Human Capital and Economic Opportunity
Global Working Group

Working Paper Series

Working Paper No. 2015-008

A Contribution to Health-Capital Theory

Titus Galama

May, 2015

Human Capital and Economic Opportunity Global Working Group
Economics Research Center
University of Chicago
1126 E. 59th Street
Chicago IL 60637
www.hceconomics.org

A contribution to health-capital theory*

Titus J. Galama[†]

First version: January 19, 2011
This version: April 23, 2015

Abstract

The Grossman model is the canonical theory of the demand for health and health investment. This paper provides strong support for the model's canonical status. Yet several authors have identified at least four significant limitations to the literature spawned by Grossman's seminal 1972 papers. I show that these criticisms are not the result of a flawed model but of an unfortunate and unnecessary choice for the functional form (linear in investment) of the health-production process, and of an incorrect interpretation of the equilibrium condition for health. I find that a generalized Grossman model, with decreasing returns in investment and endogenous longevity, addresses the limitations, and provides a remarkably successful foundation for understanding decisions regarding health.

Keywords: socioeconomic status, health, demand for health, human capital, health behavior

JEL Codes : D91, I10, I12, I14, J24

*Research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health under Award Numbers K02AG042452, R01AG037398, R01AG030824, P30AG012815 and P01AG022481. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health. I am grateful to the School of Economics of Erasmus University Rotterdam for a Visiting Professorship in the Economics of Human Capital, and to Isaac Ehrlich, Michael Grossman, Arie Kapteyn, Erik Meijer, Peter Savelyev, Eddy van Doorslaer, Hans van Kippersluis, and Arthur van Soest for extensive discussions and detailed comments.

[†]University of Southern California, Dornsife College Center for Economic and Social Research (CESR), USA, galama@usc.edu.

1 Introduction

The demand for health is one of the most central topics in Health Economics. The canonical model of the demand for health arises from Grossman (1972a, 1972b, 2000) and theoretical extensions and competing economic models are still relatively few. In Grossman's human-capital framework individuals invest in health (e.g., medical care, exercise) for the consumption benefits (health provides utility) as well as production benefits (healthy individuals have greater earnings) that good health provides. The model provides a conceptual framework for interpretation of the demand for health and health investment in relation to an individual's resource constraints, preferences, and consumption needs over the life cycle. Arguably the model has been one of the most important contributions of economics to the study of health behavior. It has provided insight into a variety of phenomena related to health, health investment, inequality in health, the relationship between health and socioeconomic status, occupational choice, etc. (e.g., Cropper, 1977; Muurinen and Le Grand, 1985; Case and Deaton, 2005) and has become the standard (textbook) framework for the economics of the derived demand for medical care.

Yet several authors have identified at least four significant limitations to the literature spawned by Grossman's seminal 1972 papers[1] (see Grossman, 2000, for a review and rebuttal of some of these limitations). A standard framework for the demand for health, health investment, and longevity, has to meet the significant challenge of providing insight into a variety of complex phenomena. Ideally it would explain the significant differences observed in health between socioeconomic status (SES) groups - often called the "SES-health gradient". In the United States, a 60-year-old top-income-quartile male reports to be in similar health as a 20-year-old bottom-income-quartile male (Case and Deaton 2005) and similar patterns hold for other measures of SES, such as education and wealth, and other indicators of health, such as disability and mortality (e.g., Cutler, Lleras-Muney and Vogl, 2011; van Doorslaer et al. 2008). Initially diverging, the disparity in health between low- and high-SES groups appears to narrow after ages 50-60. Yet, Case and Deaton (2005) have argued that health-capital models are unable to explain differences between SES groups in the rate at which health deteriorates.

Another stylized fact of the demand for medical care is that healthy individuals do not go to the doctor much: a strong negative correlation is observed between measures of health and measures of health investment. However, Wagstaff (1986a) and Zweifel and Breyer (1997) have pointed to the inability of health-capital models to predict the observed negative relation between health and the demand for medical care.

Introspection and casual observation further suggests that healthy individuals are those that began life healthy and that have invested in health over the life course. Thus one would expect that health depends on initial conditions (e.g., initial health) and the history of health investments, prices, wages, medical technology, and environmental conditions.

[1] Throughout this paper I refer to this literature as the health-capital literature.

Yet, Usher (1975) has pointed to the lack of "memory" in model solutions, and Wagstaff (1993) has argued that Grossman's empirical formulation fails to capture the dynamic character of the model. For example, the solution for health does not depend on its initial value or the histories of health investment and biological aging (see, e.g., equations 42, 45 and 47 in Grossman, 2000). Thus the model would, for example, not be able to reproduce the observation that endowed health at birth and investments in early childhood have sustained effects on adult outcomes (e.g., Heckman, 2007; Currie and Almond, 2011; Campbell et al., 2014). Clearly, the static nature of the derived solution for health is incompatible with the inherently dynamic nature of health formation.

Further, Case and Deaton (2005) note that *"...If the rate of biological deterioration is constant, which is perhaps implausible but hardly impossible, ...people will "choose" an infinite life ..."* Hence, a feature of the model is that complete health repair is possible if the rate of biological ageing is constant. Declines in health status are driven not by the rate of biological ageing, but by its rate of increase. Case and Deaton (2005) argue, however, that a technology that can effect such complete health repair is implausible.[2]

In this paper I argue that there are two fairly simple, but so far largely misunderstood, solutions to two issues that give rise to the above limitations of the Grossman model, as follows. First, I show that the condition for the "equilibrium" health stock is in fact the first-order condition for health investment. This condition should be used to determine health investment and the dynamic equation for health should be used to determine health. Second, the health-capital literature makes two important assumptions, for mathematical convenience, which lead to a degeneracy.[3] This degeneracy arises in the Grossman model under the commonly made assumptions of (i) a health-production function $f[I(t)]$ that is linear in health investment, i.e., $f[I(t)] = I(t)$, and (ii) a health-investment process in which goods/services purchased in the market $m(t)$ and own time $\tau_I(t)$ combine to produce investment $I(t)$ according to a constant returns to scale (CRTS) technology, i.e. $I(t) = \mu(t)m(t)^{\kappa_I}\tau_I(t)^{1-\kappa_I}$. Under these two assumptions the Hamiltonian of the constrained optimization problem is linear in $I(t)$, $m(t)$, and $\tau_I(t)$. Since the optimality condition for these three (related) controls are derived by taking the derivative of the Hamiltonian with respect to the controls, the controls themselves are no longer part of the optimality condition (they drop out because of the imposed linearity) and their value cannot be determined. Lacking ability to derive an expression for investment, the model essentially breaks down: without investment it is not possible to derive the paths of health, wealth, and consumption.

Ehrlich and Chuma (1990) were the first to note this indeterminacy ("bang-bang") problem. Still, the importance of this observation has gone largely unnoticed:

[2] I highlight these four criticisms because they represent, in my opinion, important critiques of the canonical theory that have not yet been satisfactorily addressed. There are other criticisms of the model. For example, I do not include criticisms that argue for useful extensions, such as the inclusion of uncertainty.

[3] A degenerate case is a limiting case in which a class of solutions changes its nature so as to belong to another, usually simpler, class. It has special features, which depart from the properties that are generic in the wider class, and the nature of the degenerate solution is generally lost under a small perturbation.

contributions to the literature that followed the publication of Ehrlich and Chuma's work in 1990 have continued to assume a health-production function with CRTS in health investment,[4] and the issue is deemed to be insufficiently important in a recent critique of the Grossman model (Zweifel, 2012) as well as in a subsequent reply to Zweifel in the model's defense (Kaestner, 2013). This may be as a consequence of the following factors. First, Ehrlich and Chuma's finding that health investment is undetermined, under the usual assumption of a CRTS health-production process, is incidental to their main contribution of introducing the demand for longevity and the authors did not explore the full implications of relaxing the assumption. Second, Ehrlich and Chuma's argument is brief and technical.[5] This has led Ried (1998) to conclude that *"...[Ehrlich and Chuma] fail to substantiate either claim [bang-bang and indeterminacy] ..."* Third, there is an incorrect notion that Ehrlich and Chuma had changed the structure of the model substantially and that *"...their results need not apply to the original model."* (Ried, 1998). Last, it is thought that introducing decreasing returns to scale (DRTS) has little impact on the basic characteristics and predictions of the model (Grossman, 2000; Kaestner, 2013). For example, it is believed that it results in individuals reaching the desired health stock gradually rather than instantaneously (e.g., Grossman, 2000, p. 364) – perhaps not a sufficiently important improvement to warrant the introduction of DRTS.

I present a theory of the demand for health, health investment and longevity based on Grossman (1972a, 1972b) and the extended version of this model by Ehrlich and Chuma (1990). Specifically, I stay as close as possible to Grossman's original formulation, but allow for the possibility that the health-production process is not strictly linear in health investment, i.e. $f[I(t)] = I(t)^{\alpha}$, where $\alpha \neq 1$, and focus on the DRTS case, i.e. $0 < \alpha < 1$. I also model endogenous longevity (as in Ehrlich and Chuma, 1990): individuals invest in the quantity of life (longevity) as well as the quality of life (health and consumption). The model is similar to Ehrlich and Chuma (1990).[6]

The contribution of this paper is as follows. I start by providing detailed proof of Ehrlich and Chuma's (1990) claim that health investment is not determined for the canonical CRTS model. This is important as Ried (1998) has concluded that Ehrlich and Chuma (1990) failed to substantiate their claim, the literature does not seem to consider the issue to be of importance, and there are some that question whether the issue exists altogether (e.g., Ried, 1998; Grossman, 2000; Laporte, 2014; Strulik, 2014). This contribution reaffirms and strengthens Ehrlich and Chuma's (1990) observation.

However, addressing the indeterminacy of investment is not sufficient to address the

[4] E.g., Bolin et al. (2001, 2002a, 2002b, 2003), Case and Deaton (2005), Erbsland, Ried and Ulrich (2002), Jacobsen (2000), Leu and Gerfin (1992), Liljas (1998), Nocera and Zweifel (1998), Ried (1996, 1998). To the best of my knowledge, the only exception is an unpublished working paper by Dustmann and Windmeijer (2000) who take Ehrlich and Chuma (1990) as their point of departure.

[5] It involves a reference to a graph with health investment on one axis and the ratio of two Lagrange multipliers on the other. The authors note that the same results hold in a discrete time setting, using a complicated proof based on the last period preceding death (see their footnote 4).

[6] Ehrlich and Chuma (1990) model DRTS as operating through the cost of investment function, whereas I model DRTS through the production function of health. Mathematically these two approaches are similar.

four major criticisms directed at the Grossman model. I therefore also make the following unique contributions. First, I argue for a reinterpretation of the "equilibrium" condition for health, one of the most central relations in the health-capital literature. Instead of determining the health stock, as is generally assumed,[7] it is a dynamic relation determining the optimal level of health investment. This condition should be used to determine health investment and the dynamic equation for health should be used to determine health.[8] I show that incorrect use of these conditions gives rise to the four major criticisms.

Second, I show that CRTS are associated with repeated "bang-bang" behavior in which at any time there exists a difference between the actual health stock and an "equilibrium" health stock that is being dissipated by instantaneous adjustments of the stock. Such bang-bang behavior, however, cannot be analyzed analytically, and, Ried (1998) has correctly suggested that it is not consistent with the notion of equilibrium. I therefore conclude that the standard CRTS assumption is significantly flawed, representing a degenerate case, i.e. a highly unusual special case (of a more general class of models) for which the model breaks down. Abandoning this unnecessary and restrictive assumption, adopting the reinterpretation of the equilibrium condition, and analyzing the resulting dynamic model I find the theory to be capable of reproducing the phenomena discussed above and of addressing the four major criticisms leveled at the Grossman model.

Third, the static relations that are often employed in empirical testing of the Grossman model would require the equilibrium health stock to be associated with the steady-state equilibrium (so that it has no "memory" of the past). However, I show that the steady state is not the solution to the optimization problem, that the Grossman model is dynamic, and that the current health state depends on past conditions (addressing the criticism of Usher, 1975, and Wagstaff, 1993). Also, the intuition that individuals reach a certain "desired" health stock gradually (for DRTS) rather than instantaneously (for CRTS) is not correct, since there exists no such desired health level for DRTS.

Fourth, despite functioning for more than 40 years as the canonical theory for the demand for health, there are as of yet no publications presenting its comparative dynamics in much detail.[9] I present detailed comparative dynamic analyses of the full DRTS Grossman model for variation in socioeconomic status (wealth, wages, and education) and health. I obtain several new results. Wealthy (high SES) individuals invest more in health,

[7] The notion of an "equilibrium" condition for health is common. See, e.g., the discussions on p. 355-356 of Grossman (2000) and p. 769-770 of Ehrlich and Chuma (1990), Cropper (1981), Muurinen (1982), Wagstaff (1986a), and the references in footnote 4.

[8] Of the references in footnotes 4 and 7, Erbsland, Ried and Ulrich (2002), Grossman (1972a, 1972b, 2000), Leu and Gerfin (1992), Muurinen (1982), Nocera and Zweifel (1998), and Ried (1996, 1998) are explicit about the use of the dynamic equation for health to determine health investment (the opposite of its intended use), whereas the remaining references are silent on the determination of investment.

[9] Ried (1998) has conducted such an analysis, but since he relies on CRTS his analysis suffers from the aforementioned degeneracy. Ehrlich and Chuma (1990) have presented directional results (broadly whether an effect is likely to be positive or negative; their Table 3). Eisenring (1999) presents a comparative dynamic analysis of a highly simplified Grossman-type model without consumption, without assets, and for fixed length of life. In short, the comparative dynamics of the Grossman model have not yet been fully analyzed, in particular since such analyses are not limited to simple directional predictions.

and as result their health deteriorates more slowly (addressing the first criticism of Case and Deaton, 2005), and they live longer. Health investment is higher and increases less rapidly with age for wealthier (higher SES) individuals if wealth (SES) enables moderate life extension (flatter investment profile) and more rapidly with age if wealth (SES) enables substantial life extension (steeper investment profile). And, healthy individuals potentially invest less in health (addressing the criticism of Wagstaff, 1986a; Zweifel and Breyer, 1997).

Fifth, an interesting consequence of the literature's focus on the degenerate case is that empirical tests of the Grossman model are still in their infancy. Thus far, tests of the Grossman model have almost exclusively relied on relations derived from the degenerate model and have employed the incorrect equilibrium condition.[10] Absent an equivalent relation for a DRTS health-production process I derive a structural dynamic relation between health and health investment to guide empirical testing.

Last, I perform numerical simulations to illustrate the properties of the theory. These simulations corroborate the analytical results presented in this paper and show that the model is capable of reproducing the rapid increase in health investment near the end of life, and that the optimal solution for length of life is finite (even for a constant biological aging rate), addressing the second criticism by Case and Deaton (2005) that health-production models are characterized by complete health repair. Thus all four criticisms are resolved.

The implications of this work extend beyond health economics. Economic theories are often formulated assuming linear relations in controls. As a result, many economic models suffer from similar forms of degeneracy, associated with indeterminacy and "bang-bang". For example, the theory of firm-investment behavior also assumes CRTS in investment and employs comparable conditions for investment and comparable notions of an "equilibrium" or "desired" level of capital (e.g., Jorgensen 1963, 1967). The theory of investment behavior in turn provides the foundation for the analysis of firms in models of economic growth. As my analyses suggest, while linearity may appear attractive for its simplicity, it can have serious and relatively hidden consequences.

The paper is organized as follows. Section 2 presents the model in continuous time and discusses the characteristics of the first-order conditions. In particular this section offers an alternative interpretation of the equilibrium condition for health. Section 3 explores the properties of a DRTS health-production process, in several ways, by: a) exploring a stylized representation of the first-order condition for health investment to gain an intuitive understanding of its properties, b) analyzing the effect of differences in socioeconomic status (wealth, wages, and education) and health on the optimal level of health investment, c) developing structural-form relations for empirical testing, and d)

[10] E.g., Erbsland, Ried, and Ulrich (2002); Gerdtham and Johannesson (1999); Gerdtham et al. (1999); Grossman (1972a); Leu and Doppmann (1986); Leu and Gerfin (1992); Nocera and Zweifel (1998); Van de Ven and Van der Gaag (1982); Wagstaff (1986a, 1993); Van Doorslaer (1987). A more recent literature has numerically (not analytically as in this paper) solved and estimated dynamic formulations based (sometimes loosely) on health-capital theory using dynamic programming techniques, and taking into account a health-production process that is subject to decreasing returns to scale in investment (e.g., Ehrlich and Yin, 2005; Fonseca et al. 2013; Gilleskie, 1998, 2010; Hugonnier, Pelgrin and St-Amour, 2013; Khwaja, 2010; Scholz and Seshadri, 2012; Yogo, 2009).

presenting numerical simulations of the model. Section 4 summarizes and concludes. The Appendix provides detailed derivations and mathematical proofs.

2 The demand for health, health investment, and longevity

I follow Grossman's basic formulation (Grossman, 1972a, 1972b, 2000) for the demand for health and health investment in continuous time, with two extensions: I allow 1) for a flexible health-production process with DRTS in investment, and 2) for length of life to be optimally chosen (as in Ehrlich and Chuma, 1990).[11]

Using continuous time optimal control (e.g., Caputo 2005) the problem can be stated as follows. Individuals maximize the life-time utility function

$$\int_0^T U[C(t), H(t)]e^{-\beta t}dt, \tag{1}$$

where T denotes length of life (endogenous), β is a subjective discount factor, and individuals derive utility $U[C(t), H(t)]$ from consumption $C(t)$ and from health $H(t)$. Time t is measured from the time individuals begin employment. Utility increases with consumption $\partial U/\partial C > 0$ and with health $\partial U/\partial H > 0$.

The objective function (1) is maximized subject to the dynamic constraints:

$$\frac{\partial H(t)}{\partial t} = f[I(t)] - d(t)H(t), \tag{2}$$

$$\frac{\partial A(t)}{\partial t} = \delta A(t) + Y[H(t)] - p_X(t)X(t) - p_m(t)m(t), \tag{3}$$

the total time budget Ω

$$\Omega = \tau_w(t) + \tau_I(t) + \tau_C(t) + s[H(t)], \tag{4}$$

and initial and end conditions: H_0, H_T, A_0 and A_T are given. Life can no longer be sustained below a minimum health level $H_{\min}$ ($H(T) = H_T \equiv H_{\min}$).

Health (equation 2) can be improved through investment in health $I(t)$ and deteriorates at the biological aging rate $d(t)$. The relation between the input, health investment $I(t)$, and the output, health improvement, is governed by the health-production function $f[I(t)]$, assumed to obey the law of diminishing marginal returns to health investment ($\partial^2 f/\partial I^2 < 0$). For simplicity of exposition I use the following simple functional form

$$f[I(t)] = I(t)^{\alpha}, \tag{5}$$

[11] In line with Grossman (1972a; 1972b) and Ehrlich and Chuma (1990) I do not incorporate uncertainty. This would unnecessarily complicate the optimization problem and is not needed to explain the stylized facts regarding health behavior discussed in this paper. For a detailed treatment of uncertainty within the Grossman model the reader is referred to Ehrlich (2000), Liljas (1998), and Ehrlich and Yin (2005).

where $0 < \alpha < 1$ (DRTS).[12] Note that for $\alpha = 1$ we retrieve the usual formulation (Grossman 1972a, 1972b).

Assets $A(t)$ (equation 3) provide a return δ (the rate of return on capital), increase with income $Y[H(t)]$ and decrease with purchases in the market of consumption goods and services $X(t)$ and health investment goods and services $m(t)$ (e.g., medical care) at prices $p_X(t)$ and $p_m(t)$, respectively. Income $Y[H(t)]$ is assumed to be increasing in health $H(t)$ as healthy individuals are more productive and earn higher wages (Currie and Madrian, 1999; Contoyannis and Rice, 2001).

