Center for European, Governance and Economic Development Research
cege

Discussion Papers

Number 172 – September 2013

HEALTH AND EDUCATION: UNDERSTANDING THE GRADIENT

Holger Strulik

Georg-August-Universität Göttingen

ISSN: 1439-2305

Health and Education: Understanding the Gradient

Holger Strulik*

First Version: December 2011. This Version: September 2013.

Abstract. This study presents a novel view on education and health behavior of individuals constrained by aging bodies. The aging process, i.e. the accumulation of health deficits over time, is built on recent insights from gerontology. The loss of body functionality, which eventually leads to death, can be accelerated by unhealthy behavior and delayed through health expenditure. Education is endogenous and determined, among others thing, by cognitive ability. The proposed theory rationalizes why better educated persons optimally choose a healthier lifestyle. The model is calibrated for the average male US citizen. In the benchmark case a difference in cognitive ability that motivates one more year of education leads to an increase of longevity by about half a year. Progress in medical technology explains why the education gradient gets larger over time.

Keywords: Schooling, Aging, Longevity, Health Expenditure, Unhealthy Behavior, Smoking.

JEL: D91, J17, J26, I12.

---

* University of Hannover, Wirtschaftswissenschaftliche Fakultät, Königsworther Platz 1, 30167 Hannover, Germany; email: strulik@vwl.uni-hannover.de. I would like to thank Carl-Johan Dalgaard, Gustav Feichtinger, Leonid and Natalia Gavrilov, Franz Hof, Wolfgang Lutz, Arnold Mitnitksi, Alexia Prskawetz, Patrick Puhani, Fidel Perez Sebastian, and Engelbert Theurl, for useful comments. This research was funded by the European Commission within the project "Long-Run Economic Perspectives of an Aging Society" (LEPAS) in the Seventh Framework Programme under the Socio-economic Sciences and Humanities theme (Grant Agreement: SSH7-CT-2009-217275).

1. Introduction

Better educated individuals are, on average, healthier and die later than less educated ones. The literature refers to the strong positive association between education and health as the education gradient or just "the gradient". According to one popular study, in the year 1990, US Americans aged 25 with any college education could expect to die 5.4 years later compared to those with only high school or less. By the year 2000 the gap increased to 7.0 more years for the better educated (Meara et al., 2008). In many other countries a similar association between education and health has been observed.[1]

One obvious explanation for the gradient is that the better educated care more about their health. They spend more on preventive care, smoke less, are less obese, and display " healthier behaviors along virtually every margin" (Cutler and Lleras-Muney, 2006). But then, of course, the question arises why do they do that? In particular the fact appears puzzling that the less educated, who are presumably less wealthy, spend more on costly unhealthy activities like smoking and eating beyond metabolic needs.

So far, the literature has suggested three different kinds of deeper explanations for the gradient: common third factors, productive efficiency and allocative efficiency. The "third factor" argument is based on the impact of general attitudes on behavior and becomes particularly intuitive if one thinks of time preference. More patient persons are presumably more willing to delay entry into the workforce for education as well as they are more willing to sacrifice pleasure from unhealthy consumption in exchange for a longer a life. The problem is that, empirically, general attitudes seem to play only a minor role for educational differences in health behavior. Cutler and Lleras-Muney (2010) estimate that attitudes like time preference account for about 10 percent of health behavior, similar to the contribution of health knowledge, whereas income (access to resources) and cognitive ability account for the greatest shares, each for about 30 percent.

The idea of productive efficiency is based on Becker's (1965) commodity theory. It postulates that less educated individuals "produce" less health out of any given inputs of time and medical care (see e.g. Grossman, 1972). Allocative efficiency, in contrast, suggests that less educated individuals use different inputs in health production, presumably because they are less well informed

[1] An incomplete list of the literature on the gradient includes Elo and Preston (1996), Contoyannis and Jones (2004), Case and Deaton (2005), Lleras-Muney (2005), Mackenbach et al. (2008), Conti et al. (2010), and Cutler and Lleras-Muney (2010). See Grossman (2006), Cutler and Lleras-Muney (2006), and Cutler et al., (2011) for surveys of the by now large literature. See Glied and Lleras-Muney (2008) and Cutler et al. (2010) on the rising education gradient.

about their "health technology" (see e.g. Kenkel, 1991). The common theme of both ideas is that less educated people behave less efficiently. If they had only access to the health technology and the knowledge of the better educated, they would care more about their health and live longer.

Acknowledging that the so far available theory certainly has a role in explaining the education gradient, the present paper offers an alternative, novel theory. Inspired by the empirical power of cognitive ability in accounting for health behavior it asks the following question. Assume that individuals share the same attitudes (preferences) and share the same allocative and productive efficiency, namely they are fully rational and perfectly foresighted. Assume that they face different returns to education. How much of the observable education gradient can then be explained by their individual-specific return to education and the implied optimal education and health behavior?

The natural explanation for idiosyncratic differences in the return to education across a population is cognitive ability. Smarter people expect a higher payoff from further education and thus educate more (see e.g. Heckman and Vytlacil, 2001). But since the return to education is potentially influenced by other factors as well, for example, by family background and school quality (e.g. Card, 1991) the present paper deals strictly speaking with "idiosyncratic differences in the return to education". Instead of using this awkward term, however, I will in the following address solely cognitive ability as one important determinant of the return to education and, through this channel, of health behavior and longevity (Deary, 2008; Der et al., 2009; Calvin et al., 2011; Kaestner and Callison, 2011).

Methodologically the present paper is related to the literature on optimal health spending and longevity, to which Ehrlich and Chuma (1990) and Hall and Jones (2007) are presumably the most popular contributions. With contrast to that literature, the present paper considers education and unhealthy consumption as individual choice variables. More importantly the present paper has the distinction of being solidly built upon recent research in gerontology. So far the literature has been built on Grossman (1972) and treated health like human capital, that is like a stock that can be accumulated by investment and that depreciates over time. Without further amendments this means that health depreciation is greater when the stock of health is large, that is when individuals are relatively young and healthy(see also Case and Deaton, 2005 for a critique of the conventional approach). In order to counteract this problem, the conventional literature assumes that the depreciation rate is increasing with age. This ad hoc solution, however, does not get down to the root of the problem. The conventional model, in which $H_{t+1} = I_H + (1 - d_t H_t)$, $\partial d_t / \partial t > 0$, predicts that of two persons of the same age the one in better health loses more health in the next time increment, i.e. $d_t H_t$ is larger for a healthier type endowed with high H_t. This prediction contradicts the gerontological evidence. The evidence in medical science shows that in any age group the loss of health next period is *lower* for those who are currently in better health. In other words the evolution of health is path dependent. At any age, persons in bad health are more likely to suffer from further health deterioration next period. This law of health deficit accumulation can be estimated with high precision (Mitnitski et al., 2006, 2007). It refutes the standard model of health accumulation.

Inspired by the medical evidence the present paper turns the economic health model upside down by stating that humans do not accumulate health but *health deficits*. The arrival of new health deficits increases with the number of deficits that a person already has. Specifically, the law of health deficit accumulation reads $\dot{D} = \mu D - E$, in which μ is the force aging. The law of deficit accumulation has a deep gerontological foundation. As suggested by McFadden (2005) it is built upon an application of reliability theory to the functioning of the human body (see Gavrilov and Gavrilova, 1991). In order to model deficit accumulation the present paper employs the so called frailty index, established by Mitnitski and Rockwood and several coauthors in a series of articles (2002a, 2002b, 2005, 2006, 2007). The frailty index counts the proportion of the total potential deficits that an individual has, at a given age. The list of potential deficits ranges form mild ones like impaired vision to severe ones like dementia. Using the frailty index Mitnitski and Rockwood estimate with an R^2 around 95 percent the rate μ at which health deficits are accumulated. In developed countries the average adult individual accumulates 3-4% more deficits from one birthday to the next. Recently, Harttgen et al. (2013) have shown a strong negative association between education and health deficits by computing the frailty index for 15 European countries and a population stratified by educational attainment.

The present study follows Dalgaard and Strulik (2013) by assuming that the factor E in the law of deficit accumulation, which operates to slow down the aging process, is partly explained by health expenditure. Additionally we will consider that unhealthy behavior speeds up the aging process (by reducing E). Unhealthy behavior and the consideration of an endogenous schooling decision are the main differences to Dalgaard and Strulik (2013). The present paper can thus be seen as the logical next step. While Dalgaard and Strulik focus on the income gradient (the

Preston curve) and ignore education, the present study considers the education gradient and its origin from health spending and unhealthy behavior.

In contrast to the health capital stock, which is a latent concept, health deficits are observable, a fact which allows for a numerical calibration of the model. The precision with which the medical literature estimates the force of aging is probably the greatest advantage of the new approach compared to the traditional approach of health capital accumulation, for which the medical literature provides little guidance on how to pin down the age-dependent health depreciation rate d_t. In the present paper, once the force of aging is calibrated according to the gerontological evidence, there are no degrees of freedom left for the determination of the other parameters of longevity. The physiological foundation of the theory ensures a reliable calibration of the model and provides therewith, perhaps for the first time, a theoretical basis for the quantitative evaluation of the education gradient.

Another deviation from conventional health economics is that the present paper takes the education decision of individuals endogenously.[2] It is shown that this treatment makes a considerable difference for the magnitude of the predicted education gradient. If education is not driven by cognitive ability (or more generally by the return to education) then there is no positive effect on health generated beyond the point at which the time spent on education is optimal from the individual viewpoint. The model thus provides an explanation why the health gradient for differential education of monozygotic twins is found to be relatively small (Fujiwara and Kawachi, 2009; Lundberg, 2013). The reason is that twins are likely to be endowed with similar cognitive ability, which trumps the length of the education period in the determination of human capital and health behavior.

The paper is organized as follows. The next section sets up the model. The standard approach of human capital accumulation is modified to account for the empirical wage for age curve and then integrated together with the opportunity of unhealthy consumption into a gerontologically founded life-cycle model. Section 3 presents the analytical results. It shows that the optimal lifestyle is governed by conditions for (i) optimal expenditure profile (on health and unhealthy goods), (ii) optimal aging (the evolution of the expenditure profile with age), (iii) optimal schooling, (iv) optimal financial management, and (v) optimal death. Most of these optimality conditions are simple enough to allow for an intuitive interpretation.

[2] A recent model in the Grossman (1972) tradition of health capital accumulation that considers exogenous education is provided by Galama and van Kippersluis (2010).

