Working Paper in Economics No. 821

Specification of the Health Production Function and its Behavioral Implications

Kristian Bolin & Michael R. Caputo

Department of Economics, May 2022

ISSN 1403-2473 (Print)
ISSN 1403-2465 (Online)

UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Specification of the Health Production Function and its Behavioral Implications

Kristian Bolin§
Department of Economics
Centre for Health Economics
University of Gothenburg
Gothenburg, Sweden
email: kristian.bolin@economics.gu.se

Michael R. Caputo
Department of Economics
University of Central Florida
P.O. Box 161400
Orlando, FL 32816-1400
email: mcaputo@ucf.edu

Abstract

The health production function of the canonical health-capital model is generalized to allow the state of health to affect the total and marginal products of health investment. If the total and marginal products of health investment are nonincreasing functions of the state of health, then the solution of the generalized model is locally qualitatively identical to that of the canonical model. Moreover, in contrast to the canonical model, the generalized model is able to rationalize the cycling of the state of health and health investment observed in some individuals. The necessary conditions on the health production function for cyclical behavior are identified as well.

Keywords: health investment; health production function; health stock; optimal control

JEL Codes: D15; I12; I18

§ Corresponding author.

1. Introduction

Actions undertaken with the ultimate objective of improving health, or components of health, range from strenuous physical training exerted by athletes whose goal is to outperform competitors, to palliative care provided to terminally ill patients. In the first case, health may be conceptualized as fitness, while in the second it is the absence of pain and suffering. In the economics literature health has typically been conceptualized as a good with stock properties, the amount of which is measured on a scale bounded below by a value corresponding to the state of being dead—this ignores the fact that, from a revealed-preference perspective, some health states are worse than death, seeing as suicides due to health concerns do occur. Adopting this way of measuring health means that a professional athlete is at one end of the scale and a patient receiving palliative care is on the other. But there is a qualitative difference between these two cases: in the first the person is already healthy, while in the second the person is about to die and is therefore provided with treatments in order to make dying as painless as possible. Between these extremes there exists combinations of states of health and actions undertaken to improve them. For example, some non-athletes engage in physical activities for health reasons, and people struck by non-fatal illnesses more or less adhere to medical treatments. Thus, a crude division of health-influencing actions exists between those undertaken in order to improve fitness and those undertaken in order to restore health.

Given the above, the first objective is to generalize the health-capital model of Grossman (1972) to allow the technology employed in the production of increments to health to depend on the rate of health investment *and* the state of health. The canonical model, in contrast, assumes that the technology depends only on the rate of health investment, and typically linearly, as in, say, Strulik (2015). The second is to qualitatively characterize the solution of the generalized model and compare and contrast it with that of the basic model.

In light of the first objective, it is useful to begin with a general discussion of why the marginal product of health investment in general depends on the state of health. In the aforementioned case of an athlete, it seems rather intuitive that improvement in fitness becomes more costly as fitness improves, i.e., as fitness improves, increasing amounts of effort are required to further increase it. Eventually, an athlete realizes that beyond some point, further increases in fitness are essentially infinitely costly. More generally, anyone making an effort to improve health beyond just being "healthy" faces the same situation in which further health increases are more costly to achieve. The opposite situation also occurs, namely, that in which further health gains are less costly as health improves. This can happen when a person is undertaking actions to restore health to a state corresponding to no active illness. Note too that people with serious and potentially life-threatening conditions, and thus low states of health, frequently face a situation in which medical treatments are either non-existent or ineffective. On the other hand, persons suffering from less severe, or trivial, manifestations of the same ailments are likely to be cured or have their health restored. More generally, as pointed out by Kaplan et al., (2010), it is a well-known fact that the response to a given treatment varies between individuals, and that such differences are to a significant extent due to health-related factors. Thus the novel stipulation introduced here, to wit, that the total and marginal products of health investment depend on the state of health, stands on firm medical ground.

It is also useful at this juncture to put matters in more formal terms. To this end, let $F(\cdot)$ be a twice continuously differentiable health production function. Given the preceding discussion, it is assumed to map the state of health H and rate of investment in health I into the gross time-rate-of-change of the state of health. Its value is thus given by $F(H,I)$. If health investment is taken to be the effort exerted to improve fitness, then $F_H(H,I)<0$ and $F_{IH}(H,I)<0$ are the corresponding assumptions. That is, if investment in health corresponds to fitness effort, then the marginal product of health is negative and health investment and the state of health are technical substitutes. If, instead, health investment is restorative, the corresponding assumptions are $F_H(H,I)>0$ and $F_{IH}(H,I)>0$, viz., the marginal product of health is positive and health investment and the state of health are technical complements.

Before turning to the medical evidence that supports the aforesaid signs of the derivatives, consider the plausibility of them from a logical or commonsense perspective. If the effort required to improve fitness were instead constant or declining with respect to the state of fitness, there would, without further assumptions, be no physiological limit to the level of fitness that would be possible to achieve and, consequently, top athletes would be far more frequent. Likewise, if the effectiveness of medical treatments were monotonically decreasing in health, one would frequently observe terminally ill patients being cured and trivially ill patients dying. Note that these cases do not rule out situations in which the marginal product of health investment changes non-monotonically over a given range of health states, making, for example, some medical treatments more effective for more severe health conditions. Even so, the ensuing summary of medical evidence demonstrates that the aforesaid signs of the derivatives largely capture the conditions facing those striving for improved fitness and those striving to restore health.

First consider the evidence that supports the claims $F_H(H,I)<0$ and $F_{IH}(H,I)<0$. In a review article about the health effects of physical activity, Warburton et al. (2006) concluded that "… the greatest improvements in health status are seen when people who are least fit become physically active". This conclusion was repeated in a more recent review article by Warburton et al. (2017) too. In addition, Fournier et al. (2010) found that pharmaceutical treatments for anxiety are more effective for patients who are severally affected than for those with a mild condition.