Goods and services $X(t)$ purchased in the market and own time inputs $\tau_C(t)$ are used in the production of consumption $C(t)$. Similarly goods and services $m(t)$ and own time inputs $\tau_I(t)$ (e.g., exercise, time spent visiting the doctor) are used in the production of health investment $I(t)$. The efficiencies of production are assumed to be a function of the consumer's stock of knowledge E (an individual's human capital exclusive of health capital [e.g., education]) as the more educated may be more efficient at investing in health (see, e.g., Grossman 2000):

$$I(t) = I[m(t), \tau_I(t); E], \tag{6}$$

$$C(t) = C[X(t), \tau_C(t); E]. \tag{7}$$

The total time available in any period Ω (equation 4) is the sum of all possible uses $\tau_w(t)$ (work), $\tau_I(t)$ (health investment), $\tau_C(t)$ (consumption) and $s[H(t)]$ (sick time; a decreasing function of health). In this formulation one can interpret $\tau_C(t)$, the own-time input into consumption $C(t)$ as representing leisure.[13]

Income $Y[H(t)]$ is taken to be a function of the wage rate $w(t)$ times the amount of time spent working $\tau_w(t)$,

$$Y[H(t)] = w(t)\left\{\Omega - \tau_I(t) - \tau_C(t) - s[H(t)]\right\}. \tag{8}$$

Thus, we have the following optimal control problem: the objective function (1) is maximized with respect to the control functions $X(t)$, $\tau_C(t)$, $m(t)$, and $\tau_I(t)$, subject to the constraints (2, 3 and 4). The Hamiltonian of this problem is:

$$\Im(t) = U[C(t), H(t)]e^{-\beta t} + q_H(t)\frac{\partial H}{\partial t} + q_A(t)\frac{\partial A}{\partial t}, \tag{9}$$

where $q_H(t)$ is the co-state variable associated with the dynamic equation (2) for the

[12] Mathematically, equation (5) is equivalent to Ehrlich and Chuma's (1990) assumption of a dual cost-of-investment function with decreasing returns to scale (their equation 5) and a linear health-production process ($\alpha = 1$ in this paper's equation 5). Conceptually, however, there is an important distinction: it is not the process of health investment (equation 6 in this paper) but the process of health production (the ultimate effect on health) that is expected to exhibit decreasing returns to scale.

[13] Because consumption consists of time inputs and purchases of goods/services in the market one can conceive leisure as a form of consumption consisting entirely or mostly of time inputs. Leisure provides utility and its cost consists of the price of goods/services utilized and the opportunity cost of time.

state variable health $H(t)$, and $q_A(t)$ is the co-state variable associated with the dynamic equation (3) for the state variable assets $A(t)$.[14]

The co-state variables $q_H(t)$, and $q_A(t)$, find a natural economic interpretation in the following standard result from Pontryagin

$$q_H(t) = \frac{\partial}{\partial H(t)} \int_t^{T^*} U(*) e^{-\beta s} ds, \tag{10}$$

$$q_A(t) = \frac{\partial}{\partial A(t)} \int_t^{T^*} U(*) e^{-\beta s} ds, \tag{11}$$

(e.g., Caputo 2005, eq. 21 p. 86), where T^* denotes optimal length of life, and $U(*)$ denotes the maximized utility function (i.e., along the optimal paths for the controls, state functions, and for the optimal length of life). Thus, for example, $q_H(t)$ represents the marginal value of remaining lifetime utility (from t onward) derived from additional health $H(t)$. I refer to the co-state functions as the "marginal value of health" and the "marginal value of wealth".

As Ehrlich and Chuma (1990) noted, a transversality condition is required for the optimal length of life T. The condition follows from the dynamic envelope theorem (e.g., Theorem 9.1, p. 232 of Caputo, 2005):

$$\frac{\partial}{\partial T} \int_t^{T^*} U(*) e^{-\beta s} ds = \frac{\partial}{\partial T} \int_0^{T} \Im(t) dt \quad = \quad \Im(T) = 0. \tag{12}$$

Thus, $\Im(T)$ represents the marginal value of remaining lifetime utility (from t onward) derived from additional longevity T), i.e. it is the "marginal value of life extension", and the age at which life extension no longer has value defines the optimal length of life T^*.

2.1 First-order conditions

Maximization of (9) with respect to the control functions $m(t)$ and $\tau_I(t)$ leads to the first-order condition for health investment $I(t)$

$$\pi_I(t) = \frac{q_H(t)}{q_A(t)}, \tag{13}$$

where $\pi_I(t)$ is the marginal cost of health investment $I(t)$

$$\pi_I(t) \equiv \frac{p_m(t) I(t)^{1-\alpha}}{\alpha[\partial I/\partial m]} = \frac{w(t) I(t)^{1-\alpha}}{\alpha[\partial I/\partial \tau_I]}. \tag{14}$$

[14] Strictly speaking, an additional term in the Hamiltonian (making it officially a Lagrangian) is needed, $\lambda_{H_{\min}}(t)\,[H(t) - H_{\min}]$, associated with the condition that $H(t) > H_{\min}$ for $t < T$ ($\lambda_{H_{\min}}(t) = 0$ if $H(t) > H_{\min}$ and $\lambda_{H_{\min}}(t) > 0$ if $H(t) \leq H_{\min}$). In practice, employing the condition entails restricting solutions to those where the constraint is not imposing (section 3.4 shows how this is implemented). Thus, for any feasible solution we have $\lambda_{H_{\min}}(t) = 0$ for all t, and the condition can be ignored.

Using (50), integrating the dynamic co-state equation for the marginal value of health $q_H(t)$ (51), and inserting the expression in (13) we obtain an alternative expression for the first-order condition for health investment $I(t)$

$$\pi_I(t) = \pi_I(0)e^{\int_0^t [d(u)+\delta]du} - \int_0^t \left[\frac{1}{q_A(0)} \frac{\partial U}{\partial H} e^{-(\beta-\delta)s} + \frac{\partial Y}{\partial H} \right] e^{\int_s^t [d(u)+\delta]du} ds, \tag{15}$$

where $q_A(0)$ is the marginal value of initial wealth.[15]

Using either the expression (15) or (16) for the first-order condition for health investment and taking the derivative with respect to time t we obtain the following expression

$$\frac{1}{q_A(0)} \frac{\partial U}{\partial H} e^{-(\beta-\delta)t} + \frac{\partial Y}{\partial H} = \sigma_H(t), \tag{17}$$

where $\sigma_H(t)$ is referred to as the user cost of health capital at the margin

$$\sigma_H(t) \quad \equiv \quad \pi_I(t)[d(t)+\delta] - \frac{\partial \pi_I}{\partial t}. \tag{18}$$

Last, maximization of (9) with respect to the control functions $X(t)$ and $\tau_C(t)$ leads to the first-order condition for consumption $C(t)$

$$\frac{\partial U}{\partial C} \quad = \quad q_A(0)\pi_C(t)e^{(\beta-\delta)t}, \tag{19}$$

where $\pi_C(t)$ is the marginal cost of consumption $C(t)$

$$\pi_C(t) \equiv \frac{p_X(t)}{\partial C/\partial X} = \frac{w(t)}{\partial C/\partial \tau_C}. \tag{20}$$

The first-order condition (13) (or the alternative forms 15, 16, and 17) determines the optimal solution for the control function health investment $I(t)$. The first-order condition (19) determines the optimal solution for the control function consumption $C(t)$.[16] The solutions for the state functions health $H(t)$ and assets $A(t)$ then follow from the dynamic equations (2) and (3). These optimal control and optimal state functions are functions of the co-states $q_A(t)$ and $q_H(t)$. The co-states are obtained by solving the dynamic co-state equations (50) and (51), imposing the life-time budget constraints for wealth and health (integrating 2 and 3), and imposing the begin and end conditions. Last, length of life T is determined by the transversality condition (12). This dynamic (step) process is explained in more detail in section 3.4.1.

[15] One can also use the final period T as point of reference, to obtain

$$\pi_I(t) = \pi_I(T)e^{-\int_t^T [d(u)+\delta]du} + \int_t^T \left[\frac{1}{q_A(0)} \frac{\partial U}{\partial H} e^{-(\beta-\delta)s} + \frac{\partial Y}{\partial H} \right] e^{-\int_t^s [d(u)+\delta]du} ds. \tag{16}$$

[16] Because the first-order condition for health investment goods / services $m(t)$ and the first-order condition for own time inputs $\tau_I(t)$ are identical (see Appendix section A) one can consider a single control function $I(t)$ (health investment) instead of two control functions $m(t)$ and $\tau_I(t)$. The same is true for consumption $C(t)$. Because of this property, the optimization problem is reduced to two control functions $I(t)$ and $C(t)$ (instead of four) and two state functions $H(t)$ and $A(t)$.

2.2 Assumptions

I make the usual assumptions of diminishing marginal utilities of consumption $\partial^2 U/\partial C^2 < 0$ and of health $\partial^2 U/\partial H^2 < 0$, and diminishing marginal production benefit of health $\partial^2 Y/\partial H^2 < 0$. In addition, for illustrative purposes and to stay as close as possible to Grossman's original formulation, I make the usual assumption of a Cobb-Douglas CRTS relation between the inputs goods/services purchased in the market and own-time and the output health investment $I(t)$ (Grossman, 1972a, 1972b)

$$I(t) \quad = \quad \mu_I(t) m(t)^{1-k_I} \tau_I(t)^{k_I}, \tag{21}$$

where $\mu_I(t)$ is an efficiency factor and $1 - k_I$ and k_I are the elasticities of investment in health $I(t)$ with respect to goods and services $m(t)$ purchased in the market and with respect to own-time $\tau_I(t)$, respectively. Using equations (14) and (21) we have

$$\pi_I(t) \quad = \quad \frac{p_m(t)^{1-k_I}\, w(t)^{k_I}}{\alpha\, k_I^{k_I} (1-k_I)^{1-k_I}\, \mu_I(t)} I(t)^{1-\alpha} \equiv \pi_I(t)^* I(t)^{1-\alpha}. \tag{22}$$

The marginal cost of health investment $\pi_I(t)$ increases with the price of investment goods and services $p_m(t)$, and the opportunity cost of not working $w(t)$, and decreases in the efficiency $\mu_I(t)$ of the use of investment inputs in the investment process (21). Because of diminishing returns to scale in investment, the marginal cost of health investment $\pi_I(t)$ is an increasing function of the level of investment $I(t)$.[17] In contrast, for the canonical CRTS health-production process ($\alpha = 1$) the marginal cost of health investment is no longer a function of the level of health investment $I(t)$.

2.3 An alternative interpretation of the first-order condition

One of the most central relations in the health-capital literature is the first-order condition (17). This relation equates the sum of the marginal consumption benefit of health $q_A(0)^{-1}\partial U/\partial H$ and the marginal production benefit of health $\partial Y/\partial H$ to the user cost of health capital $\sigma_H(t)$, and is interpreted as an equilibrium condition for the health stock $H(t)$.[18] An alternative interpretation of (17) is, however, that this relation determines the optimal level of health investment $I(t)$. The argument is as follows.

First, the first-order condition (17) is the result of maximization of the optimal control problem with respect to investment in health and hence, first and foremost, it determines the optimal level of investment $I(t)$. Optimal control theory distinguishes between control functions and state functions. Control functions are determined by the first-order conditions and state functions are determined by the dynamic equations (e.g.,

[17] This is because for DRTS the health improvement $f[I(t)] = I(t)^\alpha$ of an additional amount of investment $I(t)$ is smaller the higher is the level of investment $I(t)$ (the production function is flatter at higher levels of investment due to concavity), and as a result the effective cost of investment $\pi_I(t)$ is higher.

[18] Condition (17) is the same as equation (11) in Grossman (2000) and equation (13) in Ehrlich and Chuma (1990). See footnote 7.

Seierstad and Sydsaeter, 1977, 1987; Kirk, 1970). The first-order condition (17) is thus naturally associated with the control function health investment $I(t)$ and the dynamic equation (2) is naturally associated with the state function health $H(t)$.[19]

Second, in the health-capital literature optimal health investment $I(t)$ is assumed to be determined by the first-order condition (13), (15), or the alternative form (16).[20] However, it can be shown that the first-order conditions (13), (15) and (16) are mathematically equivalent to (17) (for proof see Appendix section D). Thus if equation (13), (15), or (16), is the first-order condition for health investment $I(t)$ (the interpretation in the health-capital literature) then equation (17) is too (and vice versa).

In the remainder of this paper I will use relations (13), (15), (16), and (17) as being equivalent. They are alternative expressions of the first-order condition for optimal investment $I(t)$, but, they are fundamentally the same. One follows from the other through differentiation or by integration.

This result may seem trivial, and it perhaps is, but the literature interprets relations (13), (15), and (16), as being distinct from the condition (17). The former three conditions are interpreted as determining optimal health-investment decisions and the latter is employed to analyze equilibrium health. Further, because the conditions for health investment (13), (15), and (16), are complex, in practice the literature employs the dynamic equation for health (2) to obtain health investment $I(t)$.[21] Thus, the literature employs a first-order condition for health investment to derive health and a dynamic relation for health to obtain health investment (the exact opposite of their intended use).[22]

The four criticisms: With these considerations in mind, let us revisit the four criticisms directed at the Grossman model (see section 1) to understand how they arise. The condition for health investment (13) provides a relation for the demand for investment $I(t)$ as a function of current prices $p_m(t)$, current wages $w(t)$, current efficiency $\mu_I(t)$ (through $\pi_I(t)$; see 22), current health $H(t)$, and current wealth $A(t)$ (through $q_H(t)$ and $q_A(t)$). The literature then uses the condition for investment (13) to obtain an expression for health. As discussed above, this is not the correct use of this relation. After all, it is a relation describing the demand for health investment as a function of health, not the other way around. Indeed, incorrect use results in an expression for health that is a function of current conditions. In other words, the resulting solution for health is

[19] Analogously, the first-order condition (19) is associated with the control variable consumption $C(t)$ and the dynamic equation (3) is associated with the state function assets $A(t)$.

[20] Condition (13), (15) (or the alternative form 16) is the same as equation (9) in Grossman (2000) and equation (8) in Ehrlich and Chuma (1990).

[21] See, e.g., Grossman (1972a, 1972b, 2000) and footnote 8.

[22] Due to CRTS, the first-order condition for health investment is not a function of investment (see 22 for $\alpha = 1$). But it is a function of health. Since the condition can not be used to determine investment, it was perhaps used to determine health, and given health, health investment can be derived from the dynamic equation for health (2). Further, human-capital theory has roots in the theory of investment behavior of the firm. The initial literature in this area also assumes CRTS and employs comparable conditions for investment and for "equilibrium" capital (e.g., Jorgensen 1967). These concepts may have carried over.

not intrinsically dynamic and does not depend on its starting point and the history of subsequent investments made, explaining the criticisms by Usher (1975) and Wagstaff (1993). Likewise, complete health repair is possible because for CRTS one can obtain conditions for which health remains constant, explaining the criticism by Case and Deaton (2005). Further, the widening of the health gradient between socioeconomic groups is a dynamic process, but the condition that is used to obtain health is not dynamic, explaining another criticism by Case and Deaton (2005).

Conversely, the dynamic relation for health (2) describes the evolution of health as a function of initial health status and the history of aging and investments made (see 87). This relation is, however, employed to determine the demand for investment, which is found to be greater for those in better health, explaining the criticism of Wagstaff (1986a) and Zweifel and Breyer (1997).

Thus, these four criticisms can be explained by the fact that the relation for health investment (13) is used to determine health, and the dynamic relation for health (2) is used to determine health investment. In the remainder of the paper, I show in more detail that employing (2) and (13) correctly, i.e. for their intended use, addresses the four criticisms.

3 The health-production process

In this section I explore the properties of the health-production process in several ways. I present a stylized representation of the first-order condition for health investment to gain an intuitive understanding of its properties (section 3.1). In particular, I contrast the characteristics of the solution for health investment under a CRTS health-production process with that of a DRTS process, and provide additional arguments to show that the CRTS case suffers from indeterminacy of health investment (as Ehrlich and Chuma, 1990, have argued). This is not just a technicality but in fact a serious problem. Essentially, the model breaks down. For this reason I prefer to refer to the issue as a mathematical degeneracy, rather than an indeterminacy.

There are at least two ways to address the degeneracy. The preferred solution, in my opinion, is to allow for a more flexible functional form of the production process, and DRTS in investment represents the more natural case.[23] I then further develop this case, showing that, together with the alternative interpretation of the equilibrium condition for health (section 2.3), it addresses the four major criticisms leveled at the Grossman model. I explore the effect of differences in socioeconomic status (wealth, wages, and education) and health on the optimal level of health investment, health, and longevity, using comparative

[23]The other solution is to maintain the assumption of CRTS in health investment and to replace the dynamic wealth equation (3) with an instantaneous budget constraint, i.e. expenditures on consumption and health investment equal earnings at each instant (Cropper, 1977; Laporte, 2014; Strulik, 2014). Such a model, however, cannot inform us about the relation between wealth and health (since wealth is absent from the model), and is a stronger departure from the Grossman model than is DRTS in health investment. It also maintains an assumption of strict linearity in health investment which is hard to defend: twice the amount of exercise or chemotherapy does not precisely double health improvement.

dynamic methods (section 3.2), derive a structural-form relation for empirical testing of the model (section 3.3), and perform numerical simulations to illustrate the workings of the model and to corroborate analytical results (section 3.4).

3.1 Stylized representations

3.1.1 CRTS: the indeterminacy of investment

In the following I show that for a CRTS health-production process the first-order condition for health investment is not a function of investment. As a result, health investment is undetermined and associated with "bang-bang" behavior.

Indeterminacy: Figure 1 provides a stylized representation of the first-order condition (13) for health investment for a CRTS (solid lines) and DRTS (dashed lines) health-production process: it graphs the marginal benefit $q_{h/a}(t) = q_H(t)/q_A(t)$ and marginal cost $\pi_I(t)$ of health investment as a function of investment $I(t)$ (left-hand side; LHS) and as a function of health $H(t)$ (right-hand side; RHS).

Consider the LHS and CRTS (solid lines) first. The marginal benefit of health investment consists of the ratio of the marginal value of health and the marginal value of wealth $q_{h/a}(t)$. This ratio, according to Ehrlich and Chuma (1990), is not a function of health investment $I(t)$.[24] This is shown as the horizontal solid line labeled $q_{h/a}(t)$. For a CRTS health-production proces, $\pi_I(t)$ is also not a function of health investment $I(t)$ (see 22 for $\alpha = 1$). This is shown as the horizontal solid line labeled $\pi_I(t)$.

Because individuals can adjust their health only gradually through health investment, the level of the health stock $H(t)^{\text{true}}$ at age t is given and provides a constraint to the optimization problem at age t. Generally the constraint provided by $H(t)^{\text{true}}$ will result in different values for the marginal benefit $q_{h/a}(t)$ (a function of health; see 10 and 11) and marginal cost of health investment $\pi_I(t)$ (not a function of health): this is depicted by the two horizontal lines having distinct levels (they do not overlap). The intersection of the two solid curves would determine the optimal level of health investment $I(t)$ but since they run in parallel this does not occur. Thus no unique optimal solution for health investment $I(t)$ exists.[25] Each optimality condition serves a particular role (see section

[24] As Ried (1998) noted, Ehrlich and Chuma (1990) do not provide proof for their claim. However, it is a standard result of Pontryagin's maximum principle and is relatively straightforward to obtain (perhaps this explains why it was omitted). From (10) and (11) we have $q_H(t) = \partial V(t)/\partial H(t) = \partial[\int_t^{T^*} U(*)e^{-\beta s}ds]/\partial H(t)$ and $q_A(t) = \partial V(t)/\partial A(t) = \partial[\int_t^{T^*} U(*)e^{-\beta s}ds]/\partial A(t)$, where $V(t)$ is the indirect utility function. Since $V(t)$ is obtained by inserting the optimal solutions in the life-time utility function, it is no longer a function of investment, or for that matter of any control or state function (except for the current and end states).