In Section 4 the model is calibrated for a 16 year old male US American in the year 2000. Section 5 present the results. The most interesting experiment is, of course, to vary the return to education, θ. It is shown that an increase of θ that motivates one year more of education results in a gain in longevity of about half a year. The study continues with a series of numerical experiments investigating whether there are alternatives to cognitive ability that could motivate the education gradient. After corroborating the main result with a series of robustness checks the paper finishes by comparing the education gradient for voluntary and compulsory education and by showing that the education gradient widens with ongoing medical technological progress.

2. Model Setup

2.1. **The objective function.** Consider a young adult at the end of the compulsory schooling period. Later on, in the numerical part, this will be a 16 year old with 9 years of education. For simplicity let the initial age at this stage be normalized to zero. The – yet to be determined – date of death is denoted by T. At each age t the person experiences utility from consumption of health-neutral goods $c(t)$ and unhealthy goods $u(t)$. We consider a minimum setup for the elaboration of the health and education nexus in which it suffices to treat health expenditure purely instrumental and not in itself utility-enhancing. Likewise, education is purely instrumental in achieving higher labor income and not in itself utility enhancing. The instantaneous utility function is assumed to be iso-elastic such that life-time utility is given by (1).

$$V = \int_0^T \mathrm{e}^{-\rho t} \left\{ v\left(\; c(t) + \beta u(t) \; \right) \right\} \; \mathrm{d}t \tag{1}$$

with $v(x) = (x^{1-\sigma} - 1)/(1-\sigma)$ for $\sigma \neq 1$ and $v(x) = \log(x)$ for $\sigma = 1$. The inverse $1/\sigma$ is the intertemporal elasticity of substitution. Consumption is measured such that x is always larger than one, implying that at each age utility is positive, a fact that makes a longer life, in principle, desirable.

The parameter β measures how pleasurable consumption of the unhealthy good is. If $\beta > 1$, the person likes unhealthy consumption better than health-neutral consumption. If $\beta < 1$, the person prefers health-neutral consumption and unhealthy goods are consumed only because they are cheaper. Later on, the parameter β is a useful device to pin down the price elasticity of demand for unhealthy goods. In the calibration, cigarettes will stand in for the unhealthy good.

2.2. **Budget constraint.** Let health expenditure be denoted by h. The price of health-neutral goods is normalized to unity, the price for health goods is denoted by p, and the price of unhealthy goods is denoted by q. Total expenditure is thus given by $e = c + ph + qu$. The – yet to be determined – length of the voluntary education period is denoted by s and the predetermined age of retirement is at R. From s to R the individual receives a wage $w(t)$ per unit of human capital. Human capital $H(s,t)$ varies with age t and the time spend on education s. We allow for aggregate productivity growth such that the wage per unit of human capital w grows at rate g_w, which is taken as given by the individual.

For simplicity there are no restrictions on the capital market; the individual can borrow or lend at rate r. An individual that holds capital k thus faces the budget constraint (2).

$$\dot{k}(t) = \chi w(t)H(s,t) + rk(t) - c(t) - ph(t) - qu(t) \tag{2}$$

with $k(0) = k_0$ and $k(T) = \bar{k}$. For simplicity we abstain from modeling a bequest motive such that k_0, and $\bar{k}$ as well as all prices are taken as given by the individual. In (2) χ is an indicator function, $\chi = 1$ for $t \in [s, R]$ and $\chi = 0$ otherwise. People are saving for consumption and health interventions after retirement. Although the arrival of health events is certainly stochastic, we follow the related literature (e.g. Ehrlich and Chuma, 1990, Hall and Jones, 2007) and treat, for simplicity, the problem deterministically. This means that the model neglects a precautionary savings motive and thus potentially underestimates the propensity to save for old age.[3]

2.3. **Health Deficit Accumulation.** Inspired by recent research in gerontology we consider a physiologically founded aging process, according to which aging is understood as increasing loss of redundancy in the human body (Gavrilov and Gavrilova, 1991, Arking, 2006). For young persons the functional capacity of organs is about tenfold higher than needed for mere survival (Fries, 1980). With preceding age and vanishing redundancy in our organism, humans become more fragile. An empirical measure of human frailty has been developed by Mitnitski and Rockwood and various coauthors in a series of articles (Mitnitski et al., 2002a,b; 2005; Rockwood and Mitnitski, 2006). They propose to compute the frailty index as the proportion of the total potential health deficits that an individual has, at a given age. As suggested by aging theory, Mitnitski et al. (2002a) confirm that the frailty index number (the number of health deficits),

[3] Strulik (2012) investigates a stochastic version of the structurally similar model of optimal aging by Dalgaard and Strulik (2013) and shows that the quantitative predictions are relatively insensitive to the consideration of death as a stochastic event.

denoted by $D(t)$ increases exponentially with age t, $D(t) = E + b\mathrm{e}^{\mu t}$. They estimate that $\mu = 0.043$ for Canadian men. This "law of increasing frailty" explains around 95% of the variation in the data, and its parameters are estimated with great precision. Conceptually, the rate of aging μ is given to the adult individual. From a physiological viewpoint, however, it can be explained by applying reliability theory to human functioning (Gavrilov and Gavrilova, 1991). Rockwood and Mitnitski (2007) show that that elderly community-dwelling people in Australia, Sweden, and the U.S. accumulate deficits in an exponential way very similar to Canadians.

In order to utilize the findings of Mitnitski and Rockwood for the present work I begin with differentiating the frailty law with respect to age, $\dot{D}(t) = \mu(D(t) - E)$.[4] Following Dalgaard and Strulik (2013) I next introduce health expenditure by assuming that E is amendable to change by way of deliberate health expenditure. Furthermore E can be diminished by unhealthy behavior. Specifically, I propose the following parsimonious refinement of the process of deficit accumulation:

$$\dot{D}(t) = \mu\left[\; D(t) - a - Ah(t)^{\gamma} + Bu(t)^{\omega} \;\right], \qquad 0 < \gamma < 1, \qquad \omega \neq 1. \tag{3}$$

The parameter a captures environmental influence on aging beyond the control of the individual, 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 efficiency of health expenditure in maintenance and repair of the human body, the parameter γ specifies the degree of decreasing returns of health expenditure. Likewise, the parameter B measures the general unhealthiness of the unhealthy good and the parameter ω (omega) measures the return in terms of deficits from unhealthy consumption. The parameter μ measures the biological rate of bodily deterioration when there is neither unhealthy consumption nor health expenditure. 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.[5]

Initial health deficits are given for the young adult, $D(0) = D_0$. Furthermore, following, Rockwood and Mitnitski (2006), we assume that a terminal frailty exists at which the individual

[4] In order to see that a larger autonomous component E implies less deficits at any given age notice that the solution of the differential equation is $D(t) = (D_0 - E)\,\mathrm{e}^{\mu t} + E = D_0\mathrm{e}^{\mu t} - E(\mathrm{e}^{\mu t} - 1)$, in which D_0 are initial health deficits.

[5] If individuals' investment in health influences E, then one may wonder if the "frailty law" should still work empirically, as E then is expected to exhibit individual-level variation. It should; but the cross-section estimate for E should be interpreted as the average level in the sample in question (see Zellner, 1969).

expires, $D(T) = \bar{D}$. Problem (1) – (3) thus constitutes a free terminal time problem in which the terminal states $k(T)$ and $D(T)$ are known.

2.4. **Education.** The length of the schooling period s is a choice variable for individuals. We take the schooling model of Bils and Klenow (2000) and introduce elements of aging such that the resulting wage for age curve can be calibrated to match the stylized facts. Specifically we assume that human capital of an individual of age t with s years of schooling is given by (4).

$$H(s,t) = \exp\left[\theta \frac{s^{1-\psi}}{1-\psi} + \eta(t-s) - \alpha_1 t\right] - \delta \exp\left[\alpha_2 t\right], \tag{4}$$

The parameter θ and ψ capture the return to education and η is the return to experience (learning on the job). For $\psi > 0$ there are decreasing return to education. The marginal return to education is given by $\theta s^{-\psi}$. Later on we consider variation of θ in order to investigate the impact of cognitive ability on education and health outcomes.

For $\alpha_1 = \delta = 0$ the schooling function boils down to the standard model (Bils and Klenow, 2000). The parameter α_1 controls for the impact of age on the return to education. The parameter δ controls the feedback of age on general human capital. For sufficiently high δ and α_2 human capital starts to decline after a certain age. The parameters α_1, α_2, and δ are conceptualized as being job specific. Most cognitive skills and motor skills start deteriorating around the age of 30 or even earlier (Nair, 2005; Skirbekk, 2004). However, so called *crystallized abilities*, i.e. the abilities to use knowledge and experience, remain relatively stable until most of adulthood and start declining after the age of 60, or even later. As a result some measures of psychological competence, like verbal skills and inductive reasoning start declining "only" around age 50 (Schaie, 1994). We would thus conceptualize an occupation that utilizes predominantly crystallized abilities (e.g. a consultant) to be reflected by a lower value of α_1 compared to an occupation that utilizes predominantly motor skills (e.g. a farm laborer). Structural change that makes occupations utilizing crystallized abilities, on average, more widespread could be described by a secular decline of α_1.

Summarizing, human capital varies with education and age. The return to education, captured by θ is a shifter of the human capital function. Together the three parameters α_1, α_2, and δ are used to estimate the empirical wage for age curve (Murphy and Welch, 1990). It turns out that α_2 is estimated to be of about the same same magnitude as the force of aging μ.