The medical evidence supporting the assertions $F_H(H,I)>0$ and $F_{IH}(H,I)>0$ is broad. For example, Herrmann et al. (2009), Montazeri (2009), Quinten et al. (2009), and Voutsadakis (2017) found that progression free cancer survival increases with quality of life and declines with comorbidities at the time of diagnosis. Similarly, Herpertz-Dahlmann et al. (1996), Zerwas et al. (2013), Dakanalis, Bartoli, et al. (2017), Dakanalis, Colmegna, et al. (2017), and Lydecker et al. (2020) determined that the likelihood of a positive outcome of eating-disorder treatments declines with baseline severity of the condition. Along the same lines, Batterham et al. (2016) provided evidence that the likelihood of a permanent weight reduction among obese patients subjected to dietary weight loss interventions decreases with body mass index at baseline. And finally, Baganz et al. (2019) found that the likelihood of remission when treated for rheumatoid arthritis decreases with the severity of the condition.

Despite numerous generalizations of the canonical model of Grossman (1972)—see Bolin and Lindgren (2016) for a recent account of such extensions—none have allowed for the health production function to depend on the rate of investment in health *and* the state of health. The generalization contemplated here is thus a rather simple one, in that it allows the total and marginal products of health investment to be functions of the state of health. While allowing the production technology to depend on the state and control variables is hardly an unconventional assumption in other fields of economics, as shown below, it is fair to conclude that an examination of its behavioral implications is both worthwhile and overdue in health economics.

A generalization of a similar flavor appeared more than one-half century ago in Treadway (1970), who studied the local qualitative properties of the neoclassical adjustment cost model of a firm when the production function depended not just on labor and the stock of capital, which is the traditional assumption, but also on the rate of investment. He showed that many of the heretofore refutable qualitative properties of the model became ambiguous under the more general assumption about the production function. Similarly, three decades ago, Levhari and Withagen (1992) studied a generalization of the archetypal nonrenewable resource extraction model in which the biomass production function depended not only on the stock of the biomass (the traditional assumption), but also on the rate of management effort. Here too, as put by Levhari and Withagen (1992, Abstract) "The analysis gives rise to results which differ substantially from the usual outcomes in the economics of renewable resources."

On the basis of the research by Treadway (1970) and Levhari and Withagen (1992), one might expect that some of the qualitative properties of the health-capital model would change as a result of the proposed generalization of the health production function. This is indeed true, but only for a certain class of health production functions. On the other hand, it is shown that if the total and marginal products of health investment are nonincreasing functions of the state of health, then the local stability of a steady state and thus the shape of the trajectories in a neighborhood of a steady state of the generalized model are qualitatively identical to those of the basic model. This implies that when $F_H(H,I) \le 0$ and $F_{IH}(H,I) \le 0$, the generalized model is unable to rationalize certain observed health behaviors, but neither can the basic model. A prominent example of such behavior is the repeated cycling between high and low states of health and their concomitant rates of investment in health that is observed in some people. In technical terms, such cyclical behavior is ruled out by the fact that a steady state of the basic model, as well as that of the generalized model when the total and marginal products of health investment are not increasing in the state of health, is a local saddle point.

A third objective, therefore, is the identification of the conditions on the health production function that are implied by cyclical behavior. It turns out that the implied conditions are not unconventional, and, in fact, mirror the conditions facing persons striving to restore their health. In particular, cyclical behavior implies that the marginal product of health is bounded above by the rate of health depreciation in a steady state, that the marginal product of health investment is an increasing function of the state of health, and that investment and health must be sufficiently complementary in a steady state. Taken together, the contributions show that the generalized model has greater reach than the archetypal model, contains it as a special case, and is no more difficult to understand.

In closing the section, note that while a simple framework is adopted in what follows, in that the resulting optimal control model has but one state variable and two control variables, general functional forms are maintained throughout. Consequently, one can be confident that the qualitative conclusions drawn are robust and not dependent on restrictive functional form assumptions. Moreover, despite the simplicity of the generalized model, the essential features of the health-capital model are retained. A further benefit of formulating the generalized health-capital model in a simple manner is that it permits a characterization of its solution by way of a phase diagram. As such, one is not required to delve into excessive technical details in order to understand the results, thereby permitting the paper to reach a wider audience.

2. A Generalized Health Capital Model and its Assumptions

To begin, let $C(t) \in \mathbb{R}_+$ be the consumption rate of a composite good at time t that does not affect an agent's state of health $H(t) \in \mathbb{R}_+$ at time t, the latter of which is hereafter referred to as health. Instantaneous preferences are represented by a felicity function $U(\cdot)$, and agents are assumed to care about consumption of the composite good and their health. Thus, the value of $U(\cdot)$ at time t is $U\big(C(t),H(t)\big)$. The state equation for health is $\dot{H}(t)=F\big(H(t),I(t)\big)-\delta H(t)$, where $\delta\in\mathbb{R}_{++}$ is a constant rate of depreciation of health, $I(t)\in\mathbb{R}_{+}$ is an agent's health investment rate—hereafter investment—at time t, and $F(\cdot)$ is a health production function, mapping health and investment to the gross time-rate-of-change of health.

Furthermore, an agent's initial health $H(0)$ is fixed at the value $H_0\in\mathbb{R}_{++}$, implying that $H(0)=H_0$ is an initial condition. Similarly, an agent's health at the time of death $T\in\mathbb{R}_{++}$ is given by $H(T)$, and is assumed to be bounded below by the value $H_T\in\mathbb{R}_{+}$, often referred to as the minimum health required for life, implying that $H(T)\geq H_T$ is a terminal condition on health. Initially, the time of death is known and fixed at a finite value $T\in\mathbb{R}_{++}$, but this assumption is later changed to allow for infinitely-lived agents—a case emphasized by Strulik (2015, pp. 310–312)—in order to determine the qualitative implications of the assumption.

Depending on the assumption made about financial markets, the budget constraint can take one of two forms. If financial markets are assumed to be perfect, then the budget constraint is of the lifetime variety. Unfortunately, this assumption introduces a considerable complication in the model in the form of a second state variable, namely wealth, and as such, does not permit analysis of the resulting optimal control model by way of a phase diagram. As a result, this assumption is eschewed in favor of the assumption that perfect financial markets do not exist.

Without loss of generality, it is assumed that the price of consumption is unity. Accordingly, the budget constraint takes the form $C(t)+p_I I(t)=Y$, where $p_I\in\mathbb{R}_{++}$ is the constant (relative) price of investment and $Y\in\mathbb{R}_{++}$ is a constant flow of (relative) income. This approach mimics that of Ferguson (2000), who sought to obtain a qualitative characterization of a simplified version of the rational addiction model via a phase diagram, Forster (2001), who studied a model with healthy and unhealthy consumption, and Strulik (2015), who derived a closed-form solution for a special case of the model studied here, among others.

