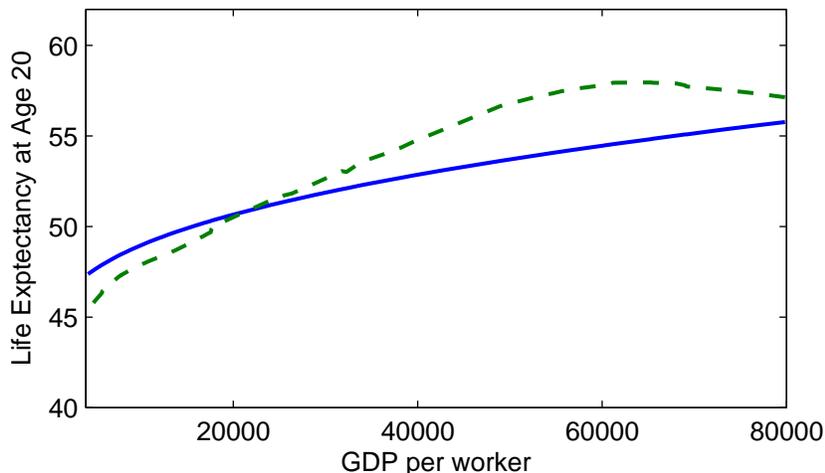
The figure compares the empirically estimated Preston curve (dotted) to the Preston curve predicted by the model (solid).

Table 1: Summary statistics, Model vs. Data

	Data	Model
LE(min)	45.4	47.4
LE(max)	58.0	55.8
LE(mean)	52.4	51.6
LE(median)	51.5	51.3
R^2		0.78

The R^2 is the coefficient of determination: the estimated income gradient consti-
tute data, whereas the model's prediction constitutes "fitted values". The reported
number thus reflects the fraction of the variation in life expectancy along the income
gradient that the model can account for.

As can be seen from the figure, the model (solid line) does a pretty good job at matching the Preston curve (dotted curve). As seen from the table, the model does underestimates the range of life expectancies spanned by the Preston curve somewhat, but at average or median income the actual income gradient and the one predicted by the model essentially coincide. The table also provides a summary measure of the goodness-of-fit of the calibrated model in the form of the coefficient of determination: the R^2 of the income-to-longevity link comes to 0.78, which means that 78% of the variation in life expectancy along the Preston curve is accounted for by our model.

While the bulk of the variation in the observed link between life expectancy and income thus is accounted for, there are systematic deviations between the Preston curve and predicted life expectancy, which accounts for the remaining 22% of the variation. There are two complementary explanations for the systematic nature of the deviations; the first is based on the assumption that "omitted variable bias" is influencing the empirically estimated income gradient, whereas the second would pertain to the case where reverse causality is thought to be important.

Consider the latter case first. If reverse causality is an issue, one may think of the estimated income gradient as the outcome from the interaction of two separate underlying schedules: an income-to-longevity schedule and a longevity-to-income schedule. The former is in theory captured by our model, which by construction does not admit the reverse line of causality. Now, if the longevity-to-income schedule has a *steeper* slope than the income-to-longevity schedule in the (income, life expectancy) space, then the estimated Preston curve will feature a slope that is strictly larger than the income-to-longevity schedule. As a result, our model (capturing only the income-to-longevity mechanism) should overestimate life expectancy at the bottom of the income distribution and underestimate it at the top.

Alternatively, suppose reverse causality is not an issue. If so, then we would interpret deviations between the Preston curve and the model's prediction as the result of omitted variables; factors that are correlated with both life expectancy and income. Theoretically, such factors could map into A (efficiency of health investments), p (relative price of health goods), or both. Our results then suggest that the price level of health in efficiency units (p/A), is higher in most of the poorest countries relative to US. But by the same token p/A must then be larger in the US compared to many of the richest countries, suggesting that the US health care system (at least in 2000) was less efficient than that of many other rich nations. Whether this is true or not is hard to say. But it remains an observable fact that the US constitutes an "outlier" in health expenditures, but not in terms of life expectancy. In any case, *if* omitted variables is the only channel affecting the Preston curve beyond the income–to-longevity channel, the observed difference between the Preston curve and the model's prediction would have to mean that the price of health in efficiency units is higher in many of the poorest places, yet lower in the richest places, relative to its level in the US.