3. Solution

3.1. **Summary.** The problem of the individual is to maximize (1) subject to (2) – (4). In order to solve the problem conveniently it turns out to be helpful to define a measure of aggregate consumption $x \equiv c + \beta u$ and replace c in (1) and (2). Details of the computation are delegated to the Appendix. From the first order conditions we obtain that the optimal solution is characterized by the following, nicely interpretable conditions.

$$u^{\omega-1} = \begin{cases} \frac{\gamma(\beta-q)A}{\omega p B} \cdot \frac{1}{h^{1-\gamma}} & \text{for } \beta > q \\ 0 & \text{for } \beta \leq q \end{cases} \tag{5}$$

$$g_x \equiv \dot{x}/x = \frac{r-\rho}{\sigma} \tag{6}$$

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

$$g_u \equiv \dot{u}/u = \frac{r-\mu}{1-\omega} \tag{8}$$

$$\begin{aligned} \int_s^R \frac{\partial}{\partial s} \mathrm{e}^{-rt} w(t) H(s,t) \mathrm{d}t &= (\theta s^{-\psi} - \eta) \exp\left(\theta \frac{s^{1-\psi}}{1-\psi} - \alpha_1 s\right) \cdot \left\{ \mathrm{e}^{(g_w - r + \eta - \alpha_1)(R-s)} - 1 \right\} \\ &= \exp\left(\theta \frac{s^{1-\psi}}{1-\psi} - \alpha_1 s\right) - \delta \mathrm{e}^{\alpha_2 s} = \mathrm{e}^{-rs} w(s) H(s,s). \end{aligned} \tag{9}$$

3.2. **Optimal Consumption Profile.** Condition (5) constrains the optimal consumption expenditure. Optimal expenditure on health and on unhealthy goods are negatively correlated if there are increasing returns of damage from unhealthy consumption ($\omega > 1$). A person who spends more on health is predicted to indulge less in health-damaging consumption. For unhealthy consumption to take place, $\beta > q$. Given that the price of health-neutral consumption has been normalized to unity, the condition requires that utility derived from a unit of unhealthy consumption exceeds that from health-neutral consumption ($\beta > 1$), or that the unhealthy good is cheaper than the health-neutral good ($q < 1$), or both. For policy it is important to note that demand of unhealthy goods is lower at higher prices and that there exists a preemptive price, $q = \beta$, which deters unhealthy consumption. If unhealthy consumption is optimal, then condition (5) furthermore predicts that its incidence is large if medical efficiency in repairing damage is large (A is large), if the resulting health damage is low (B is low), or if the price of health goods p is low.

3.3. **Optimal Aging.** The Euler equations (6)–(8) show how optimal expenditure evolves through life. Together they determine optimal aging of the person since the evolution of health deficits D depends on health behavior (h and u). Condition (6) is the familiar Euler equation for consumption, here stated for the aggregate measure of consumption x. It has the usual textbook interpretation. Equation (7) is the "Health-Euler" (Dalgaard and Strulik, 2013). Similar to the "Consumption-Euler" it suggests to postpone expenditure for health to later periods of life in favor of financial investment if the return on investment r is relatively high. If, on the other hand, the force of aging, μ is high, implying that health deficits accumulate very fast at the end of life, late-in-life health investments are a relatively ineffective way of prolonging life. It is then optimal to invest more heavily early in life. The dynamic expenditure profile for health growth is also influenced by γ, which captures the curvature of the health investment function: a larger γ implies a higher growth rate of health expenditure. Intuitively, if γ is small, diminishing returns set in rapidly, which makes it optimal to smooth health expenditure.

Although theory does not exclude the reverse, it makes sense in light of the empirical background to assume that r exceeds μ such that health expenditure rises with age. For $\omega > 1$, condition (8) prescribes that expenditure for unhealthy consumption should decrease with age. Intuitively, the damage done by, for example, alcohol consumption is relatively harmless at young age when there is plenty lot of redundancy in the body. At an advanced age, drinking a lot could be lethal and the model recommends to reduce drinking to an occasional glass of red wine. The expenditure profile requires, according to (5), that unhealthy consumption is negatively correlated with health expenditure. This implies that the expenditure profile is steep (in absolute value) if the interest rate is high and the rate of deficit accumulation is low.

3.4. **Optimal Schooling.** The optimal length of the education period s requires that the marginal loss from postponing entry into the labor market, $\mathrm{e}^{-rs}w(s)H(s,s)$, equals the marginal gain from extending education, $\int_s^R \frac{\partial}{\partial s}\mathrm{e}^{-rt}w(t)H(s,t)\mathrm{d}t$. Inserting the respective values leads to condition (9), in which g_w denotes the growth rate of the wage per unit of human capital, $w(t) = \bar{w}\exp(g_w t)$. It is assumed to be given by the rate of aggregate productivity growth and to be exogenous to the individual. The first line of (9) displays the marginal gain from education and the second line displays the marginal loss from postponing entry into the labor market.

Observe that the solution of (9) is independent from life span T, implying that – as long as retirement age R is smaller than T, which is henceforth assumed – there is no influence of longevity on education. This means that any causality behind the education gradient runs from increasing education to increasing health and longevity. Shutting off the – certainly existing – reverse causality is helpful in order to identify the size of the gradient that can be explained by increasing education.

3.5. **Optimal Financial Management.** Integrating (2) and inserting initial and terminal values provides the life-time budget

$$k_0 + W(s,R)\mathrm{e}^{-rs} - \frac{x(0)}{g_x - r}\left(\mathrm{e}^{(g_x-r)T} - 1\right) - \frac{ph(0)}{g_D}\left(\mathrm{e}^{g_D t} - 1\right) - \frac{(q-\beta)u(0)}{g_\omega}\left(\mathrm{e}^{g_\omega T} - 1\right) = \bar{k}\mathrm{e}^{-rT} \tag{10}$$

with $g_D \equiv (\gamma r - \mu)/(1-\mu)$ and $g_\omega \equiv (\omega r - \mu)/(1-\omega)$. The expression $W(s,R)$ denotes life-time labor income (human wealth) acquired between leaving school and retirement. It is obtained as (11).

$$\begin{aligned} W(s,R) = \int_s^R \mathrm{e}^{-rt} w(t) H(s,t) \mathrm{d}t = \frac{\bar{w}\exp\left(\theta s^{\frac{1-\psi}{1-\psi}} - \eta s\right)}{\eta + g_w - r - \alpha_1} \cdot \left[\mathrm{e}^{(\eta+g_w-r-\alpha_1)R} - \mathrm{e}^{(\eta+g_w-r-\alpha_1)s}\right] \\ - \frac{\bar{w}\delta}{\alpha_2 + g_w - r}\left[\mathrm{e}^{(g_w+\alpha_2-r)R} - \mathrm{e}^{(g_w+\alpha_2-r)s}\right]. \end{aligned} \tag{11}$$

3.6. **Optimal Death.** At the individually optimal time of expiry two conditions have to hold. The accumulated health deficits must have reached the terminal value $\bar{D}$ and the Lagrangian associated with Problem (1)– (4) must assume the value zero. Turning towards the first condition, integrating (3) provides $D(T)$ and thus (12).

$$\bar{D} = D(T) = D_0\mathrm{e}^{\mu T} - a\left(\mathrm{e}^{\mu T} - 1\right) - \frac{\mu A h(0)^\gamma \mathrm{e}^{\mu T}}{g_D}\left(\mathrm{e}^{g_D T} - 1\right) + \frac{\mu B u(0)^\omega \mathrm{e}^{\mu T}}{g_\omega}\left(\mathrm{e}^{g_\omega T} - 1\right). \tag{12}$$

Evaluating the Lagrangian at T provides the condition (13).

$$\begin{aligned} 0 = v(T) + \xi\bar{D}^\nu + x(T)^{-\sigma}\left\{r\bar{k} - x(T) - (q-\beta)u(T) - ph(T)\right\} \\ - \frac{x(T)^{-\sigma} ph(T)^{1-\gamma}}{\mu A\gamma}\left\{-\mu a - \mu A h(T)^\gamma + \mu B u(T)^\omega + \mu D(T)\right\} \end{aligned} \tag{13}$$

with $x(T) = x(0)\mathrm{e}^{g_c T}$, $h(T) = h(0)\mathrm{e}^{g_h T}$, $u(T) = u(0)\mathrm{e}^{g_u T}$, and $v(T) = (x(T)^{1-\sigma} - 1)/(-\sigma)$ for $\sigma \neq 1$ and $v(T) = \log(x(T)$ otherwise. Together, (5), and (9) – (13) establish 5 equations in 5 unknowns: the initial values $x(0), h(0)$, and $u(0)$, optimal education s, and the optimal age of death T.

4. Calibration

In the following calibration study we consider an average 16 year old male US American in the year 2000. The initial age is set to 16 years, corresponding to model-age zero, because individuals below roughly the age of 16 are not subject to increasing morbidity (Arking, 2006) and are presumably not well described by the law of increasing frailty. Furthermore, in many states of the US as well as in many countries around the world schooling is compulsory up to an age of about 16. This means that there is not really an *individual decision* about education below this age. An implication is that a person of model age zero has spent already 9 years on education. The effect of past education is captured by the initial endowment $\bar{w}$.

In order to calibrate the model to US data I begin with employing the Health Euler equation (7). From the data in Keehan et al. (2004) I set the growth rate of health expenditure over the life cycle g_h to 0.021. From Mitnitski et al. (2002a) I take the estimate of $\mu = 0.043$ for (Canadian) men, and finally I set $r = 0.06$ (e.g., Barro et al., 1995). This produces the estimate $\gamma = 1 - (r - \mu)/g_h = 0.19$, which squares well with the independent estimates obtained by Hall and Jones (2007).[6]

In the year 2000 the average life-expectancy of a 20 year old male US American was 75.6 years. From Mitnitski et al.'s (2002a) regression analysis I infer terminal health deficits $\bar{D} = D(75.6) = 0.1005$ and initial health deficits $D(0) = D(16) = 0.0261$. In order to get an estimate of a, I assume that before the onset of the 20th century the impact of medical technology on *adult* mortality was virtually zero. In the year 1900 the life expectancy of a 20 year old U.S. American was 62 years (Fries, 1980), implying that a 16 year old expected to live for about 46 more years. I set a such that a person who abstains from unhealthy consumption and has no access to life prolonging medical technology expects $T = 46$. From this value I get the estimate $a = 0.01427$.

The parameters entering the equation for optimal education (9) are potentially individual-specific and varying their size is the most interesting numerical experiment. In the following I estimate parameter values for the average US American. I set η to 0.05 and ψ to 0.28 according to the medium scenario in Bils and Klenow (2000). Since the benchmark American attends school for 13.5 years I adjust θ such that the marginal return to education $\theta s^{-\psi}$ equals 0.068, which

[6] 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.

is the return to schooling beyond the eighth year estimated by Psacharopoulos (1994) and since then applied by Hall and Jones (1999) and many others. This provides the estimate $\theta = 0.141$. I estimate α_1, α_2, and δ by requiring that the 16 year old reference American wants to attend school for 4.5 more years such that he acquires altogether 13.5 years of schooling (which is the US average in the year 2000 according to Turner et al., 2007) and such that the wage for age curve attains its maximum at age 55 at a value of about 1.2 times the labor income at age 30, as observed by French (2005). These three constraints provide the estimate $\alpha_1 = 0.039$, $\alpha_2 = 0.046$, and $\delta = 0.135$. Remarkably, the estimate of the decay rate of human capital α_2 is close to the value of the force of aging μ (which is 0.043).