Pulling the above information together and substituting $C(t) = Y - p_I I(t)$ in $U(\cdot)$, an agent is asserted to behave as if solving the optimal control problem

$$\max_{I(\cdot)} \int_0^T U\big(Y - p_I I(t), H(t)\big) e^{-\rho t}\, dt$$
$$\text{s.t. } \dot{H}(t) = F\big(H(t), I(t)\big) - \delta H(t),\ H(0) = H_0,\ H(T) \geq H_T. \tag{1}$$

The following assumptions are imposed on problem (1) and discussed subsequently.

(A1) For all $(C,H) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$, $U(\cdot) \in C^{(2)}$, $U_C(C,H) > 0$, $U_H(C,H) > 0$, $U(\cdot)$ is strongly concave in (C,H), and $U_{CH}(C,H) \geq 0$.

(A2) For all $(H,I) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$, $F(\cdot) \in C^{(2)}$, $F_I(H,I) > 0$, $F(\cdot)$ is strongly concave in (H,I), and $F(H,0) = 0$.

(A3) For all $\boldsymbol{\theta} \in B(\boldsymbol{\theta}^0; \varepsilon_1)$, there exists an interior and finite investment rate $I^*(t;\boldsymbol{\theta})$ and health state $H^*(t;\boldsymbol{\theta})$, and a current-value costate $\lambda^*(t;\boldsymbol{\theta})$, that satisfy the necessary conditions of problem (1), where $\boldsymbol{\theta} \stackrel{\text{def}}{=} (T, H_0, H_T, \delta, p_I, \rho, Y)$ is a parameter vector, $\boldsymbol{\theta}^0$ is a given value of $\boldsymbol{\theta}$, and $B(\boldsymbol{\theta}^0; \varepsilon_1)$ is an open 7-ball centered at $\boldsymbol{\theta}^0$ of radius $\varepsilon_1 > 0$.

(A4) For all $\boldsymbol{\beta} \in B(\boldsymbol{\beta}^0; \varepsilon_2)$ there exists an interior and finite investment rate $I^{ss}(\boldsymbol{\beta})$ and health state $H^{ss}(\boldsymbol{\beta})$, and a current-value costate $\lambda^{ss}(\boldsymbol{\beta})$, that satisfy the steady state version of the necessary conditions of problem (1), where $\boldsymbol{\beta} \stackrel{\text{def}}{=} (\delta, p_I, \rho, Y)$.

Assumption (A1) asserts that instantaneous preferences are strongly monotonic in consumption and health, that is, consumption and health are goods. It also asserts that the Hessian matrix of the felicity function is negative definite, and in particular, that $U_{CC}(C,H) < 0$ and $U_{HH}(C,H) < 0$, i.e., consumption and health exhibit diminishing marginal utility. These are ubiquitous assumptions that permit one to invoke a sufficiency theorem, which is useful when studying the qualitative properties of an optimal control problem using general functional forms, as is the case here. Finally, note that consumption and health are assumed to be unrelated or complements, which is the usual assumption.

Supposition (A2) gives the maintained assumptions on the health production function. Per usual, the marginal product of investment is assumed to be positive and the Hessian matrix of the production function with respect to health and investment is assumed to be negative definite, the latter implying that health and investment exhibit diminishing marginal productivity. The signs of the partial derivatives $F_H(H,I)$ and $F_{IH}(H,I)$, and the implications of them for the qualitative nature of the solution to problem (1), are the focus of §4–§6.

Assumption (A3) is relatively mild, in that it only asserts the existence of a solution to the necessary conditions of problem (1). But as shown in the next section, assumptions (A1) and (A2) imply that the necessary conditions are sufficient, and therefore that a solution of the necessary condition is the unique, globally optimal solution of problem (1). Finally, assumption (A4) mimics (A3), but applies to the steady state solution of problem (1)

Having dispensed with the above, the ensuing section presents basic results that are essential for a qualitative characterization of the solution to problem (1).

3. Preliminary Results

To begin, define the current-value Hamiltonian of problem (1) by

$$\Phi(H,I,\lambda) \overset{\text{def}}{=} U(Y - p_I I, H) + \lambda[F(H,I) - \delta H], \tag{2}$$

where λ is the current-value costate variable associated with health, and has the economic interpretation of the current-value shadow value of health—hereafter, shadow value of health. Given assumption (A3), Theorem 10.2 of Caputo (2005) gives the necessary conditions

$$\Phi_I(H,I,\lambda) = -p_I U_C(Y - p_I I, H) + \lambda F_I(H,I) = 0, \tag{3}$$

$$\dot{H} = \Phi_\lambda(H,I,\lambda) = F(H,I) - \delta H, \; H(0) = H_0, \tag{4}$$

$$\dot{\lambda} = \rho\lambda - \Phi_H(H,I,\lambda) = [\rho + \delta - F_H(H,I)]\lambda - U_H(Y - p_I I, H), \tag{5}$$

$$H(T) \geq H_T, \; \lambda(T) \geq 0, \; \lambda(T)[H(T) - H_T] = 0. \tag{6}$$

The above necessary conditions are also sufficient, as is now demonstrated.

The sufficiency of Eqs. (3)–(6) is established by showing that $\Phi(\cdot)$ is strongly concave in (H,I). In order to do so, first define the function $V(\cdot)$ by $V(H,I) \overset{\text{def}}{=} U(Y - p_I I, H)$. It is straightforward to show that the Hessian matrix of $V(\cdot)$ is negative definite, and thus that $V(\cdot)$ is strongly concave in (H,I). Second, note that $F(H,I) - \delta H$ is also strongly concave in (H,I) by assumption (A2). Third, it follows from Eq. (3) and assumptions (A1), (A2), and (A3) that

$\lambda^*(t;\boldsymbol{\theta}) > 0$ for all $t \in [0,T]$. And fourth, the preceding deductions show that $\Phi(\cdot)$ is a positive sum of strongly concave functions and hence strongly concave in (H,I). Consequently, by Theorem 10.4 of Caputo (2005), a solution of the necessary conditions of problem (1) is the globally optimal solution of problem (1).