In practise, of course, we have no way of knowing which of these two "stories" is more important in accounting for the left-over residual. What we do know is that they both may influence

the Preston curve, and that they together accounts for some 20% of the variation along the income gradient.

In sum, the analysis suggests that the Preston curve largely, but not exclusively, is due to the causal influence form income on longevity; 78% of the variation in life expectancy along the income gradient is accounted for by our model. This insight leads to an important conclusion. People in the poorest countries are dying earlier than citizen's in rich nations. Our analysis suggests that, to a first approximation, the sad reality seems to be that poor people spend less on health because they are poor and live shorter lives because of it. These conclusions mimic Preston's (1975) own conjecture regarding the underlying forces that generate his curve rather well. That is, the nonlinear link between income and life expectancy is to a large extent caused by lower health investments in several dimensions, ranging from clean water to medical treatment. Changes in relative prices and health technologies do matter. But in order to understand the income gradient, they do not seem to be the main culprits.

6. Concluding Remarks

In the present paper we have proposed a theory of optimal aging and death. Individuals maximize lifetime utility subject to the usual budget constraint, but also taking their physiology into account. The physiological constraint concerns the gradual emergence of health deficits that constitute the aging process. While aging and death are inevitable, individuals can invest in their health which serves to slow down aging and prolongs life. Contingent on preferred health investments the aging process as well as the time of death are determined.

The model holds strong predictions regarding optimal health spending across the life cycle, as well as for the optimal evolution of health expenditure shares. Interestingly, it is not necessarily optimal for health spending, or its expenditure share, to rise during life. The analysis has worked out under what circumstances it is optimal for spending to be rising, constant, or declining during life.

The calibrated model is able to predict spending shares and frailty across age groups, in the US, fairly well. Encouraged by these findings we used the calibrated model to elicit information about the dynamic impact on frailty and longevity from shocks to technology, income and more. One interesting outcome of the simulations were that changes in relative health prices and income have relatively modest effects on longevity. By way of contrast, the impact on longevity from

changes in health efficiency is much larger. This suggests that governments aiming to improve health outcomes might be better off focusing on e.g. health technology and on the institutional set-up of the health sector, than focusing on subsidizing prices on health investments.

In a cross-country setting the calibrated model is able to account well for the celebrated Preston curve. This suggests that most of the observed link between income and longevity across countries can be attributed to variations in health investments. The model does not fully account for the Preston curve; deviations reflect either reverse causality or omitted factors that are correlated with both income and longevity. This residual, however, only accounts for some 20 % of the variation in longevity along the income gradient; the Preston curve mostly reflects the causal impact of income on longevity.

APPENDIX A: DERIVATIONS

Derivation of (6) and (7). The Hamiltonian associated with the problem of maximizing (4) subject to (2) and (5) reads

$$H = \frac{c^{1-\sigma} - 1}{1 - \sigma} + \lambda\mu\left(D - a - Ah^\gamma\right) + \phi\left(rk + w - c - ph\right).$$

For $\sigma = 1$ the first term is replaced by $\log(c)$. The first order conditions wrt. c and h and the co-state equations are

$$c^{-\sigma} - \phi = 0 \quad \Rightarrow \quad c^{-\sigma} = \phi \quad \Rightarrow \quad \sigma\dot{c}/c = -\dot{\phi}/\phi \tag{A.1}$$

$$-\lambda\mu A\gamma h^{\gamma-1} - p\phi = 0 \tag{A.2}$$

$$\lambda\mu = \lambda\rho - \dot{\lambda} \quad \Rightarrow \quad \mu - \rho = -\dot{\lambda}/\lambda \tag{A.3}$$

$$\phi r = \phi\rho - \dot{\phi} \quad \Rightarrow \quad r - \rho = -\dot{\phi}/\phi. \tag{A.4}$$

Equation (A.4) is the well known Euler equation requiring that the shadow price of consumption (ϕ) grows at the rate of the interest rate less the time preference rate. Analogously, the Euler equation (A.3) requires that the shadow price of health grows at the rate of health deterioration (μ) less the time preference rate.