The growth rate of the wage per unit of human capital g_w, that is aggregate productivity, is set to an annual rate of one percent based on US TFP growth in 1995-2000 (Jorgenson et al., 2008). I set $R = 48 = 64 - 16$ corresponding to the average US retirement age, and adjust the initial unit wage $\bar{w}$ such that total labor income across all working ages equals \$35,320, which is the average annual pay for workers in the year 2000 (BLS, 2011). This implies $\bar{w} = 24470$. In the basic run I set $k_0 = \bar{k} = 0$, excluding bequests and inheritances on education and health behavior.

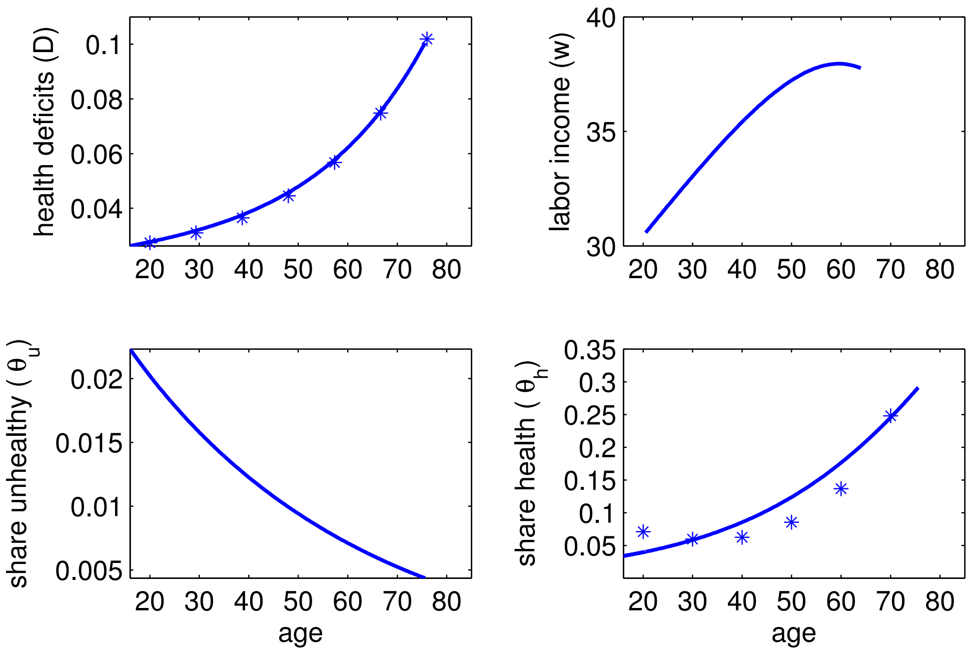
For the benchmark run I normalize the price of health p to unity and experiment with alternative values of the price of the unhealthy good. For that purpose I exploit the fact that q enters equations (5), (9), and (10)-(13) only in form of the compound $(\beta - q)$. I first calibrate $(\beta - q)$ and the remaining parameters, A, B, ρ, σ and, ω and then adjust q in order to capture a particular price elasticity of demand for the unhealthy good. The remaining parameters are estimated such that (i) the model predicts the actual accumulation of health deficits over life (as estimated by Mitnitski and Rockwood, 2002), (ii) death occurs at the moment when $\bar{D}$ health deficits have been accumulated at an age of 75.6 years, (iii) the health share of total expenditure approximates average age specific expenditure shares of American adults, and (iv) consumption of the unhealthy good costs 2.5 years of longevity.

These 2.5 lost years are explained as follows. Most of the available empirical literature on consumption of unhealthy goods is about cigarettes and tobacco. It thus seems reasonable to capture the characteristics of cigarette consumption in a benchmark case and then proceed with sensitivity analysis. Preston et al. (2010) estimate that smoking takes away 2.5 years of life-expectancy of 50 year old US males. To let the model produce this particular unhealthiness of consumption

I proceed iteratively. After running a particular specification of the model I compute how damaging the unhealthy behavior actually was by integrating its effect on deficit accumulation. From that I compute the counter-factual, that is how many extra years the individual would have lived without the unhealthy behavior. Parameters are adjusted until the model predicts a loss of 2.5 years. Given constraints (i)-(iv) I get the following estimates: $(\beta - q) = 3.9$, $B = 1.7 \cdot 10^{-7}$, $\omega = 1.4$, $A = 0.001646$, $\rho = 0.085$ and $\sigma = 1.16$. The estimate of σ fits nicely with recent empirically studies suggesting that the intertemporal elasticity of substitution is around unity (e.g. Chetty, 2006).

Finally I adjust q (and thus β). From Chaloupka and Warner (2000) we know that the price elasticity of demand for cigarettes is most likely between -0.5 and -0.3 and we alternatively consider both extreme values. For the benchmark run the price elasticity ϵ_u is set to -0.3. This leads to the estimate $q = 0.85$ and thus $\beta = 4.75$.

Figure 1: Optimal Schooling and Aging: Basic Run



Solid lines: basic run. Parameters: $a = 0.01427$, $A = 0.00165$, $\alpha_1 = 0.039$, $\alpha_2 = 0.046$, $\delta = 0.135$, $\eta = 0.05$, $\theta = 0.141$, $\bar{w} = 24470$, $g_A = 0.01$, $\mu = 0.043$, $r = 0.06$, $\rho = 0.085$, $\sigma = 1.16$, $D_0 = 0.0274$, $\bar{D} = 0.10$, $p = 1$, $\beta = 4.75$, $q = 0.85$, $B = 1.7 \cdot 10^{-7}$, $\omega = 1.40$, $k_0 = \bar{k} = 0$. Stars: data. Labor income measured in thousands of Dollars.

Figure 1 shows the implied trajectories over the life cycle of the Reference American. Stars in the health deficit panel indicate the original estimates of Mitnitski and Rockwood (2002). The model explains the actual accumulation of health deficits quite well. Stars in the lower right panel show the actual age-specific share of health expenditure. The health expenditure data is taken from Meara et al. (2004) and the data for total expenditure is taken from BLS (2002). Both are from the year 2000. The health data, however, is per person and the original expenditure data is per household. I thus converted the household data into age-specific consumption per adult by following Deaton (1997) and computing equivalence scales. Specifically I assigned household members under 18 a weight of 0.5 of adult consumption (The BLS data does not differentiate between children of different ages in the household). The model mildly underestimates health expenditure shares at young ages and overestimates it at middle ages. Altogether, however, the predicted health expenditure share fits the data reasonably well.

The lower left panel of Figure 1 shows the expenditure share of unhealthy consumption θ_u for the Reference American. When young he is predicted to spend about 2.2 percent on unhealthy goods. The expenditure share is declining until death. On average the reference American spends \$ 457 per year on unhealthy consumption, a figure that squares reasonably well with the \$ 319 that Americans spent on average for cigarettes in the year 2000 (BLS, 2002).

The upper right panel displays the calibrated invertedly-u-shaped trajectory of labor income across ages. The decline of the wage for age curve at older ages is attenuated by the fact that the wage per unit of human capital grows at 1 percent annually. As a result there is only a slight decline of labor income after age 60.

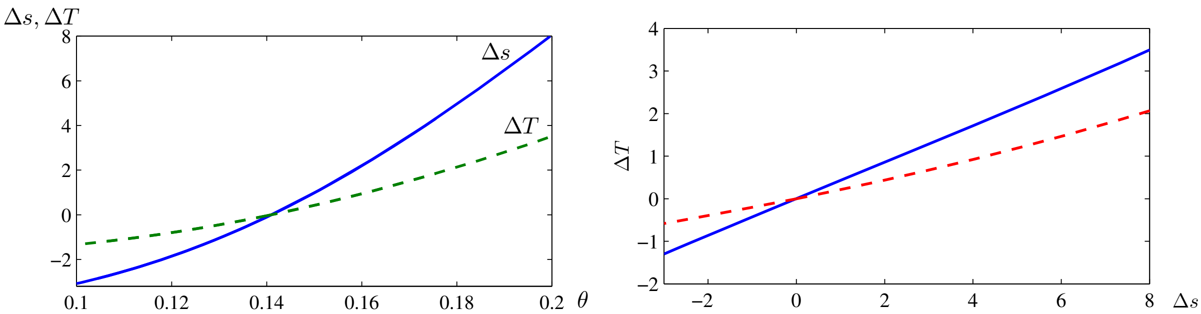
5. Results

5.1. **The Education Gradient Explained by Cognitive Ability.** The major interest of this paper is to identify the education gradient. The first experiment thus considers a person endowed with a higher return to education. The experiment is inspired by research in labor economics which acknowledges that the "return to education is not a single parameter in the population, but rather a random variable that may vary with other characteristics of individuals" (Card, 1999). Since our experiment holds constant preferences (attitudes) and time (calender year), the variation of the return to education is best explained as originating from a variation in cognitive ability. Individuals who are more or less able than the average take up more or less education (Heckman and Vytlacil, 2001). Given this notion of the return to education, the results presented below will fit nicely with the empirical evidence on cognitive ability and health behavior (Cutler and Lleras-Muney, 2010).

The first experiment is to increase θ such that the person is motivated to get one more year of education. The associated optimal changes of behavior and the resulting change of longevity are shown in the first row of Table 1. Ceteris paribus, the person educates one year longer when θ rises from 0.141 to 0.150. This motivates the person to reduce unhealthy consumption by 8.6 percent and increase health expenditure by 4.3 percent compared to the benchmark citizen. These values are calculated on the basis of average expenditure on the respective good over the life-cycle. As a consequence of the behavioral changes, the better educated person lives about half a year longer.

The result accords well with the empirical observation that education is positively associated with health through behavior as well as through income (sometimes called resources or access to health in the literature) and that both channels are about equally important (Cutler and Lleras-Muney, 2010). As explained above the model helps to identify causality. It estimates how many years in life expectancy can be explained by increasing education because the reverse causality has been shut off. Higher cognitive skills make education more worthwhile. Better educated persons earn higher income and thus aspire to enjoy consumption for a longer period of life by indulging less in unhealthy consumption and by spending more on health.