In the above proof it was shown that $\lambda^*(t;\boldsymbol{\theta}) > 0$ for all $t \in [0,T]$, i.e., health is a good according to lifetime preferences. This implies that $H^*(T;\boldsymbol{\theta}) = H_T$ by way of the transversality conditions in Eq. (6). In other words, terminal health is equal to the lower bound on health in an optimal plan. The conclusion that $H^*(T;\boldsymbol{\theta}) = H_T$, however, is predicated on the assumption of an interior solution for investment.

If instead a nonnegativity constraint on investment were introduced, then Eq. (3) would be replaced with $\lambda F_I(H,I) \le p_I U_C(Y - p_I I, H)$. But as $p_I \in \mathbb{R}_{++}$, $U_C(C,H) > 0$ by assumption (A1), and $F_I(H,I) > 0$ by assumption (A2), the sign of $\lambda^*(t;\boldsymbol{\theta})$ could be positive, negative, or zero. Given that $\lambda(T) \ge 0$ by Eq. (6), it is possible in this case that $\lambda^*(T;\boldsymbol{\theta}) = 0$, and therefore by complementary slackness, that $I^*(T;\boldsymbol{\theta}) = 0$ too. This possibility, to wit, no investment in health, a zero shadow value of health, and health at or above its lower bound at the last moment of life, will be revisited when the phase diagram is developed in the next section.

In order to construct the phase diagram, the necessary and sufficient conditions in Eqs. (3)–(6) must be reduced to a pair of ordinary differential equations and boundary conditions. This is accomplished by (i) differentiating Eq. (3) with respect to t, (ii) solving Eq. (3) for λ and substituting it in the expression obtained in step (i), and (iii) substituting Eqs. (4) and (5) in the expression obtained from step (i) and rearranging terms to get

$$\begin{aligned}\dot{I} = {} & \frac{p_I\left[U_{CH}(Y-p_I I,H) - U_C(Y-p_I I,H)F_{IH}(H,I)[F_I(H,I)]^{-1}\right][F(H,I)-\delta H]}{p_I^2 U_{CC}(Y-p_I I,H) + p_I U_C(Y-p_I I,H)F_{II}(H,I)[F_I(H,I)]^{-1}} \\ & + \frac{F_I(H,I)U_H(Y-p_I I,H) - [\rho+\delta-F_H(H,I)]p_I U_C(Y-p_I I,H)}{p_I^2 U_{CC}(Y-p_I I,H) + p_I U_C(Y-p_I I,H)F_{II}(H,I)[F_I(H,I)]^{-1}},\end{aligned} \tag{7}$$

$$\dot{H} = F(H,I) - \delta H,\ H(0) = H_0,\ H(T) = H_T. \tag{8}$$

One can also derive the above differential equations by employing Eqs. (9) and (10) from Caputo (2005, Chapter 18). Note that the arguments of the functions in Eqs. (7) and (8) have been explicitly indicated, seeing as the Jacobian matrix of Eqs. (7) and (8) will be calculated in the ensuing section. Despite the complexity of Eq. (7), a great deal can be said about the local qualitative properties of the solution to Eqs. (7) and (8), as will be demonstrated in the ensuing sections.

In closing the section, note that the steady state version of the necessary and sufficient conditions is found by setting $\dot{I}=0$ and $\dot{H}=0$ in Eqs. (7) and (8), and is given by

$$F_I(H,I)U_H(Y-p_I I,H)-[\rho+\delta-F_H(H,I)]p_I U_C(Y-p_I I,H)=0, \tag{9}$$

$$F(H,I)-\delta H=0. \tag{10}$$

Finally, recall that by assumption (A4), $H=H^{ss}(\boldsymbol{\beta})$ and $I=I^{ss}(\boldsymbol{\beta})$ are the simultaneous solution of Eqs. (9) and (10).

4. Local Stability and the Phase Portrait when $F_H(H,I)\le 0$ and $F_{IH}(H,I)\le 0$

The present section adds the stipulations $F_H(H,I)\le 0$ and $F_{IH}(H,I)\le 0$ to assumptions (A1)–(A4). As discussed in §1, said assumptions are those supported by the medical evidence when investment corresponds to fitness efforts or certain pharmaceutical treatments. The local stability of a steady state and the corresponding phase diagram follow from the properties of the Jacobian matrix of Eqs. (7) and (8) evaluated at the steady state solution $H=H^{ss}(\boldsymbol{\beta})$ and $I=I^{ss}(\boldsymbol{\beta})$. By definition, the Jacobian matrix is given by

$$\mathbf{J} \overset{\text{def}}{=} \left.\begin{bmatrix} \dfrac{\partial \dot{I}}{\partial I} & \dfrac{\partial \dot{I}}{\partial H} \\ \dfrac{\partial \dot{H}}{\partial I} & \dfrac{\partial \dot{H}}{\partial H} \end{bmatrix}\right|_{\substack{H=H^{ss}(\boldsymbol{\beta}) \\ I=I^{ss}(\boldsymbol{\beta})}} = \left.\begin{bmatrix} \underset{(+)}{\dfrac{\partial \dot{I}}{\partial I}} & \underset{(+)}{\dfrac{\partial \dot{I}}{\partial H}} \\ \underset{(+)}{F_I(H,I)} & \underset{(-)}{F_H(H,I)-\delta} \end{bmatrix}\right|_{\substack{H=H^{ss}(\boldsymbol{\beta}) \\ I=I^{ss}(\boldsymbol{\beta})}}, \tag{11}$$

where

$$\left.\frac{\partial \dot{I}}{\partial I}\right|_{\substack{H=H^{ss}(\boldsymbol{\beta}) \\ I=I^{ss}(\boldsymbol{\beta})}} = \left.\frac{p_I^2[\rho+\delta-F_H]U_{CC}+U_H F_{II}}{p_I^2 U_{CC}+p_I U_C F_{II}[F_I]^{-1}}\right|_{\substack{H=H^{ss}(\boldsymbol{\beta}) \\ I=I^{ss}(\boldsymbol{\beta})}} >0, \tag{12}$$

$$\left.\frac{\partial \dot{I}}{\partial H}\right|_{\substack{H=H^{ss}(\beta)\\ I=I^{ss}(\beta)}} = \left.\frac{p_I U_{CH}\left[2[F_H-\delta]-\rho\right]-p_I U_C F_{IH}[F_H-\delta][F_I]^{-1}}{p_I^2 U_{CC}+p_I U_C F_{II}[F_I]^{-1}}\right|_{\substack{H=H^{ss}(\beta)\\ I=I^{ss}(\beta)}} + \left.\frac{p_I U_C F_{HH}+F_I U_{HH}+U_H F_{IH}}{p_I^2 U_{CC}+p_I U_C F_{II}[F_I]^{-1}}\right|_{\substack{H=H^{ss}(\beta)\\ I=I^{ss}(\beta)}} > 0. \tag{13}$$