Log-differentiate (A.2) wrt. time and insert (A.3) and (A.4) to obtain optimal growth of health expenditure:

$$\frac{\dot{\lambda}}{\lambda} - \frac{\dot{\phi}}{\phi} = (1 - \gamma)\frac{\dot{h}}{h} \quad \Rightarrow \quad -\mu + \rho + r - \rho = (1 - \gamma)\frac{\dot{h}}{h}.$$

Observe that ρ cancels out. Intuitively, the growth rate of health expenditure depends positively on the growth rate of the the shadow price differential, i.e. the growth rate of λ/ϕ. If the shadow price of health (λ) grows at higher rate than the shadow price of consumption (ϕ), it indicates that the future contribution of health to utility is more important than the future contribution of consumption (both measured relative to current contribution) and thus it is optimal that health expenditure increases over time ($\dot{h}/h$ is positive). Since the time preference enters both equations symmetrically, it has no significance for the *growth* of health expenditure. Of course it will affect the level of health expenditure, see main text. Since r enters only the growth rate of the shadow price of consumption and μ enters only the growth rate of the shadow price of health, they do not drop out but affect growth of health expenditure with opposite sign. Specifically solving for the growth rate of expenditure we get the "Health Euler", i.e. equation (7) of the main text. As usual, (6) is obtained by inserting (A.3) into (A.1). Note also, from (A.2), that the shadow price of health ϕ is negative because the associated stock variable D is a "bad" rather than a good.

Derivation of (9a)-(9c). Begin with noting that, because g_h is optimally constant according to (7), the differential equation (2) can be rewritten as $\dot{D} = \mu(D - a - Ah(0)^\gamma \exp(\gamma g_h t))$. Given $D(0) = D_0$ the solution at time T is:

$$D(T) = D_0\exp(\mu T) - \mu Ah(0)^\gamma\exp(\mu T)\int_0^T \exp(\gamma g_h - \mu)\mathrm{d}t + \mu a\exp(\mu T)\int_0^T \exp(-\mu t)\mathrm{d}t.$$

28

At the time of expiry the boundary condition requires $D(T) = \bar{D}$. Solve the integrals in the above equation to get (9a) in the text. Next, integrate (5) and insert $k(0) = k_0$ and $k(T) = \bar{k}$ to obtain

$$\bar{k} = k_0 \exp(rT) + w \exp(rT) \int_0^T \exp(-rt)\mathrm{d}t$$

$$- c(0) \exp(rT) \int_0^T \exp[(g_c - r)t]\mathrm{d}t - ph(0) \exp(rT) \int_0^T \exp[(g_h - r)t]\mathrm{d}t.$$

Divide by $\exp(rT)$. Note that $g_h - r = (\gamma r - \mu)/(1 - \gamma) \equiv g_D$ and solve the integrals to obtain (9b) in the text.