Figure 2: The Education Gradient as a Response to Varying Cognitive Ability (θ)



Left hand side: Extra schooling and caused gain in lifespan relative to basic run for alternative θ. Right hand side: Solid line: implied education gradient for the experiment from the left hand side panel; dashed line: alternative scenario: less steeply declining returns to education, $\psi = 0.1$ (instead of 0.28).

The effect of education on life-length is almost linear. If θ is reduced from 0.141 to 0.130, as shown in the second row of Table 1, the person is motivated to attain school for one year less and as a consequence of the entailed behavior he lives 0.60 years shorter. Figure 2 investigates the education gradient more generally. The panel on the left hand side shows the desired extra years of schooling and the associated gain in longevity for alternative θ, that is for alternative returns on education. When θ varies between 0.1 and 0.2 the marginal return to education, $\theta s^{-\psi}$, evaluated at 13.5 years of education varies between 0.043 and 0.115. If θ increases from 0.14 to

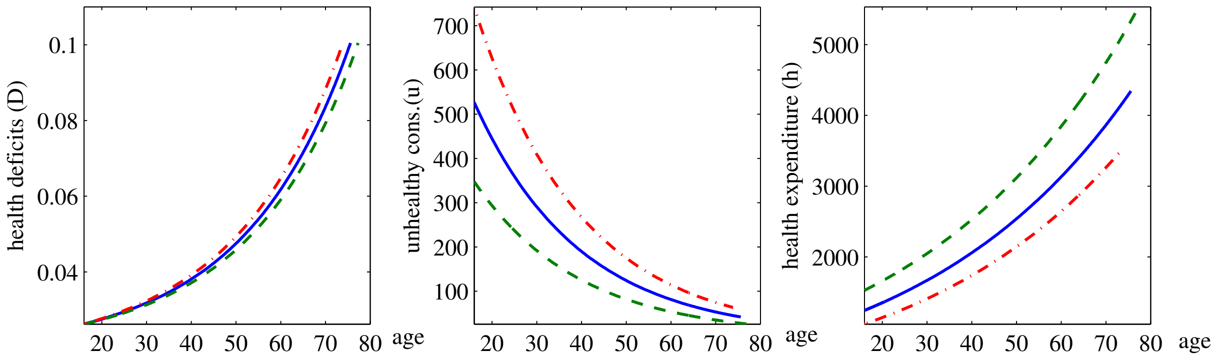
0.2 the individual obtains 8 more years of education and gains a little less than 4 more years in longevity. Compared with the observation that in 1990 a 25 year college graduate could expect to live 8 years longer than a high school dropout of the same age (Cutler and Lleras-Muney, 2010; Richards and Barry, 1998), we conclude that cognitive ability measured by θ explains about half of the education gradient, according to the benchmark model.

The panel on the right hand side of Figure 2 shows the education gradient in a Δs–ΔT diagram. The solid line reiterates the information from the panel on the left hand side. The dashed line shows sensitivity of the result with respect to decreasing returns in education. For that purpose I have set ψ to 0.1 (instead of 0.28) and re-estimated the values of α_1, α_2, δ, θ, and $\bar{w}$ such that the new scenario is comparable to the old one, meaning that the benchmark American obtains 13.5 years of education, that the marginal return to education equals 6.8 percent at this point and that the wage for age curve is the same as for the benchmark model. All other parameters are kept from the benchmark calibration. The education gradient is now less steep; 8 more years of education produce 2 more years of life expectancy. The value of θ is 0.088 for the benchmark case and it varies between 0.06 and 0.11 when education varies between -3 and 8 extra years compared to the benchmark. The implied variation of the marginal return to education is between 0.046 and 0.085. Given the lower value of ψ the marginal return to education is less steeply decreasing and there is less variation of the return to education needed in order to elicit the same variation in schooling as for the benchmark case. As a result the associated gain in longevity is smaller. In both cases, however, the education gradient is almost a straight line, which corresponds nicely with the observation that the relationship between education and health is roughly linear after 10 years of school (Cutler and Lleras-Muney, 2010).

The associated evolution of health and health behavior is presented in Figure 3. It displays optimal age-trajectories for the Reference American (solid lines), another person endowed with higher cognitive ability ($\theta = 0.173$, marginal return to schooling 8.35 percent at 13.5 years of schooling) who takes up four more years of education (dashed lines), and a third person endowed with lower cognitive ability ($\theta = 0.077$, marginal return 4.1 percent) who obtains four years less education (dash-dotted lines). The better educated persons display at any given age a better health status (less deficits). The health differences are explained by health behavior. The better educated persons spend more on health and less on unhealthy consumption at any given age.

At the aggregate level the result implies that the secular increase of the average return to

Figure 3: Education, Health, and Health Behavior



Health expenditure and unhealthy expenditure in thousands. Blue (solid) lines: basic run (from Figure 1). Green (dashed) lines: $\theta = 0.173$ (four years more education). Red (dash-dotted) lines: $\theta = 0.077$ (four years less education).

education over the last decades (Katz and Autor, 1999) may have had a causal impact on the simultaneously observed increase in life-expectancy. This suggests a mechanism linking longevity and education that reverses the causality so far proposed in the literature (Ben Porath, 1967, Cervellati and Sunde, 2005). The empirical fact that the return to education increased more for persons with high cognitive skills and for occupations requiring a lot of cognitive skills (Murnane et al., 1995) may have contributed to the disproportionate increase of longevity for the well educated.

Table 1: The Education Gradient: Alternative Mechanisms

Δs	ΔT	$\Delta u/u$	$\Delta h/h$	par. change
$+1$	$+0.56$	-8.6	$+4.3$	$\theta = 0.150$
-1	-0.60	$+9.7$	-4.2	$\theta = 0.130$
$+1$	$+0.08$	-0.3	$+0.0$	$\psi = 0.244$
-1	-0.03	$+0.0$	-0.0	$\psi = 0.338$
$+1$	$+2.6$	-31	$+21$	$\alpha_1 = 0.030$
-1	-3.0	$+43$	-17	$\alpha_1 = 0.050$
$+1$	$+1.8$	-20	$+14$	$g_w = 1.84\%$
-1	-2.0	$+28$	-11	$g_w = 0.00\%$
$+1$	$+1.1$	$+9.5$	-28	$D_0 = 0.0237, \delta = 0.027$
$+1$	$+1.0$	$+9.2$	$+1.0$	$A = 0.0215, \ \delta = 0.027$

ΔT measured in years, $\Delta u/u$ and $\Delta h/h$ measured in percent.

5.2. **Alternative Mechanisms.** Before we begin to investigate other potential drivers of the education gradient, it is worthwhile to note that two seemingly natural candidates are already excluded by theory, namely the time preference rate, ρ, and income for given education, that is $\bar{w}$.

The reason is that both parameters – while having a strong impact on life-length – leave education unaffected. To see this, reconsider equation (9) which determines optimal s and conclude that it is independent from $\bar{w}$ and ρ. The independence of the schooling decision from time-preference and the initial level of wages is not a particularity of the current approach but a standard result from the literature on optimal education (see e.g. Bils and Klenow, 2000, Card, 1999).[7]

A change of the return to education could be conceptualized as a change of θ or as a change of ψ, i.e. the curvature of the return function $f(s)$. The latter possibility is explored in the next couple of rows in Table 1. An increase of education by one year is provoked by a reduction of ψ from 0.28 to 0.244. The elicited change of health behavior and longevity however is relatively small. The reason is that the corresponding increase of income is much smaller under the ψ scenario than under the θ scenario.

The next couple of rows in Table 1 investigates the impact of age on human capital accumulation. Since variation of α_1 (in contrast to α_2 and δ) modifies the return to education and the return on experience it could be conceptualized as the impact of age on learning and cognitive abilities. A change of α_1 greatly modifies the wage for age curve and has a large impact on life-time income. As a result the change of health behavior and longevity elicited by one more year of education is quite large. One more year of education caused by declining α_1 leads to 2.6 more years of longevity. The predicted education gradient is "too steep" in the sense that there is too little variation of education needed in order to explain the full variation of longevity. A reasonable interpretation is thus that α_1 may have contributed to the education gradient on top of the variation of θ.

Besides the notion that α_1 varies at a given time across occupations, a secular decline of α_1 may have contributed to an increasing education gradient over time. The decline of farming and industrial production and the rise of the service sector has led to a decline of the importance of motor skills and a rise of the importance of crystallized abilities. In the year 2000, the Reference American is no longer farmer or industrial laborer but salesman or consultant. Since crystallized abilities decline later in life, the incentive to educate for a longer time rises, which in turn elicits more healthy behavior and increases longevity.

Another interesting macroeconomic channel, the rate of productivity growth, is investigated in

[7] Notice that actually there may well be a positive influence of time preference on schooling or vice versa from schooling on time preference and other non-cognitive traits (Fuchs, 1982; Conti, Heckman and Urzua, 2010; Kaestner and Callison, 2011). The absence of this channel in the present model helps to identify how much of the education gradient can be explained by cognitive ability.

rows 6 and 7 in Table 1. Economic growth devalues the costs of not working today and increases the benefit of working tomorrow. At a higher rate of economic growth it becomes less costly to delay entry into work-life in favor for additional education and the person educates more. Inspecting (9) we see also that, once individuals are working, economic growth operates formally like experience on the job. It makes human capital more worthwhile. Consequently individuals spend more on health and indulge less in unhealthy behavior. An increase of productivity growth from 1.0 to 1.84 percent triggers one year more of education and behavioral changes that enables the person to live about 1.8 years longer.

On the aggregate level the result is interesting since it motivates a causal effect from economic growth to education and longevity while the so far established literature has focussed on causality in the oppositive direction. The problem is, however, that aggregate productivity growth varies within a limited range, which makes it impossible to motivate a large and further increasing education gradient. In order to exploit productivity growth to rationalize the education gradient it seems more reasonable to consider occupation-specific growth of productivity. In this context, the model predicts that people are more motivated to educate longer in order to get an occupation in a high growth sector of the economy.

The other parameters of the model are, taken for themselves, not capable to motivate the gradient. Seemingly, however, the gradient could be motivated by plausible *combinations* of parameter changes. The next rows in Table 1 consider two such cases. Variation of initial health deficits D_0 have, naturally, a strong impact on longevity but, by construction, no impact on the wage for age curve and education. One could, however, argue that δ, conceptualized as the loss of human due to initial health deficits, varies with D_0. The penultimate row in Table 1 shows a combined reduction of D_0 and δ that produces an education gradient with a slope close to unity. The problem with this experiment is that it implies counterfactual health behavior. Motivated by the improved better initial health, better educated individuals are predicted to spend less on health and to indulge more unhealthy consumption.