[25] Laporte (2014) has questioned Ehrlich and Chuma's (1990) argument, as follows. She combines the first-order conditions for health investment (13) and for consumption (19) (her conditions 37 and 38) to obtain $\partial U/\partial C = q_H(t)\pi_I(t)\pi_C(t)e^{\beta t}$ (her condition 40 in my notation). This condition is identical to the first-order condition for health investment in her model, a model with an instantaneous budget constraint, for which there is no indeterminacy. There is no indeterminacy in her model because with an instantaneous budget constraint $I(t) = Y(t) - C(t)$, and since consumption $C(t)$ is determined, health investment $I(t)$

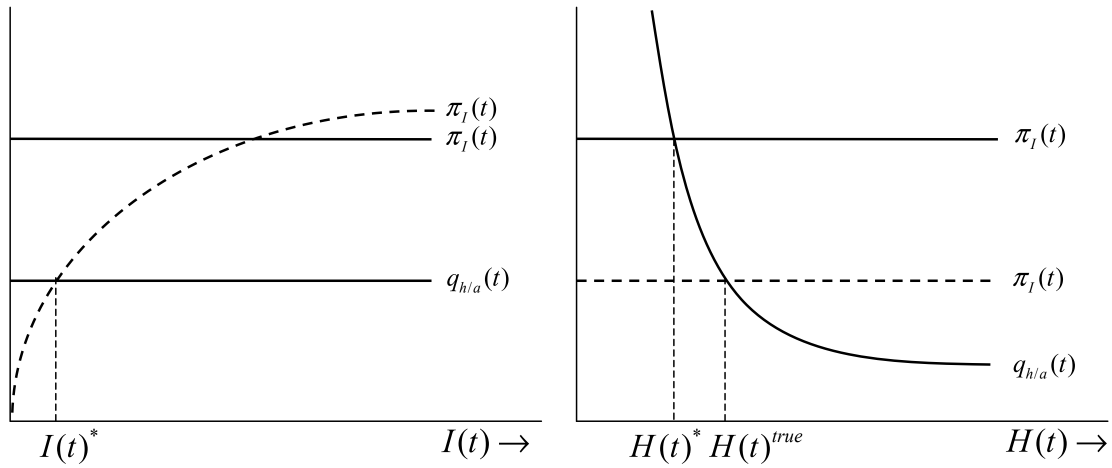

Figure 1: *Marginal benefit $q_{h/a}(t) = q_H(t)/q_A(t)$ versus marginal cost $\pi_I(t)$ of health investment for a CRTS (solid lines) and DRTS (dashed lines) health-production process.*

2.1). If one of the relations breaks down the entire model breaks down.[26]

Bang-bang: Now consider the RHS of Figure 1 and CRTS (solid lines). The marginal benefit of health investment $q_{h/a}(t)$ is downward sloping to represent the case where healthy individuals invest less in health.[27] The marginal cost of health investment $\pi_I(t)$ is independent of health (see 22). As the graph shows, for the actual level of health $H(t)^{\text{true}}$ the marginal benefit $q_{h/a}(t)$ is not equal to the marginal cost $\pi_I(t)$. Thus, there is a discrepancy between the marginal benefit $q_{h/a}(t)$ and marginal cost $\pi_I(t)$ of health investment. There is, however, a unique level of health $H(t)^*$ for which they are equal. The health-capital literature assumes this unique solution $H(t)^*$ describes the "equilibrium" or "desired" health stock.[28] If health could be adjusted to the level $H(t)^*$, condition (13) would equilibrate. But, health can adjust only gradually through investments made over time (see 2). The usual interpretation is that this results in a so-called "bang-bang" control (e.g., Ehrlich and Chuma, 1990), i.e. a control that switches abruptly between (boundary) states. For the Grossman model, however, there is no clear upper or lower boundary to health investment (the control)[29] and the implicit assumption is that health investment

is also determined. The problem with this argument, however, is that in the Grossman model with a dynamic budget constraint the first-order conditions (13) and (19) also need to hold independently, and as discussed above, the condition for optimal investment (13) does not hold for CRTS.

[26] Proof is provided in Appendix section E that the indeterminacy also holds for the alternative expressions for the first-order condition for health investment (15), (16), and (17).

[27] The opposite case is possible too, but the general result obtained here does not depend on it.

[28] The assumption is made explicitly by assuming (17) determines health. See footnote 7.

[29] Galama and Kapteyn (2011) and Galama et al. (2013), however, impose non-negative investment.

(or disinvestment, depending on whether $H(t)^{\text{true}}$ is above or below $H(t)^*$) is infinitely large for an infinitesimally small amount of time (so that $\int_t^{t+\delta t} I(s)ds$ is still finite) and health jumps instantaneously from $H(t)^{\text{true}}$ to $H(t)^*$, so that (13) equilibrates.[30] Ried (1998) correctly identifies such behavior as inconsistent with equilibrium.

Further, even for $H(t)^*$ the first-order condition for health investment (13) holds only momentarily. An instant ϵ later, the marginal cost $\pi_I(t+\epsilon)$ is no longer equal to the marginal benefit $q_{h/a}(t+\epsilon)$, since the marginal cost is exogenously determined for CRTS (see 22 for $\alpha = 1$), and there is no guarantee that it evolves in step with the marginal benefit.[31] As a result, the difference in health needs to be repeatedly dissipated (there is no end to the bang-bang behavior). Such a model cannot be analyzed analytically.[32]

Last, the different experiences of developing and developed countries suggest that the economic principle of eventually diminishing returns applies to health production. Quite modest increases in expenditures on health inputs (food, sanitation) have relatively large impacts on health in the developing world, whereas large increases in resources in the developed world have a relatively modest impact (e.g., Wagstaff, 1986b).

Thus, empirical and analytical considerations suggest the restrictive assumption of CRTS has to be avoided.

3.1.2 DRTS: the dynamic nature of investment and health

In contrast to a CRTS health-production process, for a DRTS process ($\partial^2 f/\partial I^2 < 0$) the marginal cost of health investment $\pi_I(t)$ is increasing in the level of investment (see 22). As a result, a unique level of investment $I(t)^*$ exists (see LHS of Figure 1, dashed line), for which the marginal cost of health investment $\pi_I(t)$ equals the marginal benefit of health investment $q_{h/a}(t)$ for the true level of health $H(t)^{\text{true}}$ (see RHS of Figure 1, dashed line). DRTS thus addresses the indeterminacy (e.g., Ehrlich and Chuma, 1990) as well as the "bang-bang" nature of investment, since there is no longer a discrepancy between $H^*(t)$ and $H(t)^{\text{true}}$ (they are identical).[33] In other words, for DRTS there is no "desired" health stock $H(t)^*$. Further, because a unique solution for investment $I(t)^*$ exists for every period t, the solution for health $H(t)$ can be obtained through (2; see also 87), and health is found

[30] Implicitly, by assuming (13) (or 15, 16, or 17) holds at all times for CRTS (as the literature does), the difference in health is assumed to dissipate instantaneously (or else 13 would not hold) and at no cost (the jump in investment is not accounted for in the budget constraint 3).

[31] Formal proof is provided in Appendix section F.

[32] Because of this, the mathematics field of optimal control devotes effort to finding optimal algorithms and programming methods to compute bang-bang constrained optimal-control problems (e.g., see the introduction of Kaya, Lucas and Simakov, 2004). Arrow (1968) concludes that the theory of the firm with linear investment in capital (very similar to the Grossman model) cannot be solved analytically, and presents an algorithm to describe the bang-bang solution.

[33] For those who remain skeptical about the indeterminacy, the challenge is to use the first-order condition for health investment (13, or the alternative forms 15, 16, and 17) to obtain an explicit expression for health investment $I(t)$. For DRTS, (13) with (22) provides such an expression. For CRTS ($\alpha = 1$) this relation breaks down. Analogously, (19) provides an explicit relation for consumption $C(t)$ because of diminishing marginal utility, but, if utility were linear in consumption, consumption would be undetermined.

to be a function of past levels of health investment $I(s)$ and past biological aging rates $d(s)$ $(0 \leq s < t)$ addressing the criticism of Usher (1975) and Wagstaff (1993).[34]

In the following discussion, it is useful to conceptually equate the marginal cost of investment with the level of investment, as high marginal cost implies a high level of investment (see 22), for given exogenous wages $w(t)$, prices $p_m(t)$, and efficiency of investment $\mu_I(t)$. Relation (17) in the form

$$\frac{\partial \pi_I(t)}{\partial t} \quad = \quad \pi_I(t)[d(t)+\delta] - \frac{1}{q_A(0)}\frac{\partial U}{\partial H}e^{-[\beta-\delta]t} - \frac{\partial Y}{\partial H}, \tag{23}$$

is then best thought of as a dynamic relation describing the change in the level of health investment $\partial I(t)/\partial t$ given the level of health $H(t)$ (and other factors) at age t. Further, from (2), (5) and (22) we obtain a dynamic relation for the change in the health stock

$$\frac{\partial H(t)}{\partial t} = \left(\frac{\pi_I(t)}{\pi_I(t)^*}\right)^{\alpha/(1-\alpha)} - d(t)H(t). \tag{24}$$

Relations (23) and (24) include time derivatives, illustrating the dynamic nature of the Grossman model. Both relations are functions of the marginal cost of health investment $\pi_I(t)$ and of the health stock $H(t)$. The phase diagram in Figure 2 shows the direction of motion of the system of first-order ordinary differential equations defined by (23) and (24), as a function of the marginal cost of health investment $\pi_I(t)$ (vertical axis) versus the health stock $H(t)$ (horizontal axis).

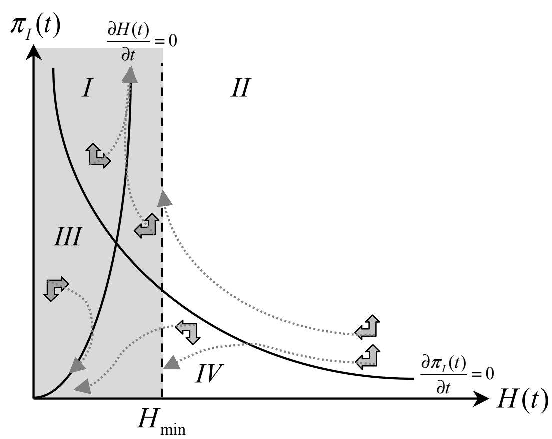

Figure 2: *Phase diagram of the marginal cost of health investment $\pi_I(t)$ versus health $H(t)$.*

[34]Current health $H(t)$ is a complicated function of past conditions through the dependence of health investment on current and past health, wealth, prices, the return to capital, the biological aging rate, the subjective discount rate, wages, and the efficiencies of health investment and consumption.

Regime switches occur when $\partial \pi_I(t)/\partial t = 0$ and $\partial H(t)/\partial t = 0$. These boundaries between regimes, so called null-clines, are shown by the thick lines in Figure 2 and are obtained by setting the derivatives to zero in (23) and (24), respectively. Because of diminishing marginal consumption $\partial^2 U/\partial H^2 < 0$ and production $\partial^2 Y/\partial H^2 < 0$ benefit, the $\partial \pi_I(t)/\partial t = 0$ null-cline is downward sloping. The $\partial H(t)/\partial t = 0$ null-cline is upward sloping. It can be either a convex or a concave relation between $\pi_I(t)$ and $H(t)$ (see 24). Here, I show a convex relation for illustrative purposes.

The two null-clines define four distinct regions I, II, III and IV. The left-up, right-up, left-down, and right-down block arrows indicate the direction of motion in the phase diagram and the grey dotted lines provide example trajectories. For example, every point in region III is associated with an evolution toward lower marginal cost of investment $\pi_I(t)$ (i.e. lower investment $I(t)$) and higher health $H(t)$.

While the null-clines are functions of age (directly and also indirectly through dependence on consumption), and therefore shift over time, the nature of the diagram is essentially unchanged as the system evolves, i.e., there are always four dynamic regions, the $\pi_I(t)/\partial t$ null-cline is always downward sloping, and the $\partial H(t)/\partial t$ null-cline is always upward sloping and intersects the origin.

The intersection of the two null-clines defines the steady state, at which both $\pi_I(t)$ and $H(t)$ would be temporarily at a stand-still. The steady state, is however of little interest as a potential solution for the system. First, it is saddle-point unstable. This is clear from visual inspection of the phase diagram: a small deviation (perturbation) from the steady state will evolve away from the steady state (the block arrows point away from the steady state), except if the deviation landed on the saddle-path (the unique trajectory in region III or region II that eventually leads to the steady state).[35] Second, the steady state is not steady: it keeps evolving / shifting. Thus, even if the trajectory was on the infinitesimally narrow (knife edge) saddle-point path trajectory, if the steady-state moves, the trajectory will no longer be on the saddle-point path. For the steady state to be fixed (constant over time) the system of first-order ordinary differential equations, given by (23) and (24), has to be of the autonomous kind: a system that does not explicitly depend on the independent variable (in this case time t). This imposes significant and unrealistic constraints on the nature of the model.[36] Third, even if the system were of the

[35] A formal proof of the instability of the steady state can be obtained by calculating the Jacobian $J(\pi_I^*, H^*)$ of the linearized system at the steady state (π_I^*, H^*) and showing that the two eigenvalues are real, non-zero, and unequal $\{\text{tr}[J(\pi_I^*, H^*)]\}^2 > 4\det[J(\pi_I^*, H^*)]$ (theorem 13.6, p. 354 of Caputo, 2005).

[36] Autonomous systems of ordinary differential equations are systems that do not explicitly depend on the independent variable (in this case time t): $\partial \mathbf{x}(t)/\partial t = f[\mathbf{x}(t)]$, where $\mathbf{x}(t)$ is a vector of N functions. A steady state is then defined by the condition $\partial \mathbf{x}(t)/\partial t = f[\mathbf{x}(t)] = \mathbf{0}$. Because the system does not explicitly depend on time t, the steady state is stationary at all times t. If the system is not of the autonomous kind, the condition $\partial \mathbf{x}(t)/\partial t = g[t, \mathbf{x}(t)] = \mathbf{0}$ holds only temporarily (since $g[t, \mathbf{x}(t)]$ depends explicitly on t), and the steady state is not stationary. The Grossman model is not autonomous, unless one makes some very strong assumptions: that utility is additively separable in consumption $C(t)$ and health $H(t)$, that the aging rate $d(t)$, the rate of return to capital δ, prices $p_m(t)$, wages $w(t)$, and the efficiency of health investment $\mu_I(t)$ are constant, and that $\beta = \delta$.

autonomous kind, any trajectory that starts at a point that is not a steady state, cannot reach a steady state in a finite amount of time (Theorem 13.4, p. 350, Caputo 2005). Thus, for the steady state to be a solution to the model, the system would have to be of the autonomous kind (which it is not), individuals would have to live infinitely long lives, and there cannot be any perturbations (e.g., health shocks), no matter how small. Thus the steady state is also not a candidate for the notion of a "desired" health level $H(t)^*$.

The starting point of the optimal trajectory for health is given by the initial condition $H(0) = H_0$. By definition, health eventually declines since end of life is determined by the minimum health level $H(T) = H_{\min}$, below which life is not sustainable, as illustrated by the shaded area.[37] Thus eventually the trajectory has to enter regions II or IV of the phase diagram so that health declines.[38] However, health investments increase with age (e.g., Grossman 1972a,b). For example, the intake of fruit and vegetables increases with age (Serdula et al. 2004; Pearson et al. 2005) and medical expenditures peak in the final phase of life (Zweifel, Felder and Meiers, 1999). This suggests that optimal solutions are best described by region II: declining health capital and increasing marginal cost of health investment (and hence increasing levels of health investment, see equations 14 and 22) with age.

3.2 Variation in socioeconomic status and health

Comparative dynamic analyses (e.g., Oniki, 1973) allow exploration of the effect of SES (wealth, wages [earnings], and education) and initial health on the marginal cost of health investment $\pi_I(t)$ and on health $H(t)$ by comparing the "perturbed" optimal trajectory with respect to the "unperturbed" (or original) optimal trajectory.

The effect of variation in an initial condition or other model parameter δZ (the "perturbation"), where $Z = \{A_0, w(t), E, H_0\}$, on the marginal cost of health investment and on health can be separated into two components[39]

$$\frac{\partial \pi_I(t)}{\partial Z} = \left.\frac{\partial \pi_I(t)}{\partial Z}\right|_T + \left.\frac{\partial \pi_I(t)}{\partial T}\right|_Z \frac{\partial T}{\partial Z}, \tag{25}$$

$$\frac{\partial H(t)}{\partial Z} = \left.\frac{\partial H(t)}{\partial Z}\right|_T + \left.\frac{\partial H(t)}{\partial T}\right|_Z \frac{\partial T}{\partial Z}, \tag{26}$$

where the first term on the RHS represents the response to variation in Z for constant T and the second term the additional response due to associated variation in T.

Apart from the earlier mentioned assumptions (see section 2.2), in the following I also assume that poorer individuals derive greater benefits from an additional increment

[37] The location of $H_{\min}$ is not known; for illustration it is shown to the right of the steady state.

[38] Note that since the null-clines shift over time the trajectory may enter regions II or IV by being overtaken by the null-clines. Thus the trajectory does not have to be in regions II or IV at all times but may reside for some time in regions I or III.

[39] Note that we can restart the problem at any time t, taking $A(t)$ and $H(t)$ as the new initial conditions. Thus the comparative dynamic results derived for, e.g., variation in initial wealth δA_0 and initial health δH_0 have greater validity, applying to variation in wealth $\delta A(t)$ and in health $\delta H(t)$ at any time $t \in [0, T)$.

of wealth than wealthier individuals, unhealthy individuals derive greater benefits from an additional increment in health than healthier individuals, individuals with shorter longevity benefit more from life extension, and poorer individuals benefit more from better health (since the stock of health and the stock of wealth are to some extent substitutable in financing consumption and leisure, e.g., Muurinen, 1982; Case and Deaton, 2005).[40]

3.2.1 Variation in initial assets

Fixed T First consider the case where longevity T is fixed. The comparative dynamic effect of initial wealth on the marginal cost of health investment, keeping length of life T fixed, is obtained by taking the derivate of (23) with respect to initial wealth A_0

$$\begin{aligned}\frac{\partial}{\partial t}\frac{\partial \pi_I(t)}{\partial A_0}\bigg|_T \quad &\approx \quad \frac{1}{q_A(0)^2}\frac{\partial U}{\partial H}e^{-(\beta-\delta)t} \times \frac{\partial q_A(0)}{\partial A_0}\bigg|_T + [d(t)+\delta] \times \frac{\partial \pi_I(t)}{\partial A_0}\bigg|_T \\ &- \left\{\frac{1}{q_A(0)}\frac{\partial^2 U}{\partial H^2}e^{-(\beta-\delta)t} + \frac{\partial^2 Y}{\partial H^2}\right\} \times \frac{\partial H(t)}{\partial A_0}\bigg|_T, \end{aligned} \tag{27}$$

where it is assumed that the indirect effects of wealth are small compared to the direct effects of wealth on health investment and health.[41] Likewise, for the health stock, the comparative dynamic effect of A_0, keeping length of life T fixed, is obtained by taking the derivative of equation (24) with respect to A_0

$$\frac{\partial}{\partial t}\frac{\partial H(t)}{\partial A_0}\bigg|_T \quad = \quad \frac{\alpha}{1-\alpha}\frac{I(t)^\alpha}{\pi_I(t)} \times \frac{\partial \pi_I(t)}{\partial A_0}\bigg|_T - d(t) \times \frac{\partial H(t)}{\partial A_0}\bigg|_T. \tag{28}$$

The LHS of Figure 3 shows the phase diagram for the motion paths of the variation of the marginal cost of health investment $\partial \pi_I(t)/\partial A_0|_T$ (y-axis) versus the variation of

[40] In other words, I assume diminishing returns to wealth, health, and longevity,

$$\begin{aligned}\frac{\partial q_A(t)}{\partial A(t)} &= \frac{\partial^2}{\partial A(t)^2}\int_t^{T^*} U(*)e^{-\beta s}ds < 0, & \frac{\partial q_H(t)}{\partial H(t)} &= \frac{\partial^2}{\partial H(t)^2}\int_t^{T^*} U(*)e^{-\beta s}ds < 0,\\ \frac{\partial \Im(T)}{\partial T} &= \frac{\partial^2}{\partial T^2}\int_t^{T^*} U(*)e^{-\beta s}ds < 0, & \frac{\partial q_A(t)}{\partial H(t)} &= \frac{\partial^2}{\partial H(t)\partial A(t)}\int_t^{T^*} U(*)e^{-\beta s}ds < 0,\end{aligned}$$

where T^* denotes optimal length of life and $U(*)$ denotes the maximized utility function. Intuitively, health is a resource and having more of it relaxes the dynamic constraint for health. But health also relaxes the dynamic constraint for wealth: being in better health reduces the need for health investment and health provides earnings. Thus health reduces the marginal value of health as well as wealth.

[41] In other words, the terms

$$\frac{1}{q_A(0)^2}\frac{\partial U}{\partial C}\frac{\frac{\partial^2 U}{\partial C\partial H}}{\frac{\partial^2 U}{\partial H^2}}e^{-(\beta-\delta)t} \times \frac{\partial q_A(0)}{\partial A_0}\bigg|_T, \qquad \text{and} \qquad \frac{1}{q_A(0)}\frac{\partial^2 U}{\partial H^2}\frac{(\frac{\partial^2 U}{\partial C\partial H})^2}{\frac{\partial^2 U}{\partial C^2}}e^{-(\beta-\delta)t} \times \frac{\partial H(t)}{\partial A_0}\bigg|_T,$$

are assumed to be small. The first term is the indirect effect of wealth on consumption and the effect that consumption in turn has on the marginal utility of health, and the second term is the indirect effect of wealth on health and the effect that health in turn has on consumption and consumption in turn on the marginal utility of health.

health $\partial H(t)/\partial A_0|_T$ (x-axis) with respect to initial wealth, for fixed T. By assumption $\partial q_A(0)/\partial A_0|_T < 0$, $\partial^2 U/\partial H^2 < 0$, and $\partial^2 Y/\partial H^2 < 0$, and we obtain the direction of motion (indicated by the block arrows) in the four dynamic regions of the phase diagram, defined by the null-clines. Since both initial health $H(0) = H_0$ and end-of-life health $H(T) = H_{\min}$ are fixed, it follows that $\partial H(0)/\partial A_0|_T = \partial H(T)/\partial A_0|_T = 0$. Thus, in the phase diagram all admissible paths begin and end at the vertical axis.