Observe that $\partial \dot{H}/\partial I\big|_{\substack{H=H^{ss}(\beta)\\ I=I^{ss}(\beta)}} > 0$ follows from assumptions (A2) and (A4), and that the sign of Eq. (12) follows from Eq. (9) and assumptions (A1), (A2), and (A4). Consequently, the sign of the elements in the first column of $\mathbf{J}$ hold regardless of the additional assumptions made in the present section or in §6. In contrast, the sign of $\partial \dot{H}/\partial H\big|_{\substack{H=H^{ss}(\beta)\\ I=I^{ss}(\beta)}} < 0$ follows from the added stipulation $F_H(H,I) \le 0$, while the sign of Eq. (13) follows from assumptions (A1) and (A2), plus both added stipulations. As a result, the sign of the elements in the second column of $\mathbf{J}$ depend on one or both of the additional assumptions introduced in this section. Finally, note that the sign pattern of the Jacobian matrix implies that $\left|\mathbf{J}\right| < 0$, i.e., the steady state is a local saddle point. The following proposition records this basic result.

Proposition 1. *Let assumptions (A1)–(A4) hold in problem (1). If, in addition,* $F_H(H,I) \le 0$ *and* $F_{IH}(H,I) \le 0$*, then the steady state solution of problem (1) is a local saddle point.*

Proposition 1 asserts that if the marginal product of health is not positive and health and investment are technical substitutes or independent for all values of health and investment, then the steady state of the generalized model is a local saddle point. But the conclusion of Proposition 1 also holds under weaker assumptions. To see this, observe that Eqs. (11)–(13) are evaluated at the steady state, implying that Proposition 1 holds if $F_H\left(H^{ss}(\beta), I^{ss}(\beta)\right) \le 0$ and $F_{IH}\left(H^{ss}(\beta), I^{ss}(\beta)\right) \le 0$. That is, if the marginal product of health is not positive and health and investment are technical substitutes or independent at the steady state, then the steady state is a local saddle point. Thus outside of a neighborhood of the steady state the marginal product of health could be positive and health and investment could be technical complements, opposite of what holds at the steady state.

Examination of Eqs. (11)–(13) shows that if $F_H(H,I) \equiv 0$, then $|\mathbf{J}| < 0$. Thus, if the production function is independent of health, then the steady state is a local saddle point. This version of the model represents a slight generalization of the usual model, in that the production function is nonlinear in investment. If in addition to the preceding stipulation it is also assumed that $F_{II}(H,I) \equiv 0$, then $|\mathbf{J}| < 0$, i.e., if the production function is also linear in investment, then the steady state is a local saddle point. Taken together, the assumptions $F_H(H,I) \equiv 0$ and $F_{II}(H,I) \equiv 0$ are those typically employed in the canonical model. These two special cases are sufficiently important and are therefore summarized in the following corollary.

Corollary 1. *Let assumptions (A1)–(A4) hold in problem (1). If, in addition, either (i) $F_H(H,I) \equiv 0$, or (ii) $F_H(H,I) \equiv 0$ and $F_{II}(H,I) \equiv 0$, then the steady state solution of problem (1) is a local saddle point.*

Proposition 1 and Corollary 1 show that the local stability of the steady state is identical in the three versions of problem (1). But this does not imply that the solution to each is identical, as there in general exist quantitative differences in the solutions because of the different assumptions placed on the health production function. In passing, note that the remark immediately following Proposition 1 applies equally as well to Corollary 1.

To construct the phase diagram, it follows from the implicit function theorem that the slope of the $\dot{I} = 0$ isocline in a neighborhood of the steady state in the HI-plane is given by the elements of row 1 of $\mathbf{J}$, in particular, by the negative of the ratio of the element in the second column to that in the first, with an analogous formula for the slope of the $\dot{H} = 0$ isocline from row 2. As a result, the $\dot{I} = 0$ isocline is downward sloping and the $\dot{H} = 0$ isocline is upward sloping in a neighborhood of the steady state. Inasmuch as $F(H,0) = 0$ by assumption (A2), the $\dot{H} = 0$ isocline emanates from the origin. Moreover, as shown in Appendix I, the $\dot{H} = 0$ isocline is upward sloping and strongly convex. Finally, by recalling that the vector field is given by the sign of the elements in $\mathbf{J}$, it is possible to draw the phase diagram corresponding to Eqs. (7) and (8). It is depicted in Figure 1 and discussed in the next section.

5. Qualitative Characterization of the Trajectories when $F_H(H,I) \le 0$ and $F_{IH}(H,I) \le 0$

Begin by considering a simple case, to wit, that of an agent who plans as if death never happens, a possibility emphasized by Strulik (2015, pp. 310–312). If one interprets the lower bound on health as the minimum required for life, then in order for this case to make sense it must be assumed that (i) the initial value of health exceeds the lower bound, i.e., $H_0 > H_T$, and (ii) the steady state value of health exceeds the lower bound, that is, $H^{ss}(\boldsymbol{\beta}) > H_T$. What is more, an additional general necessary transversality condition must be satisfied, as given in Theorem 14.9 of Caputo (2005), namely, $\lim_{t\to+\infty} \Phi(H,I,\lambda)e^{-\rho t} = 0$ along an optimal solution. This condition always holds along a solution that asymptotically approaches a steady state, such as that depicted in Figure 1 by either of the two green trajectories. Furthermore, the sufficient conditions in this case require that another transversality condition hold in the form of an inequality, as given in Theorem 14.4 of Caputo (2005)—it is assumed that said inequality holds. Accordingly, a trajectory that converges to the steady state is indeed the optimal solution to the posed optimal control problem under the present stipulations.