The final row in Table 1 shows a related case. One could imagine that technological progress leads to improvements in medical productivity A as well as to slower deterioration of labor productivity, i.e. lower δ. The reason is that better medical technology improves repair and maintenance of human capital. The final row shows a joint increase of A and decrease of δ that generates an education gradient with a slope of unity. Yet, the experiment produces again counterfactual health behavior: later decline and better repair of human capital motivates the individual to indulge more in unhealthy consumption. Higher education is counterfactually predicted to be associated with deteriorating health behavior.

5.3. **Robustness of Results.** Table 2 displays some robustness checks. It focusses on variation of θ, which appears to be the most convincing driver of the education gradient. In any experiment θ is raised from 0.141 to 0.150, i.e. the return to education at 13.5 years of education is raised from 6.8 to 7.2 percent. The first two rows, case 1 and 2, document that the education gradient operates independently from (education unrelated) income. Results are shown when, ceteris paribus, the annual wage per unit of human capital is assumed to be 50 percent higher or lower. These income variations have very strong effects on longevity, which rises by 4 years or, respectively, falls by 3 years compared to the basic run. The gradient, however, that is the gain in longevity that is associated with one more year of education, is almost the same as for the benchmark run.

Table 2: The Education Gradient: Robustness Checks

Case	Δs	ΔT	$\Delta u/u$	$\Delta h/h$	comment
1) $\Delta\bar{w} = +50\%$	$+1$	$+0.53$	-10	$+5.0$	a richer individual
2) $\Delta\bar{w} = -50\%$	$+1$	$+0.50$	-5.8	$+2.7$	a poorer one
3) $k_0 = 6\bar{w}$	$+1$	$+0.43$	-8.1	$+3.9$	inheritance (parental support)
4) $k_0 = 6\bar{w}$, $\bar{k} = 6\bar{w}$	$+1$	$+0.42$	-8.1	$+3.9$	inheritance and bequest
5) $\epsilon_q = -0.5$	$+1$	$+0.56$	-8.6	$+4.3$	higher price elast. of unhealthy good
6) $B = 10^{-6}$, $\omega = 1.16$	$+1$	$+0.69$	-16	$+3.4$	unit consumption more unhealthy
7) $B = 10^{-8}$, $\omega = 1.76$	$+1$	$+0.48$	-4.9	$+4.6$	unit consumption less unhealthy
8) $p = 2$, $A = 0.00187$	$+1$	$+0.56$	-8.6	$+4.2$	higher price of health
9) $p = 2$, $\bar{w} = 31400$	$+1$	$+0.42$	-11	$+5.5$	higher price of health

In all cases the experiment increases θ from 0.14 to 0.15. In order to match the data, re-calibration for case 5: $q = 1.4$, $\beta = 5.3$. For case 8: $A = 0.00142$, $\alpha = 1.059$. For Case 9: $A = 0.00139$, $\alpha = 1.061$. All other parameters from benchmark case (Figure 1). See text for details.

During the education period the Reference American accumulates debt of about 170 k. This relatively high accumulation of debt in young ages is an unreasonable artefact originating from the assumption that the Reference American does not benefit from supporting parents. In order to accommodate this criticism, case 3 endows the person with an inheritance of $6\bar{w}$, which is about four times the average annual labor income, a sum which can be regarded as sufficient to finance 4 or 5 years of voluntary education. The inheritance reduces the gradient by about one third to 0.42. The reason is that financial wealth, ceteris paribus, reduces the incentive to educate. The wealthy person prefers to finance larger parts of consumption and health expenditure in old age by returns on capital rather than human capital and savings from labor income. Case 4 requires, additionally, that the person leaves a bequest of $6\bar{w}$. It demonstrates that the mechanism runs mainly through the inheritance received rather then through the bequest left.

Case 5 in Table 2 adjusts prices and preferences such that the price elasticity of demand for the unhealthy good equals -0.5, an elasticity observed at the upper end of estimates from cigarette demand. The experiment keeps the difference $(\beta - q)$ from the basic model such that initial education and longevity are preserved. The result verifies the claim that it is indeed the preference-price differential rather than absolute values, which are driving the education gradient and health behavior.

With case 6 and 7 we investigate the character of the unhealthy good. Case 6 assumes that the good is much more unhealthy than cigarettes. B is raised by about factor 10. The scale parameter ω is adjusted to 1.16 such that the calibration continues to produce education, expenditure, and longevity from the basic run. The greater unhealthiness of consumption of small quantities of the good implies a higher incentive for the better educated to stay away from this good. The experiment consequently predicts a somewhat larger education gradient.

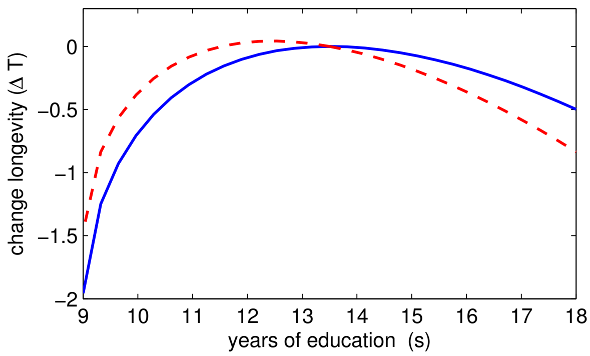
Inverting the result from above, the model produces a smaller gradient if u is generally less unhealthy. This is confirmed by case 7 in which B has been reduced by about factor 10 and the scale parameter has been increased from 1.4 to 1.77 in order to match the data. The new parameter values reflect a good for which consumption of small quantities entails relatively small effects on health whereas large excess consumption has severe consequences. It could perhaps be thought of as alcohol. The predicted education gradient is 0.43 and thus a bit smaller than for the basic run (representing cigarettes). The reason is that better educated individuals have less incentive to stay away from consuming the good.

Finally we consider the impact of the price of health. Setting $p = 2$ implies that the cost of health provisions doubles compared to the benchmark case. In order to fit the Reference American at least one parameter has to be re-calibrated such that life-span of the benchmark case is restored. Two natural candidates for that are the productivity of health expenditure A and the level of income $\bar{w}$. Case 8 considers the adjustment via higher A. In this case the predicted change of health and health behavior is virtually the same as for the benchmark run. The outcome is intuitive because the individual is compensated for the higher price of health by higher efficiency of health expenditure. The re-calibration via higher income is shown in case 9.

Now the health gradient is smaller. The individual spends somewhat more on health but chiefly responds to the higher price of health by reducing unhealthy consumption.

5.4. **Endogenous vs. Exogenous Education.** In this section I re-investigate the problem under the condition that education is not individually chosen but exogenously given. I show that the education gradient under exogenous education is relatively small after individuals achieved basic education and that it turns negative eventually. Endogenous education, driven by cognitive ability, appears to be the more powerful explanation of the contemporary education gradient.

Figure 4: Longevity Gain: Exogenous Education



The figure shows the change of longevity compared to the benchmark (75.6 years) when the individual is forced to spend s years on education. Solid line benchmark calibration. Dashed line: $\theta = 0.13$ (instead of 0.14).

The solid line in Figure 4 shows results for the Reference American when problem (1) – (3) is solved under the constraint that he has to spend s years on education. For a better understanding recall that the calibration of the benchmark model implied that the unconstrained Reference American takes up voluntarily 13.5 years of education (at a θ of 0.14 and a return to education of 6.8 percent). The model predicts that the Reference American would live 2 years less if he would be forced to take up only 9 years of education. At this low level of education the health gradient is relatively steeply increasing. Yet, naturally a maximum is reached at 13.5 years of education. If enforced education is further increasing, the health gradient turns negative. The reason is that life-time income assumes a maximum at 13.5 years of education. Extending the education period further reduces life-time income (although it increases human capital) such that the individual spends less on health provisions. The predicted decrease is relatively mild but the important point is that with already high education of the Reference American it gets harder and becomes eventually impossible to generate a positive education gradient in a model based on exogenous education.

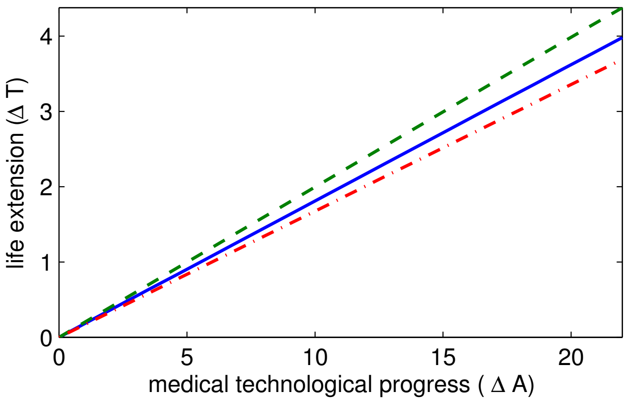
The red (dashed) line shows the result when θ is kept at 0.13 (instead of 0.14). In this case the education gradient starts declining already after 12 years of education. Forcing this person to take up more than high school would reduce life time income and longevity. These numerical experiments rationalize why the enforced education gradient is large only when average education is low. In this case it is likely that extending schooling is optimal from the individual's perspective. At a low level of education, school attendance is for most children not driven by cognitive ability but other motives, like distance to school, child labor, credit constraints, or leisure preferences. The model thus supports the empirical finding of a large health effect from compulsory schooling at low levels of schooling (e.g. Lleras-Muney, 2005). When the average level of education is already high, however, it is hard to motivate a further rising education gradient by exogenous education.

These results rationalize also why studies of monozygotic twins find relatively small health gains from education (Fujiwara and Kawachi, 2009; Lundberg, 2013). The reason is that monozygotic twins are likely to be endowed with similar cognitive ability, which trumps the length of the education period in the determination of human capital and health behavior. As shown in Figure 4, there is little variation in predicted longevity for persons sharing same θ of 0.14 who attend school between 10 and 18 years. In contrast, if schooling is motivated by cognitive ability and endogenously chosen, the benchmark model (Figure 2) predicts that eight more years of education lead to a longevity gain of four more years. These results are in line with an observed strong empirical association between cognitive ability and health and longevity (Deary, 2008; Der et al., 2009; Calvin et al., 2011).