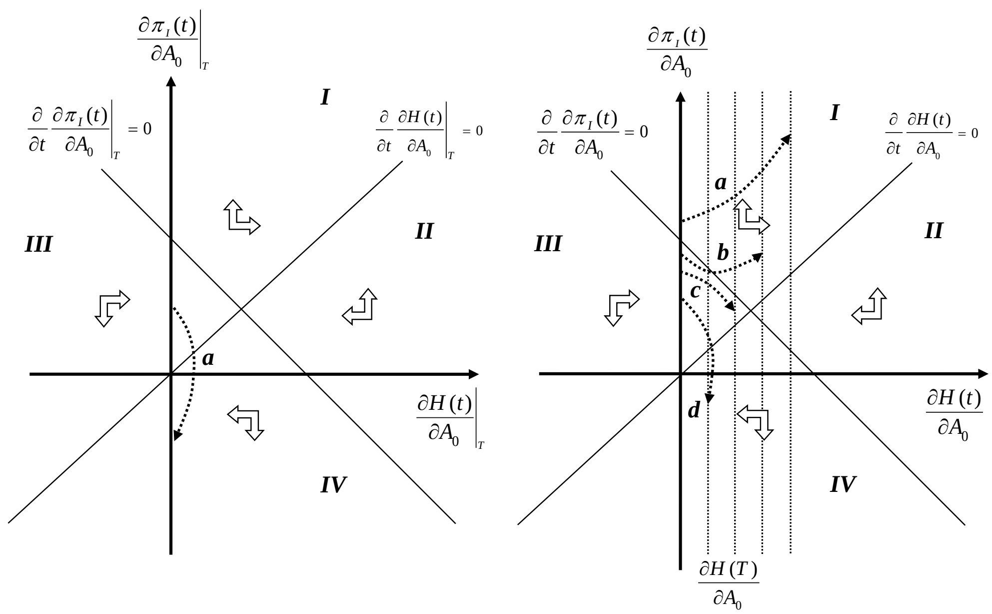

Figure 3: *The LHS shows the phase diagram of the perturbation due to variation in initial wealth δA_0, for fixed T. The RHS shows the same phase diagram allowing length of life T to be free. The four vertical dotted lines represent different values for the end point $\partial H(T)/\partial A_0$.*

Consider a path starting at the vertical axis, between the horizontal axis and the $(\partial/\partial t)\,(\partial \pi_I(t)/\partial A_0)|_T = 0$ null-cline. This path starts with $\partial \pi_I(0)/\partial A_0|_T > 0$, and could return to the vertical axis in finite time if it enters dynamic region IV at some point. This path satisfies all conditions, and an example trajectory (a) is shown for illustrative purposes.[42] Thus (see trajectory a) wealth increases the marginal cost of health investment $\partial \pi_I(t)/\partial A_0|_T > 0$ initially, but decreases it $\partial \pi_I(t)/\partial A_0|_T < 0$ eventually. In a model with a fixed life span T, health is higher at all ages, $\partial H(t)/\partial A_0|_T > 0 \quad \forall t$, except for $t = 0$ and $t = T$. Intuitively, because length of life T is fixed, any additional health investment (higher marginal cost) has to be balanced by reduced investment (lower marginal cost) to ensure that health reaches the minimum health level $H_{\min}$ over the unchanged horizon T.

[42] More complicated paths are possible (as null clines shift with time) that may temporarily enter regions I and/or II, but only those paths that start on the vertical axis above the horizontal axis and that end on the vertical axis below the horizontal axis are admissible, leading to broadly similar patterns.

Fixed length of life thus mutes the response to wealth.

Longevity Using the above results, it is possible to investigate whether life is extended as a result of additional wealth. Taking into account that in the optimum the condition $\Im(T) = 0$ has to be satisfied (see 12), we have

$$\frac{\partial T}{\partial A_0} = - \left.\frac{\partial \Im(T)}{\partial A_0}\right|_T \left\{ \left.\frac{\partial \Im(T)}{\partial T}\right|_{A_0} \right\}^{-1}, \tag{29}$$

and, using the expression for the Hamiltonian (9) we obtain (see Appendix section G)

$$\begin{aligned} \left.\frac{\partial \Im(T)}{\partial A_0}\right|_T = & \left.\frac{\partial q_A(0)}{\partial A_0}\right|_T e^{-rT} \left.\frac{\partial A(t)}{\partial t}\right|_{t=T} \\ & + \left\{ \left.\frac{\partial q_A(0)}{\partial A_0}\right|_T e^{-rT} \pi_I(T) + q_A(0) e^{-rT} \left.\frac{\partial \pi_I(T)}{\partial A_0}\right|_T \right\} \left.\frac{\partial H(t)}{\partial t}\right|_{t=T}, \end{aligned} \tag{30}$$

where I distinguish in notation between $\partial f(t)/\partial t|_{t=T}$, which represents the derivative with respect to time t at time $t = T$, and $\partial f(t)/\partial T|_{t=T}$, which represents variation with respect to the parameter T at time $t = T$.

Using (29), (30), and, by assumption, $\partial \Im(T)/\partial T|_{A_0} < 0$, wealth increases longevity $\partial T/\partial A_0 > 0$ if $\partial \Im(T)/\partial A_0|_T > 0$. In (30), both $\partial A(t)/\partial t|_{t=T}$ and $\partial H(t)/\partial t|_{t=T}$ are negative at the end of life since health approaches $H_{\min}$ from above, and assets decline in absence of a strong bequest motive. Since, by assumption, $\partial q_A(0)/\partial A_0|_T < 0$, a sufficient requirement for $\partial T/\partial A_0 > 0$ is $\partial \pi_I(T)/\partial A_0|_T \leq 0$. Indeed $\partial \pi_I(T)/\partial A_0|_T \leq 0$ (see the end point of trajectory a on the LHS of Figure 3), and thus life is extended $\partial T/\partial A_0 > 0$.

Free T Having established that wealthy individuals live longer, let's consider the more interesting case where length of life T is free. The coefficients in (27) and (28) for variation in length of life T, holding initial wealth A_0 constant, are identical to the coefficients for variation in initial wealth A_0, holding length of life T fixed. That is, we simply have to replace the partial derivatives with their total derivatives in (27) and (28) to obtain the total comparative dynamic effect for free T.[43] As a result, the phase diagram for free T is identical to that for fixed T (see the RHS of Figure 3). There is, however, one important difference: while all admissible paths start at the vertical axis, $\partial H(0)/\partial A_0 = 0$, they end to the right of the vertical axis, $\partial H(T)/\partial A_0 > 0$, and the end point $\partial H(T)/\partial A_0$ lies further to the right for greater life extension $\partial T/\partial A_0$ (indicated by the four dotted vertical lines in the figure; for proof see Appendix H). As the phase diagram shows, trajectories a, b, c and d, corresponding to four different levels of $\partial H(T)/\partial A_0$, are feasible. Moving from scenario d to c to b and finally to a, investment in health increases progressively and so does life extension (the end point $\partial H(T)/\partial A_0$ lies furthest to the right for trajectory a).

[43] I.e. replace $(\partial/\partial t)(\partial \pi_I(t)/\partial A_0)|_T$, $(\partial/\partial t)(\partial H(t)/\partial A_0)|_T$, $\partial q_A(0)/\partial A_0|_T$, $\partial \pi_I(t)/\partial A_0|_T$, and $\partial H(t)/\partial A_0|_T$ with $(\partial/\partial t)(\partial \pi_I(t)/\partial A_0)$, $(\partial/\partial t)(\partial H(t)/\partial A_0)$, $\partial q_A(0)/\partial A_0$, $\partial \pi_I(t)/\partial A_0$, and $\partial H(t)/\partial A_0$.

In all four scenarios individuals invest cumulatively more in health $\int_0^T [\partial \pi_I(t)/\partial A_0]\, dt > 0$ but they may invest less at certain ages (for example in scenario d individuals invest less late in life, compared to the unperturbed path). In a scenario where life extension is small, health investment increases less rapidly for wealthy individuals (investment is initially above and eventually beneath the unperturbed path, e.g., as in scenario d), whereas in a scenario where life extension is large, health investment increases more rapidly for wealthy individuals (health investment starts out higher and is increasingly higher at later ages, compared to the unperturbed path). Further, wealthy individuals are healthier at all ages (all paths lie to the right of the vertical axis).

Thus, additional wealth (positive δA_0) induces individuals to invest more in health. As a result, their health deteriorates slower and they live longer. This addresses the criticism of Case and Deaton (2005) that the Grossman model does not predict differences in the rate of aging between wealth groups.

3.2.2 Variation in permanent wages and education

Permanently higher wages and education operate in a similar manner to an increase in wealth δA_0, with some differences: (i) the wealth effect is muted by the increased opportunity cost of time, (ii) permanent wages w_E and education E also raise the production benefit of health, and (iii) education raises the efficiency of health investment. For details see Appendix I.

3.2.3 Variation in health

Fixed T The comparative dynamic effect of variation in initial health δH_0 is obtained by taking the derivative of (23) and (24) with respect to initial health H_0 and keeping first-order (direct) terms. The comparative dynamic relations follow simply from replacing A_0 with H_0 in (27) and (28). As a result, since $\partial q_A(0)/\partial H_0|_T < 0$ by assumption, the phase diagram for variation in initial health δH_0 (Figure 4), is similar to the phase diagram for variation in initial wealth δA_0 (LHS of Figure 3). However, an important difference with wealth is that any admissible path has to start at $\partial H(t)/\partial H_0|_{T,t=0}$, which for $t = 0$ is identical to 1. We don't know a-priori where $\partial H(t)/\partial H_0|_T = 1$ is located with respect to the steady state and Figure 4 shows two possible cases. Trajectories a, b and c are consistent with the begin and end conditions for the case where the starting point $\partial H(t)/\partial H_0|_T = 1$ is located to the left, and trajectories d and e for the case where $\partial H(t)/\partial H_0|_T = 1$ is located to the right, of the steady state (vertical dashed lines).

Trajectories a, b, and d represent an initial increase, followed by a subsequent decrease, in the marginal cost of health investment $\pi_I(t)$, with respect to the unperturbed path. Thus early in life, greater initial health may increase the demand for health investment, whereas later in life it decreases it. Trajectories c and e represent solutions where the marginal cost of health investment is lower at all times. Here better health reduces the demand for health investment at all times. Trajectories a through e all involve

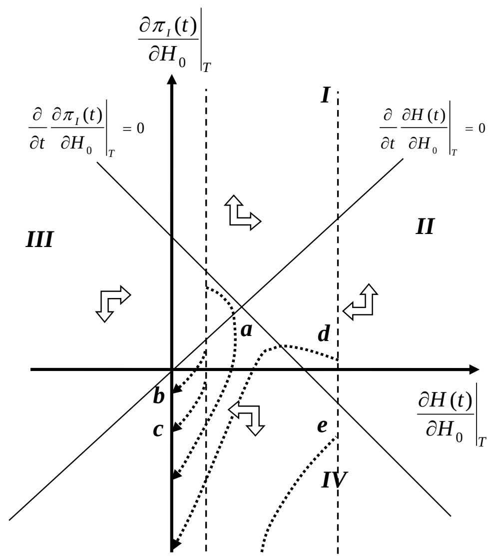

Figure 4: *Phase diagram of the perturbation due to variation in initial health δH_0, for fixed T.*

cumulatively lower health investment $\int_0^T [\partial\pi_I(t)/\partial H_0]\, dt < 0$ as initially higher health δH_0 requires less health investment over the life cycle in order for health to reach the minimum health level $H_{\min}$ within the same length of life T. However, because individuals start with better initial health $\delta H_0 > 0$, health is higher at all times (except for $t = T$).[44]

Longevity The effect of variation in initial health δH_0 on length of life T can be obtained by following the same steps as in section 3.2.1. The result is identical to replacing A_0 with H_0 in (29) and (30). By assumption, $\partial\Im(T)/\partial T|_{H_0} < 0$, and $\partial q_A(0)/\partial H_0|_T < 0$, so that life is extended if $\partial\pi_I(T)/\partial H_0|_T \leq 0$. Note that all admissible scenarios a through e, end with $\partial\pi_I(T)/\partial H_0|_T < 0$. Thus, length of life is extended $\partial T/\partial H_0 > 0$.[45]

Free T Having established that length of life is extended, $\partial T/\partial H_0 > 0$, now consider the case where T is free. Figure 5 presents the comparative dynamic results. The phase

[44]If health does not provide utility, the steady state is located at the origin. The admissible trajectory is then characterized by lower health investment at all times (similar to trajectories c and e).

[45]If, however, $\partial q_A(0)/\partial H_0|_T > 0$, the $(\partial/\partial t)(\partial\pi_I(t)/\partial H_0)|_T$ null-cline shifts downward, crossing the $(\partial/\partial t)(\partial H(t)/\partial H_0)|_T$ null cline to the left of and below the origin. In this case, admissible trajectories start with $\partial\pi_I(0)/\partial H_0|_T < 0$, and either end with $\partial\pi_I(T)/\partial H_0|_T > 0$, in which case life is not extended but reduced $\partial T/\partial H_0 < 0$ as a result of greater health, or end with $\partial\pi_I(T)/\partial H_0|_T < 0$, in which case it cannot be established that life is extended. I favor the scenario where poorer individuals derive greater benefits from health, i.e. $\partial q_A(0)/\partial H_0|_T < 0$. It is consistent with empirical evidence that worse childhood health is associated with shorter lives (e.g., Currie, 2009). Further the assumption has the natural intuitive interpretation that health and wealth are substitutable in financing consumption (see footnote 40).

diagram on the left shows feasible trajectories a through h for the case where the starting point $\partial H(t)/\partial H_0 = 1$ is located to the left of the steady state, and the phase diagram on the right shows feasible trajectories a through f for the case where the starting point $\partial H(t)/\partial H_0 = 1$ is located to the right of the steady state (dashed vertical lines in both phase diagrams). The initial condition $\partial H(t)/\partial H_0 = 1$ for $t = 0$, and the end-condition $\partial H(T)/\partial H_0 > 0$, imply that all admissible paths start and end to the right of the vertical axis. Analogous to the discussion in section 3.2.1 the end point $\partial H(T)/\partial H_0 > 0$, lies further to the right in the phase diagram, the greater life is extended $\partial T/\partial H_0$. Three example end values $\partial H(T)/\partial H_0$ are indicated by the dotted vertical lines in both diagrams.

While both phase diagrams are somewhat complicated, they show that for end points $\partial H(T)/\partial H_0$ that lie further to the right (greater degree of life extension), the deviation in the marginal cost of health investment $\partial \pi_I(t)/\partial H_0$ becomes more and more positive, with some scenarios even allowing for the possibility that healthy individuals invest more in health at every age. Whereas for end points $\partial H(T)/\partial H_0$ that lie more to the left (smaller degree of life extension), the deviation in the marginal cost of health investment $\partial \pi_I(t)/\partial H_0$ becomes more and more negative. These latter cases more closely resemble the fixed T case. Further, individuals with greater endowed health are healthier at all ages, $\partial H(t)/\partial H_0 > 0, \quad \forall t$ (all trajectories lie to the right of the vertical axis).

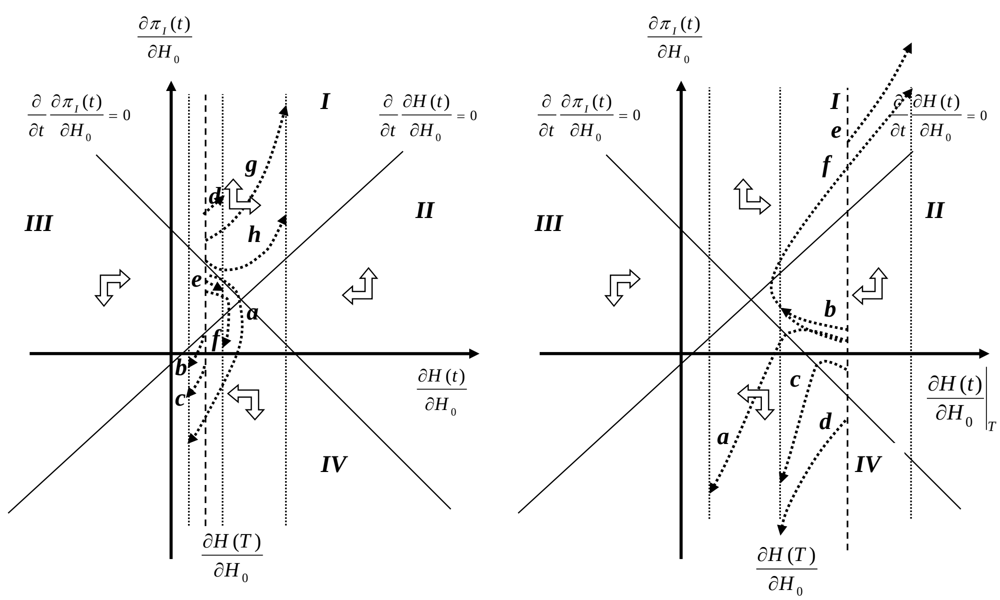

Figure 5: *Phase diagram of the perturbation due to variation in initial health δH_0, for free T, with starting point, $\partial H(t)/\partial H_0 = 1$, located to the left (LHS) and right (RHS) of the steady state.*

Thus, greater initial health potentially reduces the initial demand for health investment, $\partial I(0)/\partial H_0 < 0$. Because one can start the optimization problem at any age by redefining the initial conditions H_0 and A_0 for that age, this result holds for any age. Thus the theory can accommodate a negative relation between health and health investment. This addresses the criticism by Wagstaff (1986a) and Zweifel and Breyer (1997) that the Grossman model is unable to predict the observed strong negative relation between health and the demand for medical care. As the previous discussions and Figure 5 suggest, a negative relationship between health and health investment is more likely for a small degree of life extension afforded by additional health and for a small consumption benefit (steady state close to the origin; see footnote 44).

3.3 Structural equations

Empirical tests of health-production models have thus far been based on structural- and reduced-form equations derived under the assumption of a CRTS health-production process.[46] Because these structural- and reduced-form relations suffer from the issue of the indeterminacy of health investment and are often based on the incorrect condition (see sections 2.3 and 3.1.1), I derive in this section a structural relation for the DRTS health-production process presented in this paper. The purpose of this analysis is to show once more that the Grossman model is dynamic, that the demand for health investment is plausibly decreasing in health, and to offer some guidance for testing the model.

3.3.1 Simple functional forms

Discrete time lends itself better to empirical analyses. Appendix sections B and C present the discrete-time equivalent of the continuous-time formulation. In order to obtain expressions suitable for empirical testing we also have to assume functional forms for model functions and parameters that cannot be observed directly, such as the health-investment production process I_t and the biological aging rate d_t.

I specify the following constant relative risk aversion (CRRA) utility function:

$$U(C_t, H_t) = \frac{1}{1-\rho}\left(C_t^{\zeta} H_t^{1-\zeta}\right)^{1-\rho}, \tag{31}$$

where ζ ($0 \leq \zeta \leq 1$) is the relative "share" of consumption versus health and ρ ($\rho > 0$) the coefficient of relative risk aversion. This functional form can account for the observation that the marginal utility of consumption declines as health deteriorates (e.g., Finkelstein, Luttmer and Notowidigdo, 2013) which would rule out strongly separable functional forms, where the marginal utility of consumption is independent of health.

I make the usual assumption that sick time is a power law in health

$$s_t = \Omega\left(\frac{H_t}{H_{\min}}\right)^{-\gamma}, \tag{32}$$

where $\gamma > 0$ so that sick time decreases with health. This choice of functional form has the properties $\lim_{H_t \to \infty} s_t = 0$ and $\lim_{H_t \downarrow H_{\min}} s_t = \Omega$, where Ω is the total time budget as in (59; the discrete-time equivalent of 4).

[46]See footnote 10.