Finally note that optimal expiry requires that the Hamiltonian assumes the value of zero at T. Otherwise, it would have been optimal to live longer or die earlier. Also, at expiry $D(T) = \bar{D}$ and $k(T) = \bar{k}$. Thus the Hamiltonian reads

$$0 = H(T) = u(c(T)) + \lambda(T)\mu \left[\bar{D} - a - Ah(T)^\gamma\right] + \phi(T) \left[r\bar{k} + w - c(T) - ph(T)\right].$$
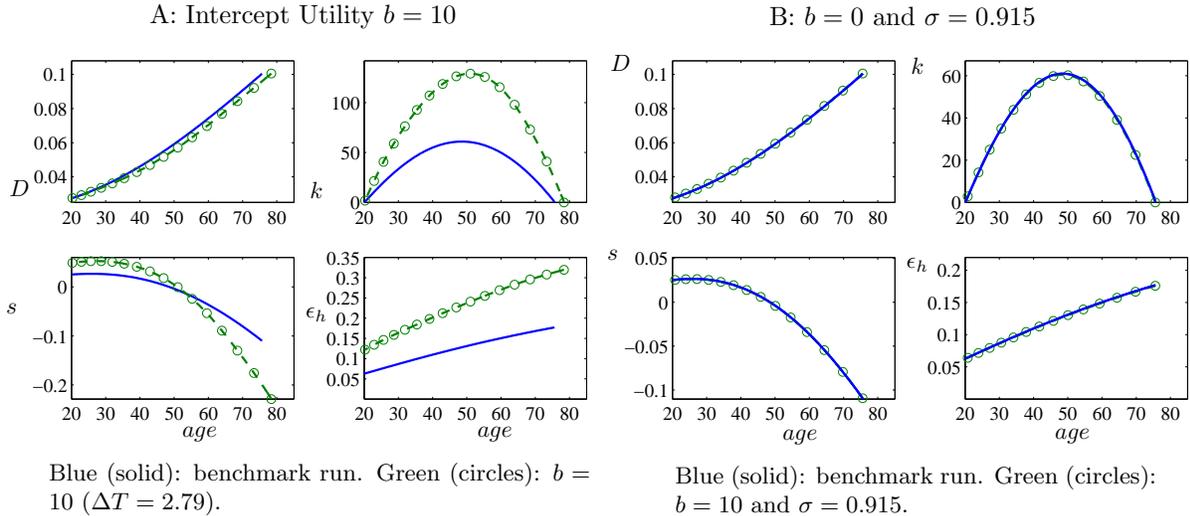
Insert $\lambda(T)$ and $\phi(T)$ from (A.1) and (A.2) to get

$$0 = u(c(T)) - \frac{p}{c(T)^\sigma} \left[\frac{(\bar{D} - a)h(T)^{1-\gamma}}{\gamma A} - \frac{h(T)}{\gamma} - \frac{w + r\bar{k}}{p} + c(T) + h(T)\right]$$

where $u_T \equiv \log c(T) + b$ in the case of log-utility and $u_T \equiv [c(T) - 1]^{1-\sigma}/(1 - \sigma) + b$ otherwise. Noting that $c(T) = c(0)\exp(g_c T)$ and $h(T) = h(0)\exp(g_h T)$ this provides (9c) in the text.

Appendix B: Preferences

Figure A.1: Health as a Luxury Good

A: Intercept Utility $b = 10$ B: $b = 0$ and $\sigma = 0.915$

Blue (solid): benchmark run. Green (circles): $b = 10$ ($\Delta T = 2.79$).

Blue (solid): benchmark run. Green (circles): $b = 10$ and $\sigma = 0.915$.

Panel A of Figure A.1 compares the benchmark run with the solution when health is a "superior good", as in Hall and Jones (2007), reflected by green circled lines showing the solution for $b = 10$. The introduction of an autonomous component in utility raises health expenditure and the slope of the health expenditure curve, as in Hall and Jones. In our case, this implies that the model predicts too much and too steeply increasing health expenditure.

The counterfactually high demand for health can be controlled for by increasing the intertemporal elasticity of substitution. A higher IRS (lower σ) implies that the individual has less preference for consumption smoothing than in the log case. Since we maintain the benchmark assumption $r = \rho$, i.e. constant level of consumption throughout life, the change in preferences can only manifest itself in a higher initial level of consumption. This also brings lower savings, a lower level of health investments and thus less longevity. As Panel B demonstrates reducing σ to 0.915 exactly counterbalances the "superior good" assumption and the predicted life cycle trajectories (green circled lines) coincide with those of the benchmark model (blue solid lines).