5.5. **Medical Technological Progress and the Health Gradient.** In the final experiment I investigate the impact of medical technological progress on life-time extension and in particular on the education gradient. It has been hypothesized that the observable secular increase of the education gradient may have its origin in technological progress because better educated persons have better access and more resources to utilize technological advances to their benefit (Cutler et al., 2010). Better educated persons are, ceteris paribus, richer (if education is driven by cognitive ability). Since the income elasticity of health spending is larger than unity (Hall and Jones, 2007), better educated persons demand more health services in order to protect or repair their human capital. We thus expect that they benefit to a larger degree from medical technological progress.

The results presented in Figure 5 confirm this expectation. The figure shows longevity gains originating from increasing efficiency of health expenditure A; ΔA is measured in percent of the

Figure 5: Medical Technological Progress and the Health Gradient



The Figure shows the gain in longevity for alternative progress of medical technology (ΔA). Blue (solid): benchmark run (13.5 years of education, $\theta = 0.1$). Green (dashed): 4 years more of education ($\theta = 0.12$). Red (dash-dotted): 7.2 years more of education ($\theta = 0.14$) The longevity gain is measured relative to the own initial life-span for both types.

benchmark run. If medical technology advances at an annual rate of 1 percent the level of A is 20 percent higher after about 18 years. The solid line shows the predicted longevity for the Reference American (endowed with a θ of 0.14, i.e. a return of education of 6.8 percent at 13.5 years of education). The dashed line shows implied longevity for a person endowed with $\theta = 0.16$ who attends school for 2 more years and experiences a return to education of 7.4 percent. The dash-dotted line reflects the longevity gain of a person with $\theta = 0.12$ who attends school for 1.8 years less, at a rate of return of 6.0 percent. Although everybody experiences an increase in longevity, the predicted gain is higher for better educated persons is higher. When A advances by 20 percent the longevity gap between a high school graduate and a college graduate widens by about one year.

6. Conclusion

This study has proposed a new view on the education gradient. It has assumed away any explanation based on attitudes, non-cognitive skills, and allocative or productive inefficiency of the uneducated. Instead it has asked how large a gradient can be motivated by optimal decisions on education and health behavior of individuals who know how their behavior affect their future health status and the time of death. The theory has been firmly built on insights from modern gerontology which allowed a reliable calibration for a "Reference American". It predicts that a person whose return to education (cognitive skills) motivate one year more of education, spends more on health and less on unhealthy behavior such that he or she lives about half a year longer.

This means that the theory explains about half of the observable education gradient. The theory motivates an almost linear education gradient, in line with the empirical observation (Cutler and Lleras-Muney, 2010).

The theory suggests that idiosyncratic differences in cognitive ability and endogenous choice of education are a more powerful explanation of the gradient than enforced exogenous education. It predicts, in line with the empirical observation from monozygotic twins (Fujiwara and Kawachi, 2009; Lundberg, 2013), that for similar cognitive ability the length of the education period contributes little to the explanation of longevity. The theory has been utilized to explain why the education gradient gets larger over time through ongoing medical technological progress. The reason is that well educated persons demand relatively more health services and thus benefit to a larger degree from health innovations.

The study has focussed on the human life cycle from young adulthood onwards. At this age, taking cognitive skills as approximately given is an appropriate simplification. But since cognitive skills seem to be malleable at younger ages (see e.g. Heckman, 2006), the present study also highlights the importance of childhood development for later life. Equipped with a high return to education, individuals are not only predicted to educate longer and earn more labor income but also to lead a healthier life and to live longer.

References

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

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

Ben-Porath, Y., 1967, The production of human capital and the life cycle of earnings, *Journal of Political Economy* 75, 352–365.

Bils, M. and Klenow, P.J., 2000, Does schooling cause growth?,*American Economic Review* 90, 1160–1183.

BLS, 2002, Consumer Expenditures in 2000, United States Department of Labor, Bureau of Labor Statistics, Report 958.

BLS, 2011, State average annual pay for 2000 and 2001, United States Department of Labor, Bureau of Labor Statistics (`http://www.bls.gov/news.release/annpay.t01.htm`).

Calvin, C. M., Deary, I. J., Fenton, C., Roberts, B. A., Der, G., Leckenby, N., and Batty, G. D., 2011, Intelligence in youth and all-cause-mortality: systematic review with meta-analysis, *International Journal of Epidemiology* 40, 626-644.

Card, D., 1999, The causal effect of education on earnings, *Handbook of Labor Economics* Vol. 3, 1801–1863, Elsevier, Amsterdam.

Case, A. and Deaton, A., 2005, Broken down by work and sex: how our health declines, in: David Wise (ed.), Analyses in the Economics of Aging, University of Chicago Press, Chicago.

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

Chaloupka, F.J. and Warner, K.E., 2000, The economics of smoking, *Handbook of Health Economics* Vol. 1, Elsevier, Amsterdam, 1539–1627.

Harttgen, K., Chatterji, S., Kowal, P., Strulik, H., Vollmer, S., 2013, Patterns in disability and frailty in older adults: Results from the Survey of Health, Ageing and Retirement in Europe (SHARE), *PlosOne*, forthcoming.

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

Conti, G. and Heckman, J. and Urzua, S., 2010, The education-health gradient, *American Economic Review* 100, 234–38.

Contoyannis, P. and Jones, A.M., 2004, Socio-economic status, health and lifestyle, *Journal of Health Economics* 23, 965–995.

Cutler, D.M. and Lleras-Muney, A., 2006, Education and health: Evaluating theories and evidence, NBER Working Paper 12352.

Cutler, D.M., Lleras-Muney, A. and Vogl, T., 2008, Socioeconomic status and health: dimensions and mechanisms, NBER Working Paper 14333.

Cutler, D.M., Lleras-Muney, A. and Vogl, T., 2011, Socioeconomic status and health: dimensions and mechanisms, in: Glied, S. and Smith, P.C. (eds), *The Oxford Handbook of Health Economics*, Oxford University Press, Oxford, pp. 124-163.

Cutler, D.M. and Lleras-Muney, A., 2010, Understanding differences in health behaviors by education, *Journal of Health Economics* 29, 1–28.

Cutler, D.M., Lange, F., Meara, E., Richards, S., and Ruhm, C.J., 2010, Explaining the rise in educational gradients in mortality, NBER Working Paper 15678.

Dalgaard, C.J. and Strulik, H., 2013, Optimal aging and death: Understanding the Preston curve, *Journal of the European Economic Association*, forthcoming.

Deary, I.L., 2008, Why do intelligent people live longer? *Nature* 456, 175-176.

Deaton, A., 1997, *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, Johns Hopkins University Press.

Der, G., Batty, G.D. and Deary, I.J., The association between IQ in adolescence and a range of health outcomes at 40 in the 1979 US National Longitudinal Study of Youth, *Intelligence* 37, 573-580.

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

Elo, I.T. and Preston, S.H., 1996, Educational differentials in mortality: United States, 1979-1985, *Social Science & Medicine* 42, 47–57.

French, E., 2005, The effects of health, wealth, and wages on labour supply and retirement behaviour, *Review of Economic Studies*72, 395-427.

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

Fujiwara, T. and Kawachi, I., 2009, Is education causally related to better health? A twin fixed-effect study in the USA, *International Journal of Epidemiology* 38, 1310-1322.

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

Glied, S. and Lleras-Muney, A., 2008, Technological innovation and inequality in health, *Demography* 45, 741–761.

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

Grossman, M., 2006, Education and nonmarket outcomes, in: *Handbook of the Economics of Education* Vol. 1, 577–633, Elsevier, Amsterdam.

Hall, R.E. and Jones, C.I., 1999, Why do some countries produce so much more output per worker than others? *Quarterly Journal of Economics* 114, 83-116.

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

Heckman, J.J. and Vytlacil, E., 2001, Identifying the role of cognitive ability in explaining the level of and change in the return of schooling, *Review of Economics and Statistics* 83, 1–12.

Heckman, J.J., 2006, Skill formation and the economics of investing in disadvantaged children, *Science* 312, 1900–1902.

Jorgenson, D.W., Ho, M.S. and Stiroh, K.J., 2008, A retrospective look at the US productivity growth resurgence, *Journal of Economic Perspectives* 22 (1), 3–24.

Kaestner, R. and Callison, K., 2011,Adolescent Cognitive and Noncognitive Correlates of Adult Health, *Journal of Human Capital* 5, 29–69.

Katz, L.F. and David, H., 1999, Changes in the wage structure and earnings inequality, *Handbook of Labor Economics* Vol. 3, 1463–1555.

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.

Kenkel, D.S., 1991, Health behavior, health knowledge, and schooling, *Journal of Political Economy* 99, 287–305.

Lleras-Muney, A., 2005, The relationship between education and adult mortality in the United States, *Review of Economic Studies* 72, 189-221.

Lundberg, P., The health returns to schooling – what can we learn from twins, *Journal of Population Economics* 26, 673-701.

Mackenbach, J.P., Stirbu, I., Roskam, A.J.R., Schaap, M.M., Menvielle, G., Leinsalu, M., and Kunst, A.E., 2008, Socioeconomic inequalities in health in 22 European countries, *New England Journal of Medicine* 358, 2468–248.

McFadden, D., 2005, Comment, in: David Wise (ed.), Analyses in the Economics of Aging, University of Chicago Press, Chicago.

Meara, E., White, C., and Cutler, D.M., 2004, Trends in medical spending by age, 1963–2000, *Health Affairs* 23, 176–183.

Meara, E.R., Richards, S., and Cutler, D.M., 2008, The gap gets bigger: changes in mortality and life expectancy, by education, 1981–2000, *Health Affairs* 27, 350–360.

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.

Mitnitski, A. and Bao, L. and Rockwood, K., 2006, Going from bad to worse: a stochastic model of transitions in deficit accumulation, in relation to mortality, *Mechanisms of Ageing and Development* 127, 490–493.

Murnane, R.J. and Willett, J.B. and Levy, F., 1995, The growing importance of cognitive skills in wage determination, *Review of Economics and Statistics* 77, 251–266.

Murphy, K. M. and Welch, F., 1990, Empirical age-earnings profiles. *Journal of Labor Economics* 8, 202-229

Nair, K. S., 2005, Aging muscle, *American Journal of Clinical Nutrition* 81, 953-963.