The optimal trajectory in this case is fully determined by whether the initial value of health is greater than or less than the steady state value of health. In the case of an initially relatively healthy agent, i.e., $H_0 > H^{ss}(\boldsymbol{\beta})$, initial investment is relatively low, and health monotonically decreases and investment monotonically increases over time, with the opposite holding for an agent who is initially relatively unhealthy. Thus, under the present stipulations, the health and investment trajectories of agents who plan to live indefinitely are monotonic and move in opposite directions over time. Consequently, agents invest less in their health as they become more healthy, refuting the claim by Zweifel (2012, p. 678) that one of the main results of the health-capital model is " … the prediction of a positive relationship between the (permanent, desired) stock of health and the derived demand for health care." Note, however, that the aforesaid refutable implication is not true in general, as shown by Caputo (2021) and in what follows.

Consider the alternative situation in which an agent plans for a known finite lifetime. In this case an agent ends life with health equal to the minimum required for life, i.e., $H^*(T;\boldsymbol{\theta}) = H_T$, as shown in §3. As was the case above, it must be assumed that $H_0 > H_T$. Furthermore, assume that the health states $\left(H_0, H_T, H^{ss}(\boldsymbol{\beta})\right)$ are ordered as $H_0 > H^{ss}(\boldsymbol{\beta}) > H_T$. Examination of Figure 1 shows that a trajectory like **C** is optimal, as it begins in a health state greater than, and terminates in a health state less than, steady state health. As a result, the optimal time path of investment is not monotonic in this case. In particular, investment increases for an initial period of time, reaches a maximum, and then declines until the terminal health state is reached. Thus health and investment move in opposite directions over time before maximum investment occurs, but then move in the same direction thereafter.

Note, in passing, that if one explicitly accounted for the nonnegativity constraint $I(t) \geq 0$ for all $t \in [0,T]$, then as shown earlier, it is possible that $\lambda^*(T;\boldsymbol{\theta}) = 0$ and $I^*(T;\boldsymbol{\theta}) = 0$. Given the preceding stipulations, this implies that a trajectory like **C** would again be optimal, but it would instead terminate along the horizontal axis with health greater than or equal to the minimum required for life.

Now consider a different ordering of $\left(H_0, H_T, H^{ss}(\boldsymbol{\beta})\right)$, say $H_0 > H_T > H^{ss}(\boldsymbol{\beta})$. In this case there are more possibilities for the optimal trajectory. For example, it is possible for a nonmonotonic trajectory like **C** to be optimal. But the optimal trajectory can also mimic the downward sloping portion of trajectories like **B** or **C**, with health declining and investment rising over time, and thus be monotonic, akin to the trajectory of an initially relatively healthy agent who plans to live forever. Furthermore, the optimal trajectory could resemble the portion of a trajectory like **C** that has a positive slope. In this case the trajectory is one in which health and investment are both monotonically declining.

In each of the preceding cases, the trajectory for health was monotonic, but not necessarily so for investment. To show that this possibility may be reversed for agents who plan for a finite lifetime, consider an ordering of $\left(H_0, H_T, H^{ss}(\boldsymbol{\beta})\right)$ such that $H^{ss}(\boldsymbol{\beta}) > H_0 > H_T$. In this case, both health and investment could be monotonically decreasing over time, as would be the case if the optimal trajectory resembled the upward sloping portions of trajectories like **C** or **D**. But it is also entirely possible that the optimal trajectory might mimic one akin to **D**. In that case, investment decreases and health increases until the maximum state of health occurs, after which health decreases along with investment.

Finally, it is important to recall that the preceding results are conditioned on the assumptions $F_H(H,I) \le 0$ and $F_{IH}(H,I) \le 0$. Consequently, under these stipulations, the generalized model is incapable of explaining the repeated cycling between high and low states of health and the concomitant cycling of investment observed in some people. The matter of cycling and the implied assumptions on the health production function are thus examined in the next section.

6. Cycling and the Health Production Function

Proposition 1 shows that if $F_H(H,I) \le 0$ and $F_{IH}(H,I) \le 0$, then the steady state is a local saddle point, that is, $|\mathbf{J}| < 0$. Consequently, said assumptions cannot lead to the cycling of health and investment. Therefore, in order to generate cycling in the generalized model, at least one of the aforesaid stipulations cannot hold. Furthermore, the eigenvalues of $\mathbf{J}$ must be complex conjugates with nonzero real parts, or equivalently, a steady state must be a local spiral node, if the generalized model is to generate cycling in a neighborhood of the steady state.[1] As a result, the assumption that the steady state is a local spiral node is maintained in the present section. The implications of this assumption are summarized in the ensuing proposition, the proof of which is relegated to Appendix II.

Proposition 2. *Let assumptions (A1)–(A4) hold in problem (1). If, in addition, the steady state solution of problem (1) is a local spiral node, then (i) it is a locally unstable spiral node, (ii) the*

[1] As is well known—see, e.g., Theorem 13.6 in Caputo (2005)—this follows because the method of linearization used here accurately predicts the qualitative nature of a steady state if the eigenvalues of the Jacobian matrix are real and unequal or are complex conjugates with nonzero real parts.

slope of the $\dot{H}=0$ *isocline is positive in a neighborhood of a steady state, (iii) the slope of the* $\dot{I}=0$ *isocline is greater than the slope of the* $\dot{H}=0$ *isocline in a neighborhood of a steady state, (iv)* $F_H\left(H^{ss}(\beta), I^{ss}(\beta)\right)-\delta<0$, *and (v)* $F_{IH}\left(H^{ss}(\beta), I^{ss}(\beta)\right)>0$ *and is bounded from below.*

Proposition 2 provides necessary conditions for a local spiral node in the generalized model. Said conditions include those that permit construction of the phase diagram for this case, as given in Figure 2. The necessary condition $F_H\left(H^{ss}(\beta), I^{ss}(\beta)\right)<\delta$ asserts that the steady state marginal product of health does not exceed the depreciation rate of health. Thus at a steady state, the marginal product of health may be less than or equal to zero—which is one of the sufficient conditions for a local saddle point steady state—or it may be positive but bounded above by the depreciation rate of health. In contrast, the necessary condition that $F_{IH}\left(H^{ss}(\beta), I^{ss}(\beta)\right)>0$ is bounded from below is the opposite of the sufficient condition for a local saddle point steady state. It asserts that investment and health are sufficiently complementary in the health production function if a steady state is a local spiral node.

The precise nature of the optimal trajectory depends on the initial and terminal values of health as well as on planned length of life, just as it did in the case of a saddle point steady state. In general, the typical trajectory spirals clockwise away from the steady state. Moreover, it exhibits time intervals in which health and investment are both decreasing, health is decreasing but investment is increasing, both are increasing, and health is increasing but investment is decreasing. Importantly, this type of solution cannot occur in the archetypal model, as shown by Corollary 1. In other words, by a generalization of the health production function to allow the total and marginal products of investment to depend on health, the reach of the standard model is extended to account for an observed pattern of heath and investment that cannot be rationalized by it.