Appendix C: Data sources

Data on life expectancy at age 20. We use two samples of data on life expectancy at the age of 20 for males: A "2000" sample and a "1980" sample. The former sample involves observations covering the period 1997–2006. In the 1980 sample we included observations if they were available sometime during the period 1970–1990; in the event several years were available we chose the data point closest to 1980. The data is available in the Demographic yearbook for 2006 (2000 sample) and in the historical supplement to the Yearbook for 1997 (1980 sample). Alternatively, the data can be obtained online at the web address:

http://unstats.un.org/unsd/demographic/sconcerns/mortality/mort2.htm

In our sample, used for estimation, the median year of observation for the "1980" sample, is 1979, with a standard deviation of 4 years. For the "2000" sample the median year of observation is 2003 with a standard deviation of 2.73 years.

Data on Labor productivity. In the regressions we employ GDP per worker (RGDPW) for 1980 and 2000, from Penn World Tables, Mark 6.3.

Data on "p". For the calibration involving shifts in the Preston curve we adjust for plausible changes in p during the period 1980–2000. We calculated the relative price as the ratio between the CPI for "Medical care" and the "all items" CPI. In this regard we obtained data from the Bureau of Labor Statistics, Consumer Price Index Detailed Report Tables; Annual Average Indexes 2000 (Tables 1A-23A), which are available from:

http://www.bls.gov/cpi/cpid00av.pdf

Data on the share of health expenditure in GDP. We obtained data on health expenditures as a share of GDP in 2000, from the World Health Organization. The data is available for free at:

http://apps.who.int/ghodata/

References

Aldy, J.E. and Viscusi, W.K., 2008, Adjusting the value of a statistical life for age and cohort effects, *Review of Economics and Statistics* 90, 573–581.

Arking, R., 2006, *The Biology of Aging: Observations and Principles*, Oxford University Press, Oxford.

Australian Institute of Health and Welfare, 2004. *Australia's health 2004*. Australia's health no. 9.

Barro, R., G. Mankiw and X. Sala-i-Martin, 1995. Capital mobility in neoclassical models of growth. *American Economic Review* 85, 103–115

Becker, G., 2007. Health as human capital: synthesis and extensions. *Oxford Economic Papers* 59, 379-410.

Bloom DE, Canning D., 2001. Cumulative causality, economic growth and the demographic transition. In: Birdsall N, Kelley AC and Sinding SW (eds). *Population Matters: Demographic Change, Economic Growth, and Poverty in the Developing World*. New York: Oxford University Press, pp. 165-97.

Bloom, D. and D. Canning, 2007. Commentary: The Preston Curve 30 years on: still sparking fires. *International Journal of Epidemiology* 36, 498499

Browning, M. and Ejrnæs, M., 2009. Consumption and children. *Review of Economics and Statistics* 91, 93–111

Chetty, R., 2006, A new method of estimating risk aversion, *American Economic Review* 96, 1821-1834.

Cervellati, M. and U. Sunde, 2005. Human capital formation, life expectancy, and the process of development, *American Economic Review* 95, 1653–1672.

Cuttler, D. Mark B. McClellan, Joseph P. Newhouse, and Dahlia Remler, 1998. Are Medical Prices Declining? Evidence from Heart Attack Treatments. *Quarterly Journal of Economics* 113, 991-1024.

Deaton, A., 2003. Health, inequality, and economic development. *Journal of Economic Literature* 41, 113–158.

Dynan, K.E. and Skinner, J. and Zeldes, S.P., 2002. The importance of bequests and life-cycle saving in capital accumulation: A new answer. *American Economic Review* 92, 274–278.