Preston, S.H., D.A. Glei, and J.R. Wilmoth, 2010, Contribution of smoking to international differences in life expectancy, in: E.M. Crimmins, S.H. Preston, and B. Cohen (eds.), *International Differences in Mortality at Older Ages: Dimensions and Sources*, Washington, D.C.: National Academies Press, pp. 105-131.

Psacharopoulos, G., 1994, Returns to investment in education: A global update, *World Development* 22, 13251343.

Richards, H., Barry, R., 1998. U.S. life tables for 1990 by sex, race, and education, *Journal of Forensic Economics* 11, 9–26.

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

Schaie, K.W., 1994, The course of adult intellectual development, *American Psychologist* 49, 304-313.

Skirbekk, V., 2004, Age and individual productivity: A literature survey, *Vienna Yearbook of Population Research*, 133–154.

Strulik, H., 2012, Optimal aging with uncertain death, University of Goettingen Cege Discussion Paper 160.

Turner, C., Tamura, R., Muholland, S., and Baier, S., 2007, Education and income of the states of the United States: 1840-2000, *Journal of Economic Growth* 12, 101-158.

Zellner, A., 1969, On the aggregation problem: A new approach to a troublesome problem. In: M. Beckmann and H.P. Künzi (eds.) : *Economic Models, Estimation and Risk Programming: Essays in honor of Gerhard Tintner*, Springer Verlag, Berlin.

6.1. **Setup of the Problem.** Integrating (3) provides the following solution.

$$D(t) = D(0)\exp(\mu t) - \int_0^t \mu a \exp(\mu(t-v))\mathrm{d}v - \mu A \int_0^t h(v)^\gamma \exp(\mu(t-v))\mathrm{d}v$$
$$+ \mu B \int_0^t u(v)^\omega \exp(\mu(t-v))\mathrm{d}v. \tag{A.1}$$

Integrating (2) and using $x \equiv c + \beta u$ we get (A.2)

$$k(t) = k(0)\exp(rt) + \int_s^R \exp(r(t-v))w(v)H(s,v)\mathrm{d}v - \int_0^t \exp(r(t-v))x(v)\mathrm{d}v$$
$$- \int_0^t \exp(r(t-v))(q-\beta)u(v)\mathrm{d}v - \int_0^t \exp(r(t-v))ph(v)\mathrm{d}v. \tag{A.2}$$

Using (A.1) and (A.2), the initial conditions $D(0) = D_0$, $k(0) = k_0$, and the terminal conditions $D(T) = \bar{D}$, $k(T) = \bar{k}$, the Lagrangian associated with problem (1)- (4) is given by

$$\max_{d,h,s,T} L = \int_0^T \mathrm{e}^{-\rho t}\frac{x^{1-\sigma}}{1-\sigma}\ \mathrm{d}t$$
$$+ \phi\left\{k_0 + \mathrm{e}^{-rs}\int_s^R \mathrm{e}^{-rt}w(t)H(s,t)\mathrm{d}t - \int_0^T \mathrm{e}^{-rt}x(t)\mathrm{d}t - \int_0^T \mathrm{e}^{-rt}ph(t)\mathrm{d}t - \int_0^T \mathrm{e}^{-rt}(q-\beta)u(t)\mathrm{d}t - \bar{k}\mathrm{e}^{-rT}\right\}$$
$$+ \lambda\left\{D_0 - \mu a\int_0^T \mathrm{e}^{-\mu t}\mathrm{d}t - \mu A\int_0^T h(t)^\gamma \mathrm{e}^{-\mu t}\mathrm{d}t + \mu B\int_0^T u(t)^\omega \mathrm{e}^{-\mu t}\mathrm{d}t - \bar{D}\mathrm{e}^{-\mu T}\right\}. \tag{A.3}$$

Using (4), $w(t)H(s,t)$ in (A.3)is determined as (A.4).

$$w(t)H(t,s) = \bar{w}\exp(g_w t)\left\{\exp\left[\theta\frac{s^{1-\psi}}{1-\psi} + \eta(t-s) - \alpha_1 t\right] - \delta\exp(\alpha_2 t)\right\}. \tag{A.4}$$

Solution. The first order conditions for consumption, health expenditure, and unhealthy consumption are:

$$0 = \mathrm{e}^{-\rho t}d^{-\sigma} - \phi\mathrm{e}^{-rt} \tag{A.5}$$
$$0 = -\phi\mathrm{e}^{-rt}p - \lambda\mu A\gamma h^{\gamma-1}e^{-\mu t}. \tag{A.6}$$
$$0 = -\phi\mathrm{e}^{-rt}(q-\beta) + \lambda\mu B\omega u(t)^{\omega-1}\mathrm{e}^{-\mu t}. \tag{A.7}$$

Differentiating (A.5) with respect to time we get (6), differentiating (A.6) with respect to time we get (7), and differentiating (A.7) with respect to time we get (8). Next, solving (A.6) for λ and using the result to substitute λ in (A.7) provides (5) in the text.

The first order condition for optimal is schooling is $\partial L/\partial s = 0$, that is

$$0 = \int_0^R \frac{\partial}{\partial s}\mathrm{e}^{-rt}w(t)H(s,t)\mathrm{d}t - \mathrm{e}^{-rt}w(s)H(s,s),$$

requiring that the gain from a marginal extension of education, the first term on the right hand side, equals the income lose from a marginal extension of education, the second term. Inserting $\partial H(s,t)/\partial s = (\theta\frac{s^{1-\psi}}{1-\psi} - \eta)\exp(\theta\frac{s^{1-\psi}}{1-\psi} + \eta(t-s) - \alpha_1 t)$, $w(t) = w(s)\exp(g_w(t-s))$, and $H(s,s) =$

$\exp(\theta \frac{s^{1-\psi}}{1-\psi} - \alpha_1 s) - \delta \exp(\alpha_2 s)$, the optimal schooling condition becomes

$$(\theta s^{-\psi} - \eta) \exp\left[\theta \frac{s^{1-\psi}}{1-\psi} + (r - \eta - g_w)s\right] \int_s^R \mathrm{e}^{(g_w - r + \eta - \alpha_1)t} \mathrm{d}t = \exp\left[\theta \frac{s^{1-\psi}}{1-\psi} - \alpha_1 s\right] - \delta \mathrm{e}^{\alpha_2 s}.$$

Solving the integral provides (9) in the text.

Two conditions have to be fulfilled at the optimal T. The first one is that $D(T) = \bar{D}$. Evaluating (A.1) at T and employing the fact of constant growth rates of h and u according to (7) and (8) this can be expressed as:

$$\bar{D} = D_0 \exp(\mu T) - \mu a \int_0^T \exp(\mu(T-t)\mathrm{d}t - \mu A \int_0^T h(0)^\gamma \exp(\gamma g_h t) \exp(\mu(T-t))\mathrm{d}t$$
$$+ \mu B \int_0^T u(0)^\omega \exp(\omega g_u t) \exp(\mu(T-t))\mathrm{d}t.$$

Solving the integrals provides (12) in the text.

The second condition for optimal death is that the Lagrangian evaluated at T assumes the value of zero, that is, using (A.3) and the Euler equations (5)-(7):

$$0 = \left(\frac{x(T)^{1-\sigma} - 1}{1-\sigma}\right) \exp(\rho T) - \xi \bar{D}^\nu \exp(\rho T)$$
$$+ \phi\left[- \exp(-rT)x(T) - (q-\beta)\exp(-rT)u(T) - p\exp(-rT)h(T) + r\exp(-rT)\bar{k}\right]$$
$$+ \lambda\left[-\mu \exp(-\mu T) - \mu A h(T)^\gamma \exp(-\mu T) + \mu B u(T)^\omega \exp(-\mu T) + \mu \bar{D} \exp(-\mu T)\right]$$

Inserting from (A.5)–(A.7) that $\phi \exp(-rT) = x(T)^{-\sigma} \exp(-\rho T)$, that $\lambda \exp(-\mu T) = -\phi \exp(-rT)$ $\cdot ph(t)^{1-\gamma}/(\mu A \gamma)$, and that $\lambda \exp(-\mu T) = \phi \exp(-rT)(q-\beta)u(t)^{1-\omega}/(\mu B \omega)$ provides (13) in the text.

Using the Euler conditions (A.5)–(A.7) the budget constraint (A.2) can be written as:

$$0 = k(0) + w(s,R) - \int_0^T d(0)\exp((g_d - r)t) - p\int_0^T h(0)\exp((g_h - r)t)$$
$$- (q-\beta)\int_0^T u(0)\exp((g_u - r)t) - \bar{k}\exp(-rT).$$

Solving the integrals provides (10) in the text. Finally human wealth $w(s,R)$ is obtained as

$$\int_s^R \exp(-rt)w(t)H(s,t)\mathrm{d}t = \bar{w}\exp\left[\theta\frac{s^{1-\psi}}{1-\psi} - \eta s\right]\int_s^R \mathrm{e}^{(\eta + g_w - r - \alpha_1)t}\mathrm{d}t - \bar{w}\delta\int_s^R \mathrm{e}^{\alpha + g_w - r)t}\mathrm{d}t.$$

Solving the integrals provides the final building block for (11) in the text.

Hamilton Approach. Alternatively, the model can be solved using optimal control theory. The current value Hamiltonian associated with problem (1)-(4) is given by

$$\mathcal{J} = \frac{x^{1-\sigma}}{1-\sigma} + \lambda_D \mu\left[D - a - Ah^\gamma - Bu^\omega\right] + \lambda_k\left[\chi w H + rk - ph - d - (q-\beta)u\right]$$

The first order conditions and costate equations are:

$$0 = x^{-\sigma} - \lambda_k \tag{A.8}$$

$$0 = -\lambda_D B\omega u^{\omega-1} - \lambda_k(q-\beta) \tag{A.9}$$

$$0 = -\lambda_D \mu A \gamma h^{\gamma-1} - \lambda_k p \tag{A.10}$$

$$\lambda_D \mu = \lambda_D \rho - \dot{\lambda}_D \tag{A.11}$$

$$\lambda_k r = \lambda_k \rho - \dot{\lambda}_k. \tag{A.12}$$

From (A.9) and (A.10) follows (5) in the text. Differentiating (A.8) with respect to age and using (A.12) provides (6) in the text. Differentiating (A.10) with respect to time and using (A.11) and (A.12) provides (7) in the text. Differentiating (A.9) with respect to time and using (A.11) and (A.12) provides (8) in the text.