Recall that the medical evidence presented in §1 supported the claim that the effect on health produced by certain behaviors or interventions aimed at improving health (in general) depends on current health, i.e., the health production function depends on investment in health *and* the state of health. Moreover, said evidence suggests that there is a meaningful qualitative difference between efforts to improve fitness or the use of certain pharmaceutical treatments to improve outcomes, versus treatments to restore health. In particular, the former are limited by the facts that the total and marginal products of health investment are nonincreasing functions of health, while the opposite is true for many interventions aimed at restoring health. Despite the fact that the archetypal health-capital model by construction does not accommodate any of these empirical facts, Proposition 1 and Corollary 1 imply that the local qualitative properties of the archetypal model and the generalized model studied herein are identical when the total and marginal products of investment are nonincreasing functions of health. As such, the archetypal model can rationalize health-related behaviors and outcomes in such circumstances despite not accommodating the underlying medical assumptions within the model. In contrast, observations of cyclical health-related behaviors and outcomes cannot be rationalized by the archetypal model, whereas Proposition 2 shows that the generalized health-capital model can indeed rationalize them and accommodate the implied medical assumptions.

In closing, note that Filed at al., (2003) and Lahti-Koski et al., (2005) presented empirical evidence suggesting that eating habits and weight are subject to cycling in significant parts of studied populations. By Proposition 2, necessary conditions for cycling in the generalized model are that at a steady state, investment and health are sufficiently strong technical complements, and that the marginal product of health is bounded above by the rate at which health depreciates. Health-production technologies that are consistent with these requirements have, indeed, been found among patients treated for conditions where cycling has been observed, namely, eating disorders and overweight and obesity problems. Thus, the generalized model developed herein is capable of rationalizing the observed cycling of eating habits and weight. What is more, cycling is rationalized without invoking rational addiction arguments, as in Dockner and Feichtinger (1993), or specific survival or utility functions, as in Suranovic and Goldfarb (2006).

7. Summary and Conclusion

Despite a half-century of research extending the Grossman (1972) model, one simple and intuitive generalization of it had yet to be contemplated, namely, allowing the health production function to depend on investment *and* health. The resulting generalized model was shown to usefully extend the reach of the prototype version but was no more difficult to understand—indeed, it permitted a qualitative characterization of its solution via phase diagrams. In particular, the generalized model was shown to qualitatively mimic the solution of the prototype model when the total and marginal products of investment were non-increasing functions of health. But unlike the archetype model, the generalized version can rationalize cyclical solutions for health and investment, in which case the steady state value of the marginal product of health was shown to be bounded above by the depreciation rate of health, and health and investment were shown to be sufficiently complementary in the steady state.

8. Appendix I: Slope of the $\dot{H}=0$ Isocline

The $\dot{H}=0$ isocline is implicitly defined by $F(H,I)-\delta H=0$, and seeing as $F_I(H,I)>0$ by assumption (A2), the implicit function theorem gives, say, $I=\hat{I}(H,\delta)$ as its $C^{(2)}$ solution. Applying the chain rule to the identity $F\left(H,\hat{I}(H,\delta)\right)-\delta H\equiv 0$ gives the slope of the $\dot{H}=0$ isocline, to wit, $\partial\hat{I}/H\equiv\left[\delta-F_H\left(H,\hat{I}(H,\delta)\right)\right]/F_I\left(H,\hat{I}(H,\delta)\right)>0$, and a second application yields

$$\begin{aligned}\frac{\partial^2\hat{I}}{\partial H^2}&\equiv\frac{-F_I\left[F_{HH}+F_{HI}\frac{\partial\hat{I}}{\partial H}\right]-[\delta-F_H]\left[F_{IH}+F_{II}\frac{\partial\hat{I}}{\partial H}\right]}{F_I^2}\\&\equiv\frac{F_I^2F_{HH}+2F_I[\delta-F_H]F_{HI}+[\delta-F_H]^2F_{II}}{-F_I^3}>0.\end{aligned}$$

The denominator of the preceding expression is negative, as $F_I(H,I)>0$ by assumption (A2). The numerator is also negative because it is a quadratic form that is constructed with the Hessian matrix of $F(\cdot)$, and the Hessian matrix is negative definite by assumption (A2).

9. Appendix II: Proof of Proposition 2

The characteristic equation for the eigenvalues (γ_1,γ_2) of $\mathbf{J}$ is $\gamma^2-\mathrm{tr}(\mathbf{J})\gamma+|\mathbf{J}|=0$, where $\mathrm{tr}(\mathbf{J})$ is the trace of $\mathbf{J}$, hence

$$\gamma_1,\gamma_2=\tfrac{1}{2}\mathrm{tr}(\mathbf{J})\pm\tfrac{1}{2}\left[[\mathrm{tr}(\mathbf{J})]^2-4|\mathbf{J}|\right]^{\frac{1}{2}}. \quad (14)$$

It follows from Eq. (22) of Caputo (2005, Chapter 18) that $\mathrm{tr}(\mathbf{J})$ is equal to the rate of time preference, that is, $\mathrm{tr}(\mathbf{J}) \stackrel{\text{def}}{=} \partial\dot{I}/\partial I\big|_{\substack{H=H^{ss}(\boldsymbol{\beta})\\ I=I^{ss}(\boldsymbol{\beta})}} + F_H\left(H^{ss}(\boldsymbol{\beta}), I^{ss}(\boldsymbol{\beta})\right) - \delta = \rho > 0$. This can be demonstrated by solving Eq. (9) for $U_H(Y - p_I I, H)$, evaluating it at $\left(H^{ss}(\boldsymbol{\beta}), I^{ss}(\boldsymbol{\beta})\right)$, and then substituting it in $\mathrm{tr}(\mathbf{J})$. As $\mathrm{tr}(\mathbf{J}) = \rho > 0$, the real parts of the eigenvalues are positive and equal to $\frac{1}{2}\rho > 0$, thereby implying that the steady state is a locally unstable spiral node. Moreover, because the eigenvalues of $\mathbf{J}$ are assumed to be complex conjugates with nonzero real parts, it follows from Eq. (14) that $|\mathbf{J}| > \frac{1}{4}\rho^2 > 0$.