Ehrlich, I. and Chuma, H., 1990. A Model of the Demand for Longevity and the Value of Life Extension. *Journal of Political Economy* 98, 761-82

Eisenring, C., 1999. Comparative dynamics in a health investment model. *Journal of Health Economics* 18, pp. 655660

Fonseca, R., P-C. Michaud, T. Galama, A. Kapteyn., 2009. On the Rise of Health Spending and Longevity. IZA discussion paper No. 4622

Fries, J.F., 1980. Aging, natural death, and the compression of morbidity. *New England Journal of Medicine* 303, 130-135.

Foster, M., 2001. The meaning of death: some simulations of a model of healthy and unhealthy consumption. *Journal of Health Economics* 20, pp. 613638

Galor O. and O. Moav, 2007. The Neolithic Origins of Contemporary Variation in Life Expectancy. Working Paper (Brown University)

Gavrilov, L.A. and Gavrilova, N.S., 1991, *The Biology of Human Life Span: A Quantitative Approach*, Harwood Academic Publishers, London.

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

Hall, R.E. and Jones, C.I., 2007. The Value of Life and the Rise in Health Spending. *The Quarterly Journal of Economics* 122, 39–72

Health Canada, 2001. *Health Expenditures in Canada by Age and Sex, 1980-81 to 2000-01*. Ottawa.

Keehan, S. P, H. Lazenby, M. Zezza, and A.C. Catlin,, 2004. Age Estimates in the National Health Accounts. *Health Care Financing Review* 1, 1–16.

Lokshin, M., 2006. Difference-based semiparametric estimation of partial linear regression models. *The Stata Journal* 6, 377-83

Mazzocco, M. and Személy, B., 2010, The decline of the US saving rate and the rise of health Expenditure, Working Paper, California Center for Population Research.

Meara, E., C. White, and D. M. Cutler, 2004. Trends in Medical Spending by Age, 1963-2000. *Health Affairs* 23, 176183.

Mitnitski, A.B. and Mogilner, A.J. and MacKnight, C. and Rockwood, K., 2002a. The accumulation of deficits with age and possible invariants of aging. *Scientific World* 2, 18161822.

Mitnitski, A.B. and Mogilner, A.J. and MacKnight, C. and Rockwood, K., 2002b. The mortality rate as a function of accumulated deficits in a frailty index. *Mechanisms of ageing and development* 123, 1457-1460.

Mitnitski, A. and Song, X. and Skoog, I. and Broe, GA and Cox, J.L. and Grunfeld, E. and Rockwood, K., 2005. Relative fitness and frailty of elderly men and women in developed countries and their relationship with mortality. *Journal of the American Geriatrics Society* 53, 2184–2189.

New Zealand Ministry of Health, 2004. *Population Ageing and Health Expenditure: New Zealand 2002–2051*.Public Health Intelligence Occasional Bulletin No 22

Preston, S.H., 1975. The changing relation between mortality and level of economic development. *Population Studies* 29, 231–248

Pritchett, L. and Summers, L.H., 1996. Wealthier is healthier. *Journal of Human Resources* 31, 841–868

Ried, W., 1998. Comparative dynamic analysis of the Grossman model. *Journal Health Economics* 17, pp. 383-425.

Rockwood, K. and A.B. Mitnitski, A.B., 2006, Limits to deficit accumulation in elderly people. *Mechanisms of ageing and development* 127, 494–496.

Rockwood, K. and Mitnitski, A.B., 2007, Frailty in relation to the accumulation of deficits. *Journals of Gerontology Series A: Biological and Medical Sciences* 62, 722

Stigler G.J and G.S. Becker, 1977. De Gustibus Non Est Disputandum. *American Economic Review* 67, 76-90

Strehler, B.L. and Mildvan, A.S., 1960. General theory of mortality and aging. *Science* 132, 14-21.

Zweifel, P. and Felder, S. and Meiers, M., 1999. Ageing of population and health care expenditure: a red herring? *Health Economics* 8, 485–496.