Using equation (61) we have:

$$\frac{\partial Y_t}{\partial H_t} \equiv \varphi_{H_t} = w_t \gamma \Omega H_{\min}^{\gamma} H_t^{-(1+\gamma)} \equiv w_t \Omega^* H_t^{-(1+\gamma)}. \tag{33}$$

Investment in health I_t and consumption C_t are assumed to be produced by combining own time and goods/services purchased in the market according to a Cobb-Douglas CRTS production function (see equations 105 and 109).

Following Grossman (1972a, 1972b) I assume that the more educated are more efficient consumers and producers of health investment

$$\mu_{I_t} = \mu_{I_0} e^{\rho_I E}, \tag{34}$$

where E is the level of education and ρ_I is a constant.

Further, I assume a Mincer-type wage equation in which the more educated and the more experienced earn higher wages (Mincer, 1974)

$$w_t = w_E e^{\rho_w E + \beta_x x_t - \beta_{x^2} x_t^2}, \tag{35}$$

where education E is expressed in years of schooling, x_t is years of working experience, and ρ_w, β_x and β_{x^2} are constants, assumed to be positive.

Lastly, following Wagstaff (1986a) and Cropper (1981) I assume the biological aging rate d_t to be of the form

$$d_t = d_\bullet e^{\beta_t t + \boldsymbol{\beta_\xi \xi_t}}, \tag{36}$$

where $d_\bullet \equiv d_0 e^{-\boldsymbol{\beta_\xi \xi_0}}$ and $\boldsymbol{\xi_t}$ is a vector of environmental variables (e.g., working and living conditions, hazardous environment, etc) that affect the biological aging rate. The vector $\boldsymbol{\xi_t}$ may include other exogenous variables that affect the biological aging rate, such as education (Muurinen, 1982).

3.3.2 Structural relation between health and health investment

A structural relation for the demand for health investment goods and services m_t (e.g., medical care) can be obtained from the first-order conditions for health investment (76) and for consumption (78), using the functional relations defined in the previous section 3.3.1 (see Appendix J for details)

$$b_{it}^1 m_{it}^{1-\alpha} - (1-\alpha) m_{it}^{1-\alpha} \widetilde{m_{it}} \quad = \quad b_{it}^2 H_{it}^{-1/\chi} + b_{it}^3 H_{it}^{-(1+\gamma)}, \tag{37}$$

where I have defined the following functions

$$b_{it}^1 \equiv \left[d_\bullet e^{\beta_t t_i + \boldsymbol{\beta_\xi \xi_{it}}} + \delta - (1-\alpha k_I)\widetilde{p_{m_{it}}} - \alpha k_I \widetilde{w_{it}} \right], \tag{38}$$

$$b_{it}^2 \equiv b_*^2 \left(q_{0i}^A\right)^{-1/\rho\chi} e^{\alpha \rho_I E_i} p_{m_{it}}^{-(1-\alpha k_I)} w_{it}^{-[k_C(1/\rho\chi - 1) + \alpha k_I]} p_{X_{it}}^{-(1-k_C)(1/\rho\chi - 1)} \left(\frac{1+\beta_i}{1+\delta}\right)^{-t_i/\rho\chi} \tag{39}$$

$$b_{it}^3 \equiv b_*^3 e^{\alpha \rho_I E_i} p_{m_{it}}^{-(1-\alpha k_I)} w_{it}^{1-\alpha k_I}, \tag{40}$$

and the following constants

$$
\begin{aligned}
b_*^2 &\equiv [(1-\zeta)\Lambda]^{1/\chi}\, \alpha k_I^{\alpha k_I}(1-k_I)^{1-\alpha k_I}\mu_{I_0}^{\alpha}\left[k_C^{k_C}(1-k_C)^{1-k_C}\,\mu_{C_t}\right]^{1/\rho\chi-1}, && (41)\\
b_*^3 &\equiv \alpha k_I^{\alpha k_I}(1-k_I)^{1-\alpha k_I}\mu_{I_0}^{\alpha}\Omega_*, && (42)\\
\Lambda &\equiv \zeta^{\frac{1-\rho}{\rho}}(\zeta/1-\zeta)^{1-\chi}, && (43)\\
\chi &\equiv (1+\rho\zeta-\zeta)/\rho, && (44)
\end{aligned}
$$

where the subscript i indexes the i^{th} individual, and where the notation $\widetilde{f_t}$ is used to denote the relative change $\widetilde{f_t} \equiv 1-\frac{f_{t-1}}{f_t}$ in a function f_t. Further, I have assumed small relative changes (much smaller than one) in the price of medical care $\widetilde{p_{m_{it}}}$, wages $\widetilde{w_{it}}$ and the efficiency of the health investment process $\widetilde{\mu_{I_{it}}}$ and, for simplicity, assumed a constant discount factor $\beta_t = \beta$ and constant rate of return to capital $\delta_t = \delta$.[47]

The structural form (37) of the first-order condition for health investment describes a dynamic relationship between the demand for health investment goods / services m_t (e.g., medical care), the relative change in the demand for health investment goods / services $\widetilde{m_t}$ and the health stock H_t. For slow changes in the demand for health investment goods / services with time (small $\widetilde{m_t}$), the demand for health investment goods / services m_t falls with the level of health H_t. This is further reflected in the elasticity of health investment goods / services with respect to health H_t, which, for small $\widetilde{m_t}$, is negative (and a function of the health stock H_t)

$$
\sigma_{m_t,H_t} = \frac{\partial m_t}{\partial H_t}\frac{H_t}{m_t} = -\frac{1}{1-\alpha}\left[\frac{\frac{1}{\chi}b_t^2 H_t^{-1/\chi} + (1+\gamma)b_t^3 H_t^{-(1+\gamma)}}{b_t^2 H_t^{-1/\chi} + b_t^3 H_t^{-(1+\gamma)}}\right], \tag{45}
$$

where I have suppressed the index i for the individual.[48] This addresses the criticism by Wagstaff (1986a) and Zweifel and Breyer (1997) that health-capital models are unable to predict a negative relation between health and health investment.[49] Appendix J provides additional discussion of these relations.

In practice, estimating (37) is not straightforward. Galama et al. (2012) estimate, after some simplifying assumptions, a linearized version of (37). They obtain a significantly negative relation between measures of medical care and measures of health. However, when controlling for the endogeneity of health, using childhood health and parental smoking during childhood as instruments, the coefficients become statistically

[47] A similar expression for time inputs $\tau_{I_{it}}$ can be obtained using (108). Further, one can substitute (35) for the wage rate w_{it} to obtain an expression in terms of years of schooling E_i and years of experience x_{it}.

[48] For the pure investment (PI) model ($\partial U/\partial H = 0$) we have $\sigma^{PI}_{m_t,H_t} = -(1+\gamma)/(1-\alpha)$, and for the pure consumption (PC) model ($\partial Y/\partial H = 0$) $\sigma^{PC}_{m_t,H_t} = -1/\chi(1-\alpha)$. Note, that in the pure investment model health does not provide utility, $\zeta = 1$ (see equation 31) and $b_{it}^2 = 0$, whereas in the pure consumption model health does not provide a production benefit, $\varphi_{H_{it}} = 0$ and $b_{it}^3 = 0$.

[49] Note, however, that the comparative dynamic analysis for health (section 3.2.3) also allows for the possibility that health increases the demand for health investment.

insignificant, bringing into question the robustness of prior estimates of the relation between health investment and health from the literature.

A more promising approach may be to numerically solve and estimate the model using dynamic programming techniques. Indeed, this appears to be the approach taken by the recent literature (see references to such dynamic programming efforts in footnote 10).

3.4 Numerical simulations

In this section I present simulations of the model with a DRTS health-production process and a simple step process. I first discuss the step process for fixed length of life (section 3.4.1). I then illustrate the properties of the model with numerical simulations accounting for endogenous length of life (section 3.4.2). The purpose of these analyses is to provide intuition into the dynamics of the model, and to corroborate results obtained from the dynamic analyses in sections 3.1 and 3.2, namely: 1) that the model can replicate the stylized fact that health investment increases with age, and 2) that length of life is finite, even for a constant deterioration rate (addressing the criticism of Case and Deaton, 2005).

3.4.1 Step process and fixed length of life

We start with the initial condition for health H_0. Initial consumption C_0 then follows from the first-order condition for consumption (78), which, for the assumed functional forms in section 3.3.1, can be written as

$$C_t \quad = \quad \left[\frac{q_0^A}{\zeta} \pi_{C_t} \frac{\prod_{j=1}^{t}(1+\beta_j)}{\prod_{j=1}^{t}(1+\delta_j)} \right]^{-1/\rho\chi} H_t^{\frac{\chi-1}{\chi}}, \tag{46}$$

where π_{C_t} is given by (110). Initial consumption C_0 is a function of initial health H_0, the price of goods and services p_{X_0}, the wage rate w_0 (the opportunity cost of not working) and the marginal value of initial wealth q_0^A. Next, the initial level of health investment I_0 follows from the initial marginal cost of health investment π_{I_0} (see expressions 72 and 106). The initial level of health investment I_0 is a function of the price of goods and services p_{m_0}, the wage rate w_0, education E, and the initial marginal value of wealth q_0^A and of health q_0^H. Thus, given exogenous education, prices and wage rates, the initial level of health investment I_0 and the initial level of consumption C_0 are functions of the as of yet undetermined initial values of q_0^A and q_0^H.

Health in the next period H_1 is determined by the dynamic equation (57). Assets in the next period A_1 follow from the initial condition for assets A_0 and the dynamic equation for assets (58). For the assumed functional forms in section 3.3.1 we have

$$A_{t+1} = (1+\delta_t)A_t + w_t \left[\Omega - \tau_{I_t}^* I_t - \tau_{C_t}^* C_t - s_t \right] - p_{X_t} X_t^* C_t - p_{m_t} m_t^* I_t, \tag{47}$$

where s_t, m_t^*, $\tau_{I_t}^*$, X_t^*, $\tau_{C_t}^*$ are defined in (32), (107), (108), (111) and (112).

Consumption C_1 follows from the first-order condition for consumption (46), health investment I_1 follows from the first-order condition for health investment (74), (75), or (76), which for the assumed functional forms in section 3.3.1 can be expressed as

$$\pi_{I_t} = \frac{1}{1-d_t}\left[\pi_{I_{t-1}}(1+\delta_t) - \frac{1-\zeta}{q_0^A} C_t^{1-\rho\chi} H_t^{\rho(\chi-1)-1} - w_t \Omega_* H_t^{-(1+\gamma)}\right]. \qquad (48)$$

Health H_2 and assets A_2 in the next period are determined by the dynamic equations (57) and (47) and so on. The solutions for consumption C_t, health investment I_t, health H_t, and wealth A_t, for every period t are functions of the initial marginal values of wealth q_0^A and health q_0^H. In the final period, the end conditions for the final level of health $H_T = H_{\min}$ and the final level of assets A_T determine q_0^A and q_0^H.[50] Section 3.4.2 details how this is implemented.

As pointed out by Ehrlich and Chuma (1990), length of life T is exogenous (fixed) in the absence of an additional (transversality) condition (12).[51] In discrete time this condition consists of maximizing the "indirect utility function" V_T (63) with respect to length of life T.

3.4.2 Simulations with endogenous longevity

In this section I simulate the model for a particular set of parameter values. The purpose of this exercise is to illustrate some properties of the model. Other parameter choices are possible and a full exploration of the model's properties would require exploring a wide range of parameter values. Ultimately one would like to estimate the model with panel data to test its ability to describe human behavior. This is beyond the scope of this paper.

Figure 6 shows the results of model simulations using the step process and equations presented in the above section 3.4.1. In the simulations I have used a period step size of one tenth of a year and assumed annual wages of the form

$$w_t = 10e^{\{1.31383\times 10^{-3}[70(t-20)-(t-20)^2]\}} \text{ \$ (thousands)}, \qquad (49)$$

starting at age $t = 20$ when the individual begins work life until she retires at a fixed retirement age $R = 65$. This Mincer-type wage equation starts with annual wages of \$ 10,000 per year at age $t = 20$ and peaks at \$ 50,000 per year at age $t = 55$ after which wages gradually decline till the age of retirement $R = 65$ (earnings are zero after retirement). In addition I use the following parameters: $\alpha = 0.5$, $\gamma = 10$ (sick time increases significantly only upon approaching end of life, i.e., as H_t approaches $H_{\min}$), $H_0 = 100$, $H_T \equiv H_{\min} = 15$, $A_0 = A_T = 0$ \$ (thousands) (no bequests), $\Omega = 0.1$ year (the total time available in a period equals the time step size), $k_I = k_C = 0$ (health investment

[50] Alternatively one could start with the final period $t = T - 1$ and use recursive back substitution.

[51] Grossman (1998) and Ried (1998) have argued that length of life is determined in an iterative process and that the condition that health at the end of life H_T equal the minimum health stock $H_{\min}$ is sufficient. However, as the preceding discussion shows, the end conditions A_T and H_T are needed to determine the co-state functions q_0^A and q_0^H for fixed T. Thus an additional condition is needed for optimal T.

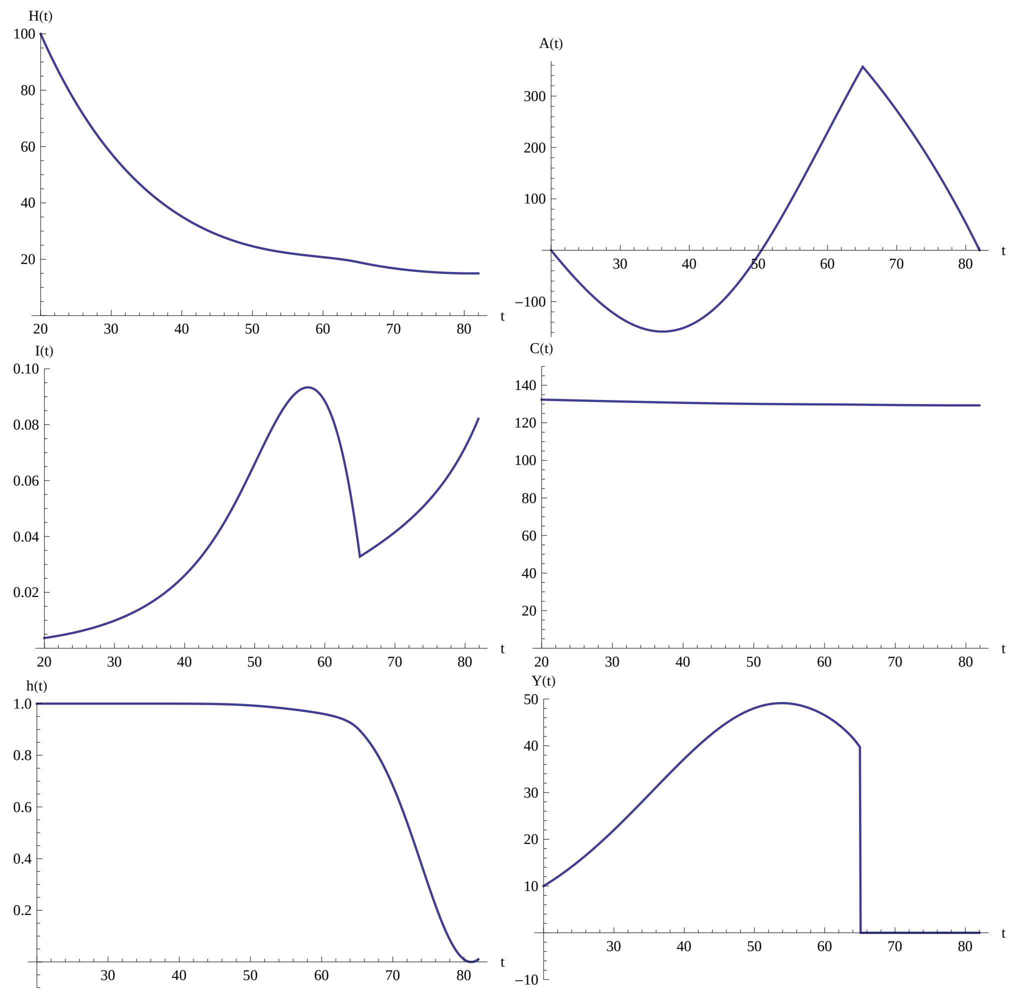

Figure 6: *Health ($H(t)$; top-left), assets ($A(t)$; \$ thousands; top-right), health investment ($I(t)$; center-left), consumption ($C(t)$; center-right), healthy time ($h(t)$; fraction of total time Ω; bottom-left), annual earnings ($Y(t)$; \$ thousands per year; bottom-right)*

I_t and consumption C_t consist of purchases in the market, not of own time inputs),[52] $p_{m_t} = 0.2$ \$ per medical good/service unit, $p_{X_t} = 0.2$ \$ per consumption good/service unit, $\mu_{I_t} = 0.01$, $\mu_{C_t} = 1$, $\rho = 0.8$, $\zeta = 0.95$ (high relative "weight" of consumption versus health in providing utility), a constant aging rate $d_t = d_0 = 0.06$ (per year), a constant return to capital $\delta_t = \delta_0 = 0.03$ (per year) and a constant subjective discount factor $\beta_t = \beta_0 = 0.03$ (per year).

[52]This simplification helps avoid corner solutions in which the time budget constraint is not satisfied. This is because for this choice healthy time $h_t = \Omega - s_t$ is always positive after retirement, even as s_t approaches Ω as H_t approaches $H_{\min}$. After retirement, income Y_t and time spend working τ_{w_t} are zero. Further, for $k_I = k_C = 0$ no time is devoted to health investment, $\tau_{I_t} = 0$, or to consumption, $\tau_{C_t} = 0$.

I start with the initial values for health H_0 and assets A_0 and employ the Nelder-Mead method (also called the downhill simplex or amoeba method; Nelder and Mead, 1965) to iteratively determine the initial marginal value of wealth q_0^A and of health q_0^H that satisfy the end conditions A_T and H_T. I use the usual values $\alpha_{NM} = 1$, $\gamma_{NM} = 2$, $\rho_{NM} = 0.5$ and $\sigma_{NM} = 0.5$ for the Nelder-Mead reflection, expansion, contraction and shrink coefficients, respectively. Optimal length of life is determined by maximizing the "indirect utility function" V_T (63) with respect to length of life T. Solutions for which health drops below $H_{\min}$ are excluded (see footnote 14). I find $T = 82.0$ years.

Health H_t (top-left panel of Figure 6) gradually declines with age t and life ends at age $T = 82.0$ years. Health deteriorates somewhat slower during the ages 50 to 65, coinciding with increased levels of health investment I_t (center-left panel of Figure 6). The demand for health investment consists of two components. The first is driven by the production benefit of health and follows a hump shaped pattern similar to the earnings profile Y_t (bottom-right panel of Figure 6). Health investment improves health, reduces sick time, and thereby improves earnings Y_t. The second is driven by the desire of individuals to be healthy (consumption benefit) and to extend life. This component gradually increases with age. Thus the simulation corroborates the dynamic analysis, suggesting that solutions are feasible in which health investment increases near the end of life.

At a price $p_{m_t} = 0.2$ \$ per medical good/services unit her expenditures on health investment goods/services $p_{m_t} m_t$ peak at about \$ 1,800 per year at around age 55. The fact that such humped-shape profiles are generally not observed in medical expenditure data sets, at least not as sizeable as the simulation shows, suggests that the production benefit of health (compared to the consumption benefit of health) may be smaller in real life than is simulated.

The individual's assets A_t (top-right panel in Figure 6) initially deplete till about age 50 as she borrows to fund her consumption C_t (center-right panel of Figure 6) and health investment I_t needs. She builds up savings between ages 50 and the age of retirement (65) and depletes these savings by end of life. Consumption is relatively constant with age. At a price $p_{X_t} = 0.2$ \$ per consumption good/service unit her expenditures on consumption goods/services $p_{X_t} X_t$ are about \$ 28,000 per year.

Healthy time h_t (bottom-left panel of Figure 6) starts to decline rapidly around the age of retirement. While some of this can be explained by a drop in health investment I_t at retirement, this is mostly due to the steep functional relation assumed between health H_t and sick time s_t (equation 32 for $\gamma = 10$).

The simulations further show that solutions are feasible for which the biological aging rate is constant, despite the common perception that the biological aging rate needs to increase with age in order to ensure that health falls with age and life is finite (e.g., Grossman, 1972a, 2000; Ehrlich and Chuma, 1990; Case and Deaton, 2005). This addresses the criticism that if the rate of biological deterioration is constant, people will life infinitely long lives (Case and Deaton, 2005).