Recall from §4 that $\partial\dot{I}/\partial I\big|_{\substack{H=H^{ss}(\boldsymbol{\beta})\\ I=I^{ss}(\boldsymbol{\beta})}} > 0$ and $\partial\dot{H}/\partial I\big|_{\substack{H=H^{ss}(\boldsymbol{\beta})\\ I=I^{ss}(\boldsymbol{\beta})}} > 0$ hold in general under assumptions (A1)–(A4). Because $|\mathbf{J}| > 0$, it follows from Eq. (11) that

$$-\frac{\partial\dot{I}/\partial H}{\underbrace{\partial\dot{I}/\partial I}_{+}}\Bigg|_{\substack{H=H^{ss}(\boldsymbol{\beta})\\ I=I^{ss}(\boldsymbol{\beta})}} > -\frac{\partial\dot{H}/\partial H}{\underbrace{\partial\dot{H}/\partial I}_{+}}\Bigg|_{\substack{H=H^{ss}(\boldsymbol{\beta})\\ I=I^{ss}(\boldsymbol{\beta})}}. \tag{15}$$

By the implicit function theorem, Eq. (15) asserts that the slope of the $\dot{I} = 0$ isocline is greater than the slope of the $\dot{H} = 0$ isocline in a neighborhood of a steady state.

Because the $\dot{H} = 0$ isocline cuts through the origin in the HI-plane, as shown in §4, for there to exist a steady state that satisfies Assumption (A4), the $\dot{H} = 0$ isocline must have a positive slope in a neighborhood of a steady state, i.e., the right-hand side of Eq. (15) must be positive. Seeing as $\partial\dot{H}/\partial I\big|_{\substack{H=H^{ss}(\boldsymbol{\beta})\\ I=I^{ss}(\boldsymbol{\beta})}} > 0$, this implies that $\partial\dot{H}/\partial H\big|_{\substack{H=H^{ss}(\boldsymbol{\beta})\\ I=I^{ss}(\boldsymbol{\beta})}} = F_H\left(H^{ss}(\boldsymbol{\beta}), I^{ss}(\boldsymbol{\beta})\right) - \delta < 0$ from Eq. (15). It then follows from $\partial\dot{I}/\partial I\big|_{\substack{H=H^{ss}(\boldsymbol{\beta})\\ I=I^{ss}(\boldsymbol{\beta})}} > 0$ and Eq. (15) that the slope of the $\dot{I} = 0$ isocline is positive and cuts the $\dot{H} = 0$ isocline from below in a neighborhood of a steady state.

The condition $|\mathbf{J}| > \frac{1}{4}\rho^2 > 0$ places a restriction on $F_{IH}\left(H^{ss}(\boldsymbol{\beta}), I^{ss}(\boldsymbol{\beta})\right)$. To see this, use Eq. (11) to show that $|\mathbf{J}| > \frac{1}{4}\rho^2$ is equivalent to

$$\frac{\partial\dot{I}}{\partial I}\frac{\partial\dot{H}}{\partial H}\Bigg|_{\substack{H=H^{ss}(\boldsymbol{\beta})\\ I=I^{ss}(\boldsymbol{\beta})}} - \frac{\partial\dot{I}}{\partial H}\frac{\partial\dot{H}}{\partial I}\Bigg|_{\substack{H=H^{ss}(\boldsymbol{\beta})\\ I=I^{ss}(\boldsymbol{\beta})}} > \frac{1}{4}\rho^2. \tag{16}$$

Now use Eqs. (11)–(13) to rewrite Eq. (16) in the form

$$
\begin{aligned}
F_{IH}\underbrace{\left[p_I U_C[F_H-\delta]-U_H F_I\right]}_{-} < \underbrace{\tfrac{1}{4}\rho^2\left[p_I^2 U_{CC}+p_I U_C F_{II}[F_I]^{-1}\right]}_{-} + \underbrace{p_I U_C F_I F_{HH}}_{-} + \underbrace{F_I^2 U_{HH}}_{-} \\
+\underbrace{p_I U_{CH}\left[2[F_H-\delta]-\rho\right]F_I}_{-} - \underbrace{p_I^2[\rho+\delta-F_H][F_H-\delta]U_{CC}}_{+} - \underbrace{U_H F_{II}[F_H-\delta]}_{+},
\end{aligned} \tag{17}
$$

where all terms are evaluated at $\left(H^{ss}(\beta), I^{ss}(\beta)\right)$. Equation (17) places a positive lower bound on the magnitude of $F_{IH}\left(H^{ss}(\beta), I^{ss}(\beta)\right)$ and shows that $F_{IH}\left(H^{ss}(\beta), I^{ss}(\beta)\right) > 0$ is necessary seeing as $F_H\left(H^{ss}(\beta), I^{ss}(\beta)\right) < \delta$. *Q.E.D.*

10. References

Baganz, L., Richter, A., Albrecht, K., Schneider, M., Burmester, G.R., Zink, A., Strangfeld, A., 2019. Are Prognostic Factors Adequately Selected to Guide Treatment Decisions in Patients With Rheumatoid Arthritis? A Collaborative Analysis From Three Observational Cohorts. *Semin. Arthritis Rheum.* 48, 976–982.

Batterham, M., Tapsell, L.C., Charlton, K.E., 2016. Baseline Characteristics Associated With Different BMI Trajectories in Weight Loss Trials: A Case for Better Targeting of Interventions. *Eur. J. Clin. Nutr.* 70, 207–211

Bolin, K., Lindgren, B., 2016. Non-Monotonic Health Behaviours—Implications for Individual Health-Related Behaviour in a Demand-for-Health Framework. *J. Health Econ.* 50, 9–26.

Caputo, M.R., 2005. *Foundations of Dynamic Economic Analysis: Optimal Control Theory and Applications*. Cambridge University Press, Cambridge.

Caputo, M.R., 2021. New Insights in the Canonical Model of Health Capital. *Math. Soc. Sci.* 110, 23–33.

Dakanalis A., Bartoli F., Caslini M., Crocamo C., Zanetti M.A., Riva G., Clerici M., Carrà G., 2017. Validity and Clinical Utility of the DSM-5 Severity Specifier for Bulimia Nervosa: Results From a Multisite Sample of Patients Who Received Evidence