4 Discussion and conclusions

The Grossman model is the canonical theory of the demand for health and health investment. This paper shows that much of the criticism directed at the model is not the result of a flawed model but of an unfortunate and unnecessary choice for the functional form (linear in investment) of the health-production process, and of an incorrect interpretation of the equilibrium condition for health. The linear relation between the production of health and health investment leads to an indeterminacy of health investment. Lacking ability to derive an expression for investment, the model essentially breaks down: without investment it is not possible to derive the paths of health, wealth, and consumption. In addition, I show that the "equilibrium" condition for health is instead an optimality condition for health investment. This condition should be used to determine health investment and the dynamic equation for health should be used to determine health. Employing a diminishing returns to scale production function, adopting the reinterpretation of the equilibrium condition, modeling endogenous longevity, and analyzing the resulting dynamics, the theory is capable of addressing the major criticisms leveled at the Grossman model.

This paper also presents the first detailed comparative dynamic analyses of the Grossman model. These show that high SES individuals invest more in health, and as a result their health declines more slowly and they live longer. As the health of lower SES individuals deteriorates faster, they start to invest more in health (if the demand for investment is higher for those in poor health). This, and mortality selection, as those in poor health live shorter lives, leading to an apparently healthier lower SES population in old age, may explain the narrowing of the SES-health gradient around ages 50 to 60. Thus the model can reproduce the essential features of the SES-health gradient. Comparative dynamic analyses also highlight an important role for life extension. For example, health investment is higher and increases less rapidly with age for higher SES individuals if SES enables moderate life extension (flatter investment profile) and more rapidly with age if SES enables substantial life extension (steeper investment profile).

These analyses support the Grossman model's canonical status, showing that it provides a remarkably successful foundation for understanding decisions regarding health. Yet, further development is required in at least three directions. First, medical care explains only a relatively small part of the SES-health gradient, suggesting a role for a more comprehensive theory of health production that includes additional health behaviors. Second, a childhood phase needs to be incorporated, in recognition of the importance of childhood endowments and investments in the determination of later-life socioeconomic and health outcomes. Third, a unified theory of joint investment in skill (or human) capital and in health capital could provide a basis for a theory of the relationship between education and health.

References

Arrow, K. J. (1968), "Optimal capital policy with irreversible investment". In: *Value, capital, and growth: Papers in Honour of Sir John Hicks*, J.N. Wolfe (editor), Edinburgh: Edinburgh University Press, pp. 1-20.

Bolin, K., Jacobson, L. and Lindgren, B. (2001), "The family as the producer of health - When spouses are Nash bargainers", Journal of Health Economics, 20: pp. 349-362.

Bolin, K., Jacobson, L. and Lindgren, B. (2002a), "Employer investments in employee health – implications for the family as health producer", Journal of Health Economics, 21: pp. 563-583.

Bolin, K., Jacobson, L. and Lindgren, B. (2002b), "The family as the health producer – When spouses act strategically", Journal of Health Economics, 21: pp. 475-495.

Bolin, K., Lindgren, B., Lindstrom, M., and Nystedt, P. (2003), "Investments in social capital – implications of social interactions for the production of health", Social Science and Medicine, 56(12): pp. 2379-2390.

Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., and Pan, Y. (2014), "Early childhood investments substantially boost adult health", Science, 343(6178): pp. 1478-1485.

Caputo, M.R. (2005), *Foundations of Dynamic Economic Analysis*, Cambridge University Press.

Case, A. and Deaton, A. (2005), "Broken down by work and sex: how our health declines", in: David A. Wise (ed.), *Analyses in the Economics of Aging*, The University of Chicago Press, Chicago, pp. 185-212.

Contoyannis, P. and Rice, N. (2001), "The impact of health on wages: evidence from the British Household Panel Survey", Empirical Economics, 26: pp. 599-622.

Contoyannis, P., Jones, A.M. and Rice, N. (2004), "The dynamics of health in the British Household Panel Survey", Journal of Applied Econometrics, 19(4): pp. 473-503.

Cropper, M.L. (1977), "Health, investment in health, and occupational choice", Journal of Political Economy, 85: pp. 1273-1294.

Cropper, M.L. (1981), "Measuring the benefits from reduced morbidity", The American Economic Review, *Papers and Proceedings of the Ninety-Third Annual Meeting of the American Economic Association*, 71(2): pp. 235-240.

Currie, J. and Madrian, B.C. (1999), "Health, health insurance and the labour market", In: Ashenfelter O, Card, D. (eds.), *Handbook of Labour Economics*, 3: pp. 3309-3415, Amsterdam: Elsevier Science B.V.

Currie, J. (2009). "Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development", Journal of Economic Literature 47(1): pp. 87-122.

Currie, J., and Almond, D. (2011). "Human capital development before age five", *Handbook of labor economics*, 4: pp. 1314-1486.

Cutler, D.M., Lleras-Muney, A. and Vogl, T. (2011), "Socioeconomic status and health: dimensions and mechanisms", Oxford Handbook of Health Economics.

Dustmann, C., and Windmeijer, F. (2000), "Wages and the demand for health – A life cycle analysis", IZA Discussion Paper 171, Germany.

Eisenring, C. (1999), "Comparative dynamics in a health investment model", Journal of Health Economics, 18(5): pp. 655-660.

Ehrlich, I. and Chuma, H. (1990), "A Model of the demand for longevity and the value of life extension", Journal of Political Economy, 98(4): pp. 761-782.

Ehrlich, I. (2000), "Uncertain lifetime, life protection, and the value of life saving", Journal of Health Economics, 19(3): pp. 341-367.

Ehrlich, I. and Yin, Y. (2005), "Explaining diversities in age-specific life expectancies and values of life saving: a numerical analysis", The Journal of Risk and Uncertainty, 31(2): pp. 129-162.

Erbsland, M., Ried, W. and Ulrich, V. (2002), "Health, health care, and the environment: econometric evidence from German micro data", In: *Econometric Analysis of Health Data*, John Wiley & Sons, Ltd, pp. 25-36.

Finkelstein, A., Luttmer, E. F. P., Notowidigdo, M. J. (2013), "What good is wealth without health? The effect of health on the marginal utility of consumption", Journal of the European Economic Association, 11: pp. 221-258.

Fonseca, R., Michaud, P.-C, Kapteyn, A., and Galama, T. J. (2013), "Accounting for rise of health spending and longevity", IZA Discussion Paper, 7622.

Galama, T.J. and Kapteyn, A. (2011), “Grossman’s missing health threshold”, Journal of Health Economics, 30(5): pp. 1044-1056.

Galama, T. J., Hullegie, P., Meijer, E., and Outcault, S. (2012), “Is there empirical evidence for decreasing returns to scale in a health capital model?”, Health Economics, 21(9): pp. 1080-1100.

Galama, T.J., Kapteyn, A., Fonseca, F., Michaud, P.C. (2013), “A health production model with endogenous retirement”, Health Economics, 22: pp. 883-902.

Gerdtham, U. G., and Johannesson, M. (1999), “New estimates of the demand for health: Results based on a categorical health measure and Swedish micro data.” Social Science and Medicine, 49: pp. 1325-1332.

Gerdtham, U. G., Johannesson, M., Lundberg, L., and Isacson, D. (1999), “The demand for health: Results from new measures of health capital.” European Journal of Political Economy, 15: pp. 501-521.

Gilleskie, D. (1998), “A dynamic stochastic model of medical care use and work absence.” Econometrica, 66(1): pp. 1-45.

Gilleskie, D. (2010), “Work absences and doctor visits during an illness episode: The differential role of preferences, production, and policies among men and women.” Journal of Econometrics, 156(1): pp. 148-163.

Grossman, M. (1972a), “The Demand for health - A theoretical and empirical investigation”, New York: National Bureau of Economic Research.

Grossman, M. (1972b), “On the concept of health capital and the demand for health”, Journal of Political Economy, 80(2): pp. 223-255.

Grossman, M. (1998), “On optimal length of life”, Journal of Health Economics, 17: pp. 499-509.

Grossman, M. (2000), “The human capital model” In: Culyer and J. P. Newhouse (Eds.), *Handbook of Health Economics*, 1, pp. 347-408. Amsterdam, The Netherlands: Elsevier Science.

Heckman, J.J. (2007), “The economics, technology, and neuroscience of human capability formation”, Proceedings of the national Academy of Sciences, 104(33): pp. 13250-13255.

Hugonnier, J., Pelgrin, F., and St-Amour, P. (2013), “Health and (other) asset holdings.”

Review of Economic Studies, 80(2): pp. 663-710.

Jorgenson, D. W. (1963), "Capital theory and investment behavior." The American Economic Review, pp. 247-259.

Jorgenson, D. (1967), "The theory of investment behavior." In *Determinants of investment behavior*, pp. 129-188, NBER.

Kaestner, R. (2013), "The Grossman model after 40 years: a reply to Peter Zweifel." European Journal of Health Economics, 14: pp. 357-360.

Kaya, C. Y., Lucas, S. K., and Simakov, S. T. (2004), "Computations for bang-bang constrained optimal control using a mathematical programming formulation", Optimal Control Applications and Methods, 25(6): pp. 295-308.

Khwaja, A. (2010), "Estimating willingness to pay for Medicare using a dynamic life-cycle model of demand for health insurance", Journal of Econometrics, 156(1): pp. 130-147.

Kirk, D. (1970), *Optimal control theory: an introduction*, Prentice-Hall.

Laporte, A. (2014), "Should the Grossman model retain its iconic status in Health Economics?", Canadian Centre for Health Economics Working Paper 2014-04.

Leu, R.E. and Doppman, R.J. (1986), "Gesundheitszustandsmessung und nachfrage nach gesundheitsleistungen", In: Wille, E. (Hrsg.), *Informations- und planungsprobleme in offentlichen aufgabenbereichen*, Frankfurt am Main/Bern/New York: Lang 1986, pp. 1-90.

Leu, R.E. and Gerfin, M. (1992), "Die nachfrage nach gesundheit - ein empirisher test des Grossman-modells (Demand for health - an empirical test of the Grossman model)", In: Oberender, P. (Ed.), *Steuerungsprobleme im gesundheitswesen*, Baden-Baden: Nomos, pp. 61-78.

Liljas, B. (1998), "The demand for health with uncertainty and insurance", Journal of Health Economics, 17(2): pp. 153-170.

Mincer, J.A. (1974), "Schooling, experience, and earnings", Columbia University Press.

Muurinen, J-M. (1982), "Demand for health: a generalized Grossman model", Journal of Health Economics, 1: pp. 5-28.

Muurinen, J-M., and Le Grand, J. (1985), "The Economic analysis of inequalities in health", Social Science and Medicine, 20(10): pp. 1029-1035.

Nelder, J.A. and Mead, R. (1965), "A simplex method for function minimization", Computer Journal, 7: pp. 308–313.

Nocera, S. and Zweifel, P. (1998), "The demand for health: an empirical test of the Grossman model using panel data", In: Zweifel, P. (Ed.), *Health, the medical profession and regulation*, Kluwer academic publishers, Boston/Dordrecht/London, pp. 35-49.

Oniki, H. (1973), "Comparative dynamics (sensitivity analysis) in optimal control theory", Journal of Economic Theory, 6: pp. 265–283.

Pearson, T., Russell, J., Campbell, M. J., and Barker, M. E. (2005), "Do 'food deserts' influence fruit and vegetable consumption?" A cross-sectional study." *Appetite*, 45(2), 195-197.

Ried, W. (1996), "Willingness to pay and cost of illness for changes in health capital depreciation", Health Economics, 5: pp. 447-468.

Ried, W. (1998), "Comparative dynamic analysis of the full Grossman model", Journal of Health Economics, 17: pp. 383-425.

Scholz, J. K., and Seshadri, A. (2012), "The interplay of wealth, retirement decisions. Policy and economic shocks." Michigan Retirement Research Center Research Paper No. WP, 2012-271.

Seierstad, A. and Sydsaeter, K. (1977), "Sufficient conditions in optimal control theory", International Economic Review, 18(2): pp. 367-391.

Seierstad, A. and K. Sydsaeter (1987), "Optimal control theory with economic applications", In: Bliss, C.J. and Intriligator, M.D. (Eds), *Advanced Textbooks in Economics*, Volume 24, Elsevier, North Holland.

Serdula, M. K., Gillespie, C., Kettel-Khan, L., Farris, R., Seymour, J., and Denny, C. (2004), "Trends in fruit and vegetable consumption among adults in the United States: behavioral risk factor surveillance system, 1994-2000." American Journal of Public Health, 94(6): pp. 1014-1018.

Strulik, H. (2014), "A closed-form solution for the health capital model", CEGE Discussion Papers, 222.

Sydsaeter, K., Strom, A. and Berck, P. (2005), "Economists' mathematical manual", ISBN-10 3-540-26088-9, 4th ed., Springer Berlin, Heidelberg, New York.

Usher, D. (1975), "Comments on the correlation between health and schooling", In: N.E. Terleckyj, ed., *Household Production and Consumption* (Columbia University Press for the National Bureau of Economic Research, New York), pp. 212-220.

Van de Ven, W. and van der Gaag, J. (1982), "Health as an unobservable: A MIMIC-model of the demand for health care", Journal of Health Economics, 1: pp. 157-183.

Van Doorslaer, E.K. (1987), "Health, knowledge and the demand for medical care", Assen, 171, ISBN 90-232-2335-7.

Van Doorslaer, E.K., Van Kippersluis, H., O'Donnell, O., Van Ourti, T. (2008), "Socioeconomic differences in health over the life cycle: evidence and explanations", Netspar Panel paper prepared for the Netspar Panel Day 16 October, 2008, Tilburg.

Wagstaff, A. (1986a), "The demand for health: some new empirical evidence", Journal of Health Economics, 5: pp. 195-233.

Wagstaff, A. (1986b), "The demand for health: theory and applications", Journal of Epidemiology and Community Health, 40: pp. 1-11

Wagstaff, A. (1993), "The demand for health: an empirical reformulation of the Grossman model", Health Economics 2: pp. 189-198.

Yogo, M. (2009), "Portfolio choice in retirement: Health risk and the demand for annuities, housing, and risky assets." PARC Working Papers, Population Aging Research Center.

Zweifel, P. and Breyer, F. (1997), *Health Economics*, Oxford University Press, New York.

Zweifel, P., Felder, S., and Meiers, M. (1999), "Ageing of population and health care expenditures: a red herring? ", Health Economics, 8(6): pp. 485-496.

Zweifel, P. (2012), "The Grossman model after 40 years." European Journal of Health Economics, 13: pp. 677-682.

A First-order conditions: continuous time

The first-order necessary conditions for the optimal control problem, consisting of maximizing the objective function (1) with respect to the control functions $X(t)$, $\tau_C(t)$, $m(t)$, and $\tau_I(t)$, subject to the constraints (2, 3 and 4) and begin and end conditions, follow from Pontryagin's maximum principle (e.g., Caputo, 2005). The Hamiltonian is given by (9). For the co-state variable $q_A(t)$ associated with assets we have

$$\begin{aligned} \frac{\partial q_A}{\partial t} &= -\frac{\partial \Im}{\partial A} = -q_A(t)\delta \Leftrightarrow \\ q_A(t) &= q_A(0)e^{-\delta t}. \end{aligned} \tag{50}$$

The co-state variable $q_H(t)$ associated with health capital follows from

$$\begin{aligned} \frac{\partial q_H}{\partial t} &= -\frac{\partial \Im}{\partial H} \\ &= q_H(t)d(t) - \frac{\partial U}{\partial H}e^{-\beta t} - q_A(t)\frac{\partial Y}{\partial H}. \end{aligned} \tag{51}$$

The first-order condition for investment in health (13) follows from optimizing with respect to health investment goods and services $m(t)$ and time inputs $\tau_I(t)$:

$$\frac{\partial \Im}{\partial m} = 0 \Leftrightarrow \quad q_H(t)\frac{\alpha}{I(t)^{1-\alpha}}\frac{\partial I}{\partial m} - q_A(t)p_m(t) = 0, \tag{52}$$

$$\frac{\partial \Im}{\partial \tau_I} = 0 \Leftrightarrow \quad q_H(t)\frac{\alpha}{I(t)^{1-\alpha}}\frac{\partial I}{\partial \tau_I} - q_A(t)w(t) = 0. \tag{53}$$

The first-order condition for consumption (19) follows from optimizing with respect to consumption goods and services $X(t)$ and time inputs $\tau_C(t)$:

$$\frac{\partial \Im}{\partial X} = 0 \Leftrightarrow \quad \frac{\partial U}{\partial C}\frac{\partial C}{\partial X}e^{-\beta t} - q_A(t)p_X(t) = 0, \tag{54}$$

$$\frac{\partial \Im}{\partial \tau_C} = 0 \Leftrightarrow \quad \frac{\partial U}{\partial C}\frac{\partial C}{\partial \tau_C}e^{-\beta t} - q_A(t)w(t) = 0. \tag{55}$$

B Model formulation: discrete time

Using discrete time optimal control (e.g., Sydsaeter, Strom and Berck, 2005) the problem can be stated as follows.[53] Individuals maximize the life-time utility function

$$\sum_{t=0}^{T-1} \frac{U(C_t, H_t)}{\prod_{k=1}^{t}(1+\beta_k)}, \tag{56}$$

[53] Notation for the discrete time version of the model follows that of the continuous time version, with time replaced by an index. For example, $A(t)$ becomes A_t.

where individuals live for T (endogenous) periods, β_k is a subjective discount factor and individuals derive utility $U(C_t, H_t)$ from consumption C_t and from health H_t.

The objective function (56) is maximized subject to the dynamic constraints:

$$
\begin{aligned}
H_{t+1} &= f(I_t) + (1 - d_t) H_t, && (57)\\
A_{t+1} &= (1 + \delta_t) A_t + Y(H_t) - p_{X_t} X_t - p_{m_t} m_t, && (58)
\end{aligned}
$$

the total time budget Ω_t

$$\Omega_t = \tau_{w_t} + \tau_{I_t} + \tau_{C_t} + s(H_t), \tag{59}$$

and initial and end conditions: H_0, H_T, A_0 and A_T are given. Individuals live for T periods and die at the end of period $T - 1$. Life cannot be sustained below a minimum health level $H_{\min}$, and the individual dies when $H_T = H_{\min}$.

For simplicity I assume

$$f(I_t) = I_t^{\alpha}, \tag{60}$$

where $0 < \alpha < 1$ (DRTS).

Income $Y(H_t)$ equals the wage rate w_t times the amount of time spent working τ_{w_t},

$$Y(H_t) = w_t \left[\Omega_t - \tau_{I_t} - \tau_{C_t} - s(H_t)\right]. \tag{61}$$

Thus, we have the following optimal control problem: the objective function (56) is maximized with respect to the control functions X_t, τ_{C_t}, m_t and τ_{I_t} and subject to the constraints (57, 58 and 59). The Hamiltonian of this problem is:

$$\Im_t = \frac{U(C_t, H_t)}{\prod_{k=1}^{t}(1 + \beta_k)} + q_t^H H_{t+1} + q_t^A A_{t+1}, \quad t = 0, \ldots T - 1 \tag{62}$$

where q_t^H is the adjoint variable associated with the dynamic equation (57) for the state variable health H_t and q_t^A is the adjoint variable associated with the dynamic equation (58) for the state variable assets A_t.

The optimal control problem presented so far is formulated for a fixed length of life T (see, e.g., Seierstad and Sydsaeter, 1977, 1987; Kirk, 1970; see also section 3.4.1). To allow for differential longevity one needs to optimize over all possible lengths of life T. One way to achieve this is by first solving the optimal control problem conditional on length of life T (i.e., for a fixed exogenous T), inserting the optimal solutions for consumption C_t^* and health H_t^* (denoted by $*$) into the "indirect utility function"

$$V_T \equiv \sum_{t=0}^{T-1} \frac{U(C_t^*, H_t^*)}{\prod_{k=1}^{t}(1 + \beta_k)}, \tag{63}$$

and maximizing V_T with respect to T.[54]

[54] This is mathematically equivalent to the condition utilized by Ehrlich and Chuma (1990) (in continuous time) that the Hamiltonian equal zero at the end of life $\Im(T) = 0$ (see 12).

C First-order conditions: discrete time

Associated with the Hamiltonian (equation 62) we have the following conditions:

$$
\begin{aligned}
q^A_{t-1} &= \frac{\partial \Im_t}{\partial A_t} \Rightarrow \\
q^A_{t-1} &= (1+\delta_t) q^A_t \Leftrightarrow \\
q^A_t &= \frac{q^A_0}{\prod_{k=1}^{t}(1+\delta_k)}, && (64) \\
q^H_{t-1} &= \frac{\partial \Im_t}{\partial H_t} \Rightarrow \\
q^H_{t-1} &= q^H_t (1-d_t) + \frac{\partial U(C_t,H_t)/\partial H_t}{\prod_{k=1}^{t}(1+\beta_k)} + q^A_0 \frac{\partial Y(H_t)/\partial H_t}{\prod_{k=1}^{t}(1+\delta_k)} \Leftrightarrow && (65) \\
q^H_t &= -\sum_{i=1}^{t} \left[\frac{\partial U(C_i,H_i)/\partial H_i}{\prod_{j=1}^{i}(1+\beta_j)} + q^A_0 \frac{\partial Y(H_i)/\partial H_i}{\prod_{j=1}^{i}(1+\delta_j)} \right] \frac{1}{\prod_{k=i}^{t}(1-d_k)} \\
&+ \frac{q^H_0}{\prod_{k=1}^{t}(1-d_k)} && (66) \\
&= \sum_{i=t+1}^{T} \left[\frac{\partial U(C_i,H_i)/\partial H_i}{\prod_{j=1}^{i}(1+\beta_j)} + q^A_0 \frac{\partial Y(H_i)/\partial H_i}{\prod_{j=1}^{i}(1+\delta_j)} \right] \prod_{k=t+1}^{i-1} (1-d_k) \\
&+ q^H_T \prod_{k=t+1}^{T} (1-d_k) && (67) \\
\frac{\partial \Im_t}{\partial X_t} &= 0 \Rightarrow \\
\frac{\partial U(C_t,H_t)}{\partial C_t} &= q^A_0 \frac{p_{X_t}}{\partial C_t/\partial X_t} \frac{\prod_{j=1}^{t}(1+\beta_j)}{\prod_{j=1}^{t}(1+\delta_j)} \\
&\equiv q^A_0 \pi_{C_t} \frac{\prod_{j=1}^{t}(1+\beta_j)}{\prod_{j=1}^{t}(1+\delta_j)}, && (68) \\
\frac{\partial \Im_t}{\partial \tau_{C_t}} &= 0 \Rightarrow \\
\frac{\partial U(C_t,H_t)}{\partial C_t} &= q^A_0 \frac{w_t}{\partial C_t/\partial \tau_{C_t}} \frac{\prod_{j=1}^{t}(1+\beta_j)}{\prod_{j=1}^{t}(1+\delta_j)} \\
&\equiv q^A_0 \pi_{C_t} \frac{\prod_{j=1}^{t}(1+\beta_j)}{\prod_{j=1}^{t}(1+\delta_j)}, && (69)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \Im_t}{\partial m_t} &= 0 \Rightarrow \\
q_t^H &= q_0^A \left\{ \frac{p_{m_t} I_t^{1-\alpha}}{\alpha[\partial I_t/\partial m_t]} \right\} \frac{1}{\prod_{j=1}^{t}(1+\delta_j)} \\
&\equiv q_0^A \, \pi_{I_t} \frac{1}{\prod_{j=1}^{t}(1+\delta_j)}, \qquad (70) \\
\frac{\partial \Im_t}{\partial \tau_{I_t}} &= 0 \Rightarrow \\
q_t^H &= q_0^A \left\{ \frac{w_t I_t^{1-\alpha}}{\alpha[\partial I_t/\partial \tau_{I_t}]} \right\} \frac{1}{\prod_{j=1}^{t}(1+\delta_j)} \\
&\equiv q_0^A \, \pi_{I_t} \frac{1}{\prod_{j=1}^{t}(1+\delta_j)}, \qquad (71)
\end{aligned}
$$

where I have used the following definitions

$$
\sum_{k}^{k-1}(\bullet) \equiv 0,
$$

$$
\prod_{k}^{k-1}(\bullet) \equiv 1.
$$

Combining (70) and (71) we obtain the first-order condition for health investment

$$
\pi_{I_t} = \frac{q_t^H}{q_t^A}, \qquad (72)
$$

where π_{I_t} is the marginal cost of health investment I_t

$$
\pi_{I_t} \equiv \frac{p_{m_t} I_t^{1-\alpha}}{\alpha[\partial I_t/\partial m_t]} = \frac{w_t I_t^{1-\alpha}}{\alpha[\partial I_t/\partial \tau_{I_t}]}. \qquad (73)
$$

Using (64), (66), and (72), we obtain alternative expressions for the first-order condition for health investment I_t, with the initial point $t=0$ as point of reference

$$
\begin{aligned}
\frac{\pi_{I_t}}{\prod_{k=1}^{t}(1+\delta_k)} &= -\sum_{i=1}^{t} \left[\frac{\partial U(C_i,H_i)/\partial H_i}{q_0^A \prod_{j=1}^{i}(1+\beta_j)} + \frac{\partial Y(H_i)/\partial H_i}{\prod_{j=1}^{i}(1+\delta_j)} \right] \frac{1}{\prod_{k=i}^{t}(1-d_k)} \\
&+ \frac{\pi_{I_0}}{\prod_{k=1}^{t}(1-d_k)}, \qquad (74)
\end{aligned}
$$

or with the final period $T-1$ as point of reference

$$
\begin{aligned}
\frac{\pi_{I_t}}{\prod_{k=1}^{t}(1+\delta_k)} &= \sum_{i=t+1}^{T-1} \left[\frac{\partial U(C_i,H_i)/\partial H_i}{q_0^A \prod_{j=1}^{i}(1+\beta_j)} + \frac{\partial Y(H_i)/\partial H_i}{\prod_{j=1}^{i}(1+\delta_j)} \right] \prod_{k=t+1}^{i-1}(1-d_k) \\
&+ \frac{\pi_{I_{T-1}} \prod_{k=t+1}^{T-1}(1-d_k)}{\prod_{k=1}^{T-1}(1+\delta_k)}. \qquad (75)
\end{aligned}
$$

Using either (74) or (75) and taking the difference between period t and $t-1$ we obtain the following expression

$$\sigma_{H_t} \quad = \quad \frac{1}{q_0^A}\frac{\partial U(C_t,H_t)}{\partial H_t}\frac{\prod_{j=1}^{t}(1+\delta_j)}{\prod_{j=1}^{t}(1+\beta_j)}+\frac{\partial Y(H_t)}{\partial H_t}, \tag{76}$$

where σ_{H_t} is the user cost of health capital at the margin

$$\sigma_{H_t} \quad \equiv \quad \pi_{I_t}(d_t+\delta_t)-\Delta\pi_{I_t}(1+\delta_t), \tag{77}$$

and $\Delta\pi_{I_t} \equiv \pi_{I_t} - \pi_{I_{t-1}}$.

Combining (68) and (69) we obtain the first-order condition for consumption C_t

$$\frac{\partial U(C_t,H_t)}{\partial C_t} \quad = \quad q_0^A \pi_{C_t}\frac{\prod_{j=1}^{t}(1+\beta_j)}{\prod_{j=1}^{t}(1+\delta_j)}, \tag{78}$$

where π_{C_t} is the marginal cost of consumption C_t

$$\pi_{C_t} \equiv \frac{p_{X_t}}{\partial C_t/\partial X_t} = \frac{w_t}{\partial C_t/\partial \tau_{C_t}}. \tag{79}$$

The first-order condition (72) (or the alternative forms 74, 75 and 76) determines the optimal solution for the control function health investment I_t. The first-order condition (78) determines the optimal solution for the control function consumption C_t. The solutions for the state functions health H_t and assets A_t then follow from the dynamic equations (57) and (58). Length of life T is determined by maximizing the indirect utility function V_T (see 63) with respect to T.

D Mathematical equivalency of first-order conditions

Using (50), integrating the dynamic co-state equation for the marginal value of health $q_H(t)$ (51), and inserting the expressions in (13) we obtain the alternative expressions (15) or (16) for the first-order condition for health investment. Naturally, one can also perform the reverse step. Thus,

$$(13) \Leftrightarrow (15), \tag{80}$$

$$(13) \Leftrightarrow (16). \tag{81}$$

Using the Leibniz integral rule to differentiate the first-order condition for health investment (15) or the alternative expression (16) with respect to t one obtains (17). In other words

$$(15) \Rightarrow (17), \tag{82}$$

$$(16) \Rightarrow (17). \tag{83}$$

Conversely, one can rewrite (17) as a first-order differential equation in $\pi_I(t)$. which can be integrated to provide

$$\begin{aligned}\pi_I(t) &= \pi_I(t')e^{\int_{t'}^{t}[d(u)+\delta]du} \\ &- \int_{t'}^{t}\left[\frac{1}{q_A(0)}\frac{\partial U}{\partial H}e^{-[\beta-\delta]s}+\frac{\partial Y}{\partial H}\right]e^{\int_s^t[d(u)+\delta]du}ds. \end{aligned} \tag{84}$$

For $t'=0$ we obtain (15) and for $t'=T$ we obtain (16). Thus we have

$$(15) \Leftarrow (17), \tag{85}$$

$$(16) \Leftarrow (17). \tag{86}$$

Naturally, the equivalency of the first-order conditions, established above for continuous time, is also true in the discrete time formulation. From (64), (66), and (72), one obtains (74) or (75), and one can also make the reverse step. Taking the difference between period t and $t-1$ (analogous to differentiating with respect to time in continuous time) of either expression (74) or (75) one arrives at (76). And, using recursive backward or forward substitution of relation (76) one arrives at (74) or (75).

E Indeterminacy of health investment

One can also establish the indeterminacy of health investment for CRTS by considering the alternative expression for the first-order condition for health investment (15). This condition equates the current marginal cost of investment in health $\pi_I(t)$ (LHS) with a function of the current and all *past* values of the marginal utility of health $\partial U(s)/\partial H(s)$ and of the marginal production benefit of health $\partial Y(s)/\partial H(s)$ $(0 \leq s \leq t)$ (RHS). The LHS of (15) is not a function of health investment as the marginal cost of health investment $\pi_I(t)$ is independent of the level of investment for a CRTS health-production process. The RHS of (15) is also not a function of current investment $I(t)$ because the marginal utility of health $\partial U(s)/\partial H(s)$ and the marginal production benefit of health $\partial Y(s)/\partial H(s)$ are functions of the health stock $H(s)$ $(0 \leq s \leq t)$ which in turn is a function of past but not current health investment $I(s)$ $(s < t)$.[55] Since the consumption $q_A(0)^{-1}\partial U/\partial H$ and production $\partial Y/\partial H$ benefits of health are independent of the level of current health

[55]This can be seen as follows. The evolution of the health stock $H(t)$ is determined by the dynamic equation (2) which can be written (using 5) as

$$H(t) = H(0)e^{-\int_0^t d(u)du} + \int_0^t I(s)^{\alpha}e^{-\int_s^t d(u)du}ds. \tag{87}$$

Health $H(t)$ is a function of past health investment $I(s)$ but not of current health investment $I(t)$ $(0 \leq s < t)$. That health is not a function of current investment is clear from the formulation of the model, where past but not current investment contributes to current health. In continuous time this means that the domain of s in (87) does not include the end point t: $s \in [0, t)$. In discrete time it is easier to see this. The evolution of the health stock H_t is determined by the dynamic equation (57) which can investment $I(t)$, the first-order condition (15) is not a function of current health investment $I(t)$ and the level of health investment is not determined.[56]

Using similar reasoning one can also proof the indeterminacy of health investment using the alternative conditions (16) and (17).

F Repeated bang-bang behavior

It is of interest to see if we can somehow recover health investment from the first-order condition for health investment (13) for the case of CRTS ($\alpha = 1$). To this end, differentiate both sides of (13) separately as many times as needed till health investment $I(t)$ appears. After differentiating the condition twice it is possible to obtain an expression for the level of health investment $I(t)^{\#}$ for which the (now) twice-differentiated relation holds (i.e. both sides equilibrate)

$$\begin{aligned} I(t)^{\#} &= d(t)H(t) + \phi(t)\left\{\frac{\partial^2 \pi_I}{\partial t^2} - q_{h/a}(t)\frac{\partial d}{\partial t} - \frac{\partial q_{h/a}(t)}{\partial t}[d(t)+\delta]\right\} \\ &+ \phi(t)\left\{\frac{1}{q_A(0)}\frac{\frac{\partial^2 U}{\partial C \partial H}}{\frac{\partial^2 U}{\partial C^2}} e^{-(\beta-\delta)t}\frac{\partial U}{\partial C}\left[\frac{\frac{\partial \pi_C}{\partial t}}{\pi_C(t)} - (\beta-\delta)\right]\right\} \\ &- \phi(t)\left\{\frac{\partial w}{\partial t}\frac{\partial s}{\partial H} + (\beta-\delta)\frac{1}{q_A(0)}\frac{\partial U}{\partial H}e^{-(\beta-\delta)t}\right\} \end{aligned} \quad (89)$$

be written (using 60) as

$$H_t = H_0 \prod_{j=0}^{t-1}(1-d_j) + \sum_{j=0}^{t-1} I_j^{\alpha} \prod_{i=j+1}^{t-1}(1-d_i). \quad (88)$$

In words, health H_t is a function of past I_j but not of current health investment I_t (since $j < t$).

[56] Grossman (2000) has questioned the indeterminacy of health investment, noting (in a discrete time setting) that the first-order condition for health investment (75; equivalent to 16 in continuous time) equates the current marginal cost of investment in health π_{I_t} (LHS) with a function of all *future* values of the marginal utility of health $\partial U_s/\partial H_s$ and of the marginal production benefit of health $\partial Y_s/\partial H_s$ ($t < s \leq T-1$) (RHS). These in turn are functions of health and health is a function of all *past* values of health investment I_s ($0 \leq s < t$; see equation 88; equivalent to 87 in continuous time). Thus the RHS of the first-order condition for health investment (75) is a function of current health investment I_t (and, in fact, all future and all past values as well) and hence a solution for health investment I_t ought to exist. This apparent discrepancy can be reconciled by noting that implicit in the first-order condition for health investment (75) is the use of the final period $t = T-1$ as the point of reference, while the relation (88) for the health stock uses the initial period $t = 0$ as the point of reference. Consistently using the initial period $t = 0$ as the point of reference, i.e., using the form (74) instead of (75) for the first-order condition for health investment, one finds that the RHS of (74) is not a function of current investment as the health stock is a function of past but not current health investment I_s ($s < t$). Likewise, consistently using the final period $t = T$ as the point of reference, i.e., using the alternative expression $H_t = H_T/[\prod_{i=t}^{T-1}(1-d_i)] - \sum_{j=t}^{T-1} I_j^{\alpha}/[\prod_{i=t}^{j}(1-d_i)]$ and comparing this with the first-order condition (75) one finds that the first-order condition is independent of current health investment I_t.

where

$$\phi(t) \equiv \left\{ w(t) \frac{\partial^2 s}{\partial H^2} + \left[\frac{\left(\frac{\partial^2 U}{\partial C \partial H}\right)^2}{\frac{\partial^2 U}{\partial C^2}} - \frac{\partial^2 U}{\partial H^2} \right] \frac{1}{q_A(0)} e^{-(\beta-\delta)t} \right\}^{-1}, \tag{90}$$

$$\frac{\partial q_{h/a}}{\partial t} \equiv q_{h/a}(t)[d(t)+\delta] - \frac{1}{q_A(0)} \frac{\partial U}{\partial H} e^{-(\beta-\delta)t} - \frac{\partial Y}{\partial H}, \tag{91}$$

and

$$q_{h/a}(t) \equiv \frac{q_H(t)}{q_A(t)}. \tag{92}$$

Integrating (89) twice one finds that for this unique level of investment $I(t)^{\#}$ the condition $\pi_I(t) = q_{h/a}(t) + K_1 + K_2 t$ holds, i.e. unless $K_1 + K_2 t = 0$, the first-order condition for health investment (13) does not hold.[57] Once more we cannot determine the level of health investment as $I(t)^{\#}$ is not consistent with (13). By definition, however, (13) holds for $H^*(t)$ so that $K_1 + K_2 t = 0$ for $H^*(t)$ (as Figure 1 illustrates). From this discussion it might seem that $H(t)^*$ and $I(t)^{\#}$ are plausible solutions for the canonical CRTS Grossman model. However, given $H(t)^*$ and $I(t)^{\#}$ the first-order condition for health investment (13) holds only momentarily, since an instant ϵ later, true health $H(t+\epsilon)^{\text{true}}$, evolving according to $\partial H/\partial t = I(t)^{\#} - d(t)H^*(t)$ (see 2 for $\alpha = 1$), is no longer at the "equilibrium" level $H(t+\epsilon)^*$, assumed to evolve according to (13) (or the alternative forms 15, 16, or 17). As a result, the difference in health needs to be repeatedly dissipated and there is no end to the bang-bang behavior.

G Longevity

Note that

$$\begin{aligned} \left.\frac{\partial \Im(T)}{\partial A_0}\right|_T = {} & \left.\frac{\partial \Im(T)}{\partial I(T)} \frac{\partial I(T)}{\partial A_0}\right|_T + \left.\frac{\partial \Im(T)}{\partial C(T)} \frac{\partial C(T)}{\partial A_0}\right|_T + \left.\frac{\partial \Im(T)}{\partial A(T)} \frac{\partial A(T)}{\partial A_0}\right|_T \\ & + \left.\frac{\partial \Im(T)}{\partial H(T)} \frac{\partial H(T)}{\partial A_0}\right|_T + \left.\frac{\partial \Im(T)}{\partial q_A(T)} \frac{\partial q_A(T)}{\partial A_0}\right|_T + \left.\frac{\partial \Im(T)}{\partial q_H(T)} \frac{\partial q_H(T)}{\partial A_0}\right|_T, \end{aligned} \tag{93}$$

where $\partial \Im(T)/\partial I(T) = 0$, $\partial \Im(T)/\partial C(T) = 0$, which follows from the first-order conditions, $\partial A(T)/\partial A_0|_T = \partial H(T)/\partial A_0|_T = 0$, since $A(T)$ and $H(T)$ are fixed, $\partial \Im(T)/\partial q_H(T) = \partial H(t)/\partial t|_{t=T}$ and $\partial \Im(T)/\partial q_A(T) = \partial A(t)/\partial t|_{t=T}$, and I have used (13) and (50).

[57] Note that the literature has not previously derived and does in fact not employ the expression $I(t)^{\#}$ for the level of health investment.

H $\partial H(T)/\partial A_0$ is positive and increases with δT.

First, solve the state equation for health (2):

$$H(t) = H(0)e^{-\int_0^t d(x)dx} + \int_0^t I(s)^\alpha e^{-\int_s^t d(x)dx} ds. \tag{94}$$

Then take the derivative of (94) with respect to T to obtain

$$\left.\frac{\partial H(t)}{\partial T}\right|_{A_0} = \int_0^t \left\{ \alpha I(s)^{\alpha-1} \left.\frac{\partial I(s)}{\partial T}\right|_{A_0} \right\} e^{-\int_s^t d(x)dx} ds. \tag{95}$$

Now take the derivative of (94) with respect to T for $t = T$

$$\begin{aligned}\left.\frac{\partial H(T)}{\partial T}\right|_{A_0} &= \left.\frac{\partial H_{\min}}{\partial T}\right|_{A_0} = 0 \\ &= I(T)^\alpha - d(T)H(T) + \left.\frac{\partial H(t)}{\partial T}\right|_{A_0,t=T} \\ &= \left.\frac{\partial H(t)}{\partial t}\right|_{A_0,t=T} + \left.\frac{\partial H(t)}{\partial T}\right|_{A_0,t=T},\end{aligned} \tag{96}$$

where we distinguish in notation between $\partial H(t)/\partial t|_{A_0,t=T}$, which represents the derivative with respect to time t at $t = T$, and $\partial H(t)/\partial T|_{A_0,t=T}$, which represents variation with respect to the parameter T at $t = T$.

The derivative of health with respect to time at $t = T$ is negative since $H(t)$ approaches $H_{\min}$ from above. Thus we have

$$\left.\frac{\partial H(t)}{\partial T}\right|_{A_0,t=T} = -\left.\frac{\partial H(t)}{\partial t}\right|_{A_0,t=T} > 0.$$

Intuitively, if length-of-life is extended to $T + \delta T$ the health stock has to be higher at the previous point of death T, and it is higher by exactly the change in health over a small period of time. Thus, $\partial H(T)/\partial A_0 = \partial H(T)/\partial A_0|_T + (\partial H(t)/\partial T|_{A_0,t=T})(\partial T/\partial A_0) = (\partial H(t)/\partial T|_{A_0,t=T})(\partial T/\partial A_0) > 0$.

Further, consider equation (26) for $t = T$ and $Z = A_0$ and note that for fixed T, $\partial H(T)/\partial A_0|_T = 0$ (see earlier discussion), so that the value of the total differential $\partial H(T)/\partial A_0$ is determined by the second term on the RHS of (26). Since $\partial H(t)/\partial T|_{A_0,t=T} = -\,\partial H(t)/\partial t|_{A_0,t=T} = -[I(T)^\alpha - d(T)H(T)]$ (compare with 96), it is the same for any trajectory as it represents the negative of the derivate with respect to time t at $t = T$ of the unperturbed (unchanged) path. Thus the end point $\partial H(T)/\partial A_0$ is proportional to the degree of life extension afforded by additional wealth $\partial T/\partial A_0$ and lies further to the right in the phase diagram for greater $\partial T/\partial A_0$.

I Variation in permanent wages and education

Permanent wages w_E The comparative dynamic effect of a permanent increase in the wage rate $w(t)$, through, e.g., an increase in the parameter w_E in (35), on the marginal cost of health investment $\pi_I(t)$ can be obtained by taking the derivate of (23) with respect to w_E and keeping first-order terms (total differentials, free T):

$$\begin{aligned}\frac{\partial}{\partial t}\frac{\partial \pi_{I(t)}}{\partial w_E} \approx\;& \frac{w(t)}{w_E}\frac{\partial s}{\partial H} + \frac{1}{q_A(0)^2}\frac{\partial U}{\partial H}e^{-(\beta-\delta)t} \times \frac{\partial q_A(0)}{\partial w_E} + [d(t)+\delta] \times \frac{\partial \pi_I(t)}{\partial w_E} \\ -\;& \left\{\frac{1}{q_A(0)}\frac{\partial^2 U}{\partial H^2}e^{-(\beta-\delta)t} + \frac{\partial^2 Y}{\partial H^2}\right\} \times \frac{\partial H(t)}{\partial w_E}. \qquad (97)\end{aligned}$$

The first term on the RHS of (97) represents a wealth effect. Permanently higher wages raise the production benefit of health, as health is more valuable in reducing sick time (freeing time for work) when wages are higher. In addition there is the usual wealth effect (second term on the RHS). Both wealth terms are negative since sick time decreases with health $\partial s/\partial H < 0$, and $\partial q_A(0)/\partial w_E|_T < 0$ because w_E raises lifetime earnings (permanent income) and relaxes the budget constraint (3). Variation in permanent wages δw_E is thus distinct from variation in wealth δA_0 in that it not only raises the consumption benefit of health (as is the case for variation in δA_0) but also the production benefit of health (which is not the case for variation in δA_0).

Likewise, the comparative dynamic effect of a permanent increase in the wage rate on health $H(t)$ is obtained by taking the derivative of (24) with respect to w_E and keeping first-order terms

$$\frac{\partial}{\partial t}\frac{\partial H(t)}{\partial w_E} \approx -\frac{\alpha}{1-\alpha}\frac{\kappa_I}{w_E}I(t)^\alpha + \frac{\alpha}{1-\alpha}\frac{I(t)^\alpha}{\pi_I(t)} \times \frac{\partial \pi_I(t)}{\partial w_E} - d(t) \times \frac{\partial H(t)}{\partial w_E}, \qquad (98)$$

where the first term on the RHS of equation (98) represents the negative effect of the opportunity cost of time on health investment, and in turn on health.

The corresponding phase diagram is shown in Figure 7. It is nearly identical to the phase diagram for variation in initial wealth δA_0 (RHS of Figure 3), except that the $(\partial/\partial t)(\partial H(t)/\partial w_E)$ null cline crosses the vertical $\partial \pi_I(t)/\partial w_E$ axis at $\pi_I(t)\kappa_I/w_E$ and not at the origin. This term represents the effect of a permanent increase in wages w_E on the opportunity cost of investing time in health. The $(\partial/\partial t)(\partial \pi_I(t)/\partial w_E)$ null cline crosses the vertical $\partial \pi_I(t)/\partial w_E$ axis at $(-\partial s/\partial H)\,(w(t)/w_E)/\,[d(t)+\delta] - \left[q_A(0)^{-2}\partial U/\partial H e^{-(\beta-\delta)t}\right]\partial q_A(0)/\partial w_E$. This expression represents a wealth, or permanent income, effect: permanently higher wages increase the production benefit of health (first term) and increases wealth, thereby raising the consumption benefit of health (second term; operating through $\partial q_A(0)/\partial w_E < 0$). In the scenario depicted in Figure 7, it is assumed that the opportunity cost of time effect is small compared to the wealth / permanent income effect.

Following similar steps as in section 3.2.1 for variation in wealth, we first need to establish whether length of life is extended as a result of a permanent increase in income.

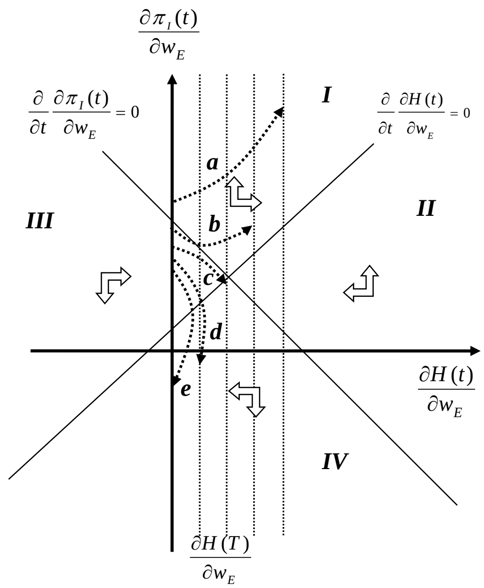

Figure 7: *Phase diagram of the perturbation due to variation in permanent wages δw_E, allowing length of life T to be optimally chosen. The four vertical dotted lines represent different potential values for the end point $\partial H(T)/\partial w_E$.*

This can be accomplished by considering the fixed T case. The comparative dynamic effect of a permanent increase in the wage rate w_E on longevity can be obtained by replacing A_0 with w_E in (29) and (30). Since $\partial q_A(0)/\partial w_E|_T < 0$ (diminishing returns to wealth / permanent income), it follows that, similar to the case for variation in initial wealth δA_0 (see section 3.2.1), a sufficient condition for life extension in response to positive variation in permanent wages is $\partial \pi_I(T)/\partial w_E|_T \leq 0$.

For fixed T all admissible paths in the phase diagram have to start and end at the vertical axis, since $H(0)$ and $H(T)$ are fixed.[58] Trajectory e in the phase diagram of Figure 7 is consistent with these conditions and the trajectory is characterized by $\partial \pi_I(T)/\partial w_E|_T < 0$. Thus length of life is extended $\partial T/\partial w_E > 0$.

Considering the T free case, the reasoning is identical to the discussion in section 3.2.1. Following the logic outlined there, I find that trajectories a, b, c, and d are consistent with life extension. The greater life is extended as a result of greater permanent income, the further to the right is the trajectory's end point $\partial H(T)/\partial w_E$. Example trajectory a is associated with a large increase in the marginal value of health $\partial \pi_I(t)/\partial w_E$ and in health $\partial H(t)/\partial w_E$, compared to the unperturbed trajectory, and this trajectory is associated with the greatest gain in longevity $\partial T/\partial w_E$. Trajectory b and c represent an intermediary

[58]As discussed before, the coefficients of the comparative dynamic equations (97) and (98) are identical for the partial differentials, $\partial \pi_I(t)/\partial w_E|_T$ and $\partial H(t)/\partial w_E|_T$, for fixed T, and for the total differentials, $\partial \pi_I(t)/\partial w_E$ and $\partial H(t)/\partial w_E$, for free T. We can thus use the same phase diagram for both cases.

case, and trajectory d a case of limited response (the latter most closely resembles the fixed T case, represented by trajectory e). Trajectory f is incompatible with live extension and ruled out.

These results rely on the assumption that the opportunity cost effect is smaller than the wealth / permanent income effect. If, however, the opportunity cost effect is substantial, the $(\partial/\partial t)(\partial H(t)/\partial w_E)$ null cline is shifted further upward in the phase diagram of Figure 7 than shown. A trajectory similar to e might then end up above the $\partial H(t)/\partial w_E$ axis with a positive end value of $\partial \pi_I(T)/\partial w_E|_T$, in which case we cannot unambiguously establish that length of life increases.[59]

If the opportunity cost is very high, outweighing the wealth / permanent income effect, the $\partial H(t)/\partial w_E$ null cline could even cross the vertical $\partial \pi_I(t)/\partial w_E$ axis above the location where the $\partial \pi_I(t)/\partial w_E$ null cline crosses the vertical $\partial \pi_I(t)/\partial w_E$ axis. In such a scenario (not shown), for fixed T, any admissible trajectory is characterized by $\partial \pi_I(T)/\partial w_E|_T > 0$, and we cannot unambiguously establish that length of life increases. While theoretically we cannot rule out this scenario, empirical evidence suggests that a permanent wage change affects health positively, while a transitory wage increase affects health negatively (e.g., Contoyannis, Jones and Rice, 2004), and that high-income individuals are generally in better health than low-income individuals. Thus, in practice it appears the opportunity cost effect is not large.

Education E The comparative dynamic effect of an increase in education E (see 34 and 35), on the marginal cost of health investment $\pi_I(t)$ is obtained by taking the derivate of (23) with respect to E and keeping first-order terms (total differentials, free T):

$$\begin{aligned}\frac{\partial}{\partial t}\frac{\partial \pi_I(t)}{\partial E} \quad &\approx \quad \rho_w w(t)\frac{\partial s}{\partial H} + \frac{1}{q_A(0)^2}\frac{\partial U}{\partial H}e^{-(\beta-\delta)t} \times \frac{\partial q_A(0)}{\partial E} + [d(t)+\delta] \times \frac{\partial \pi_I(t)}{\partial E} \\ &- \quad \left\{\frac{1}{q_A(0)}\frac{\partial^2 U}{\partial H^2}e^{-(\beta-\delta)t} + \frac{\partial^2 Y}{\partial H^2}\right\} \times \frac{\partial H(t)}{\partial E}. \end{aligned} \tag{100}$$

Likewise, the comparative dynamic effect of an increase in education on health $H(t)$ is obtained by taking the derivative of (24) with respect to E and keeping first-order terms:

$$\frac{\partial}{\partial t}\frac{\partial H(t)}{\partial E} \quad \approx \quad \frac{\alpha}{1-\alpha}I(t)^\alpha\left[\rho_I - \kappa_I \rho_w\right] + \frac{\alpha}{1-\alpha}\frac{I(t)^\alpha}{\pi_I(t)} \times \frac{\partial \pi_I(t)}{\partial E} - d(t) \times \frac{\partial H(t)}{\partial E}. \tag{101}$$

Contrasting the results of the comparative dynamics for education E (equations 100 and 101) with those obtained for permanent income w_E (equations 97 and 98) we observe

[59] It is possible that length of life is still extended $\partial T/\partial w_E > 0$, even if $\partial \pi_I(T)/\partial w_E|_T > 0$, as long as

$$\left.\frac{\partial q_A(0)}{\partial w_E}\right|_T e^{-\delta T}\pi_I(T) + q_A(0)e^{-rT}\left.\frac{\partial \pi_I(T)}{\partial w_E}\right|_T < 0 \tag{99}$$

(see expression 30). It is not clear from the phase diagram that this condition holds, hence we cannot establish whether life is extended.

that permanent wages w_E and education E operate in the same way. This should come as no surprise, as they both increase permanent wages. There is however one important difference: the first term on the RHS of (101) represents both the effect of education on the efficiency of health investment ρ_I (the educated are assumed to be more efficient producers of health) and the effect of education on the opportunity cost of time $\kappa_I \rho_w$. The efficiency effect of education reduces the opportunity cost of time effect. The phase diagram for the effect of variation in education δE is essentially the same as for variation in permanent income w_E, shown in Figure 7, and replacing w_E by E (for this reason I do not provide a separate phase diagram). Given strong empirical support for a positive association between education and health, it could be that the efficiency effect dominates, in which case the $(\partial/\partial t)(\partial H(t)/\partial E)$ null cline would cross the vertical $\partial \pi_I(t)/\partial E$ axis below, instead of above, the origin. This would make the case for variation in education δE stronger (compared to the case for variation in permanent wages w_E) in ensuring that the condition $\partial \pi_I(t)/\partial E|_T \leq 0$ is obtained and hence length of life is extended.

J Structural relations for empirical testing

From the utility function (31) and the first-order conditions (76) and (78) it follows that

$$C_t = \frac{\zeta}{1-\zeta} \frac{\sigma_{H_t} - \varphi_{H_t}}{\pi_{C_t}} H_t, \tag{102}$$

and

$$\begin{aligned} H_t &= (1-\zeta)\Lambda(q_0^A)^{-1/\rho} \left(\sigma_{H_t} - \varphi_{H_t}\right)^{-\chi} \pi_{C_t}^{\chi - 1/\rho} \frac{\prod_{j=1}^t (1+\beta_j)^{-1/\rho}}{\prod_{j=1}^t (1+\delta_j)^{-1/\rho}}, && (103) \\ C_t &= \zeta\Lambda(q_0^A)^{-1/\rho} \left(\sigma_{H_t} - \varphi_{H_t}\right)^{1-\chi} \pi_{C_t}^{\chi - 1/\rho - 1} \frac{\prod_{j=1}^t (1+\beta_j)^{-1/\rho}}{\prod_{j=1}^t (1+\delta_j)^{-1/\rho}}, && (104) \end{aligned}$$

where Λ and χ are defined in (43) and (44).

Analogous to (21) define

$$I_t = \mu_{I_t} m_t^{1-k_I} \tau_{I_t}^{k_I}. \tag{105}$$

Using equations (73) and (105) we obtain

$$\begin{aligned} \pi_{I_t} &= \frac{p_{m_t}^{1-k_I} w_t^{k_I}}{\alpha\, k_I^{k_I} (1-k_I)^{1-k_I} \mu_{I_t}} I_t^{1-\alpha} \equiv \pi_{I_t}^* I_t^{1-\alpha}, && (106) \\ m_t &= \left(\frac{1-k_I}{k_I}\right)^{k_I} \mu_{I_t}^{-1} p_{m_t}^{-k_I} w_t^{k_I} I_t \equiv m_{I_t}^* I_t, && (107) \\ \tau_{I_t} &= \left(\frac{1-k_I}{k_I}\right)^{-(1-k_I)} \mu_{I_t}^{-1} p_{m_t}^{1-k_I} w_t^{-(1-k_I)} I_t \equiv \tau_{I_t}^* I_t. && (108) \end{aligned}$$

Analogously to health investment I_t (equation 105), consumption C_t is assumed to be produced by combining own time and goods/services purchased in the market according to a Cobb-Douglas CRTS production function

$$C_t = \mu_{C_t} X_t^{1-k_C} \tau_{C_t}^{k_C}, \tag{109}$$

where μ_{C_t} is an efficiency factor and $1-k_C$ and k_C are the elasticities of consumption C_t with respect to goods and services X_t purchased in the market and with respect to own-time τ_{C_t}, respectively.

Using equations (79) and (109) we have

$$\pi_{C_t} = \frac{p_{X_t}^{1-k_C}\, w_t^{k_C}}{k_C^{k_C}(1-k_C)^{1-k_C}\,\mu_{C_t}}, \tag{110}$$

$$X_t = \left(\frac{1-k_C}{k_C}\right)^{k_C} \mu_{C_t}^{-1} p_{X_t}^{-k_C} w_t^{k_C} C_t \equiv X_{C_t}^{*} C_t, \tag{111}$$

$$\tau_{C_t} = \left(\frac{1-k_C}{k_C}\right)^{-(1-k_C)} \mu_{C_t}^{-1} p_{X_t}^{1-k_C} w_t^{-(1-k_C)} C_t \equiv \tau_{C_t}^{*} C_t. \tag{112}$$

From (77), (33), (103), (106) and (110) follows

$$a_t^1 I_t^{1-\alpha} - (1-\alpha) I_t^{1-\alpha} \widetilde{I_t} = a_t^2 H_t^{-1/\chi} + a_t^3 H_t^{-(1+\gamma)}, \tag{113}$$

where

$$a_t^1 \equiv \left[d_t + \delta_t - (1-k_I)\widetilde{p_{m_t}} - k_I \widetilde{w_t} + \widetilde{\mu_{I_t}}\right], \tag{114}$$

$$a_t^2 \equiv \left[(1-\zeta)\Lambda(q_0^A)^{-1/\rho}\right]^{1/\chi} (\pi_{I_t}^{*})^{-1} (\pi_{C_t})^{1-1/\rho\chi} \frac{\prod_{j=1}^{t}(1+\beta_j)^{-1/\rho\chi}}{\prod_{j=1}^{t}(1+\delta_j)^{-1/\rho\chi}}, \tag{115}$$

$$a_t^3 \equiv w_t (\pi_{I_t}^{*})^{-1} \Omega_{*}, \tag{116}$$

where the notation $\widetilde{f_t}$ is used to denote the relative change $\widetilde{f_t} \equiv 1 - \frac{f_{t-1}}{f_t}$ in a function f_t and we have assumed small relative changes in the price of medical care $\widetilde{p_{m_t}}$, wages $\widetilde{w_t}$ and the efficiency of the health investment process $\widetilde{\mu_{I_t}}$.

Using (107) and (113) and the functional relations defined in section 3.3.1 we obtain a structural relation (37) between health investment goods and services m_t purchased in the market and the stock of health H_t.

Assuming that both medical goods / services m_t and time input τ_{I_t} increase health investment suggests $0 \le k_I \le 1$ (see equation 105), and if education E increases the efficiency of medical care then $\rho_I > 0$ (see equation 34). Similarly we have $0 \le k_C \le 1$ (see equation 109). For these assumptions and small changes $\widetilde{m_t}$, the demand for health investment goods/services m_t (see relations 37 and 38) decreases with the biological aging rate d_t (and hence with environmental factors that are detrimental to health $\boldsymbol{\xi_t}$), the rate of return to capital δ_t (an opportunity cost – individuals can invest in health or in the stock market) and increases with price increases $\widetilde{p_{m_t}}$ and wage increases $\widetilde{w_t}$ (it is better to invest in health now when prices p_{m_t} and the opportunity cost of time w_t are higher in the future). In addition, due to the *consumption* aspect of health (health provides utility) the demand for health investment goods/services m_t (see relations 37 and 39) increases with wealth (the shadow price of wealth q_0^A is a decreasing function of wealth and life-time earnings), education E (through assumed greater efficiency of health investment with the level of education) and decreases with the price of health investment goods/services p_{m_t}. For $\rho\chi < 1$ the demand for health investment goods/services m_t decreases with the price of consumption goods/services p_{X_t} (for $\rho\chi > 1$ it increases) and with the wage rate w_t (opportunity cost of time) (for $\rho\chi > 1$ the effect of the wage rate w_t is ambiguous). And, due to the *production* aspect of health (health increases earnings) the demand for health investment goods/services m_t (see relations 37 and 40) increases with education E (through assumed greater efficiency of health investment with the level of education) and the wage rate w_t (a higher wage rate increases the marginal production benefit of health, and this outweighs the opportunity cost of time associated with health investment) and decreases with the price of health investment goods/services p_{m_t}.

The above discussion masks important effects of earnings and education. In this model of perfect certainty an evolutionary wage change (along an individual's wage profile) does not affect the shadow price of wealth q_0^A as the change is fully anticipated by the individual. Thus comparing panel data for a single individual may reveal a higher wage rate w_t to be associated with a lower demand for health investment goods / services m_t due to a higher opportunity cost of time. However, comparing across individuals, those who currently have a higher wage rate will in most cases also have higher life-time earnings and thus a lower shadow price of wealth q_0^A. This wealth effect increases the demand for health investment goods / services and competes with the opportunity cost of time effect. Similarly, to account for the effect of education it is important not only to consider the possible effect of a higher efficiency of health investment (the parameter ρ_I), as in the structural relation (37), but also the effect that education has on earnings (opportunity cost of time effect; see equation 35) and in turn on wealth (wealth effect). Plausibly, the wealth effect dominates the opportunity cost effect. For example, Dustmann and Windmeijer (2000) and Contoyannis, Jones and Rice (2004) find a positive effect on health from a permanent wage increase and a negative effect from a transitory wage increase. We expect then that the effect of education and earnings is to increase the demand for health investment goods / services through a wealth effect that may dominate the opportunity cost of time effect associated with higher earnings.[60] Thus, in testing the theory it will be important to account for wealth.

[60] Further, one may be tempted to conclude that individuals invest less in health during middle and old age because of the high opportunity cost of time associated with high earnings at these ages (see equation 35). However, as health deteriorates with age the demand for health investment increases (see sections 3.2 and 3.3). If the latter effect dominates, the model is capable of reproducing the observation that young individuals invest little, the middle-aged invest more, and the elderly invest most in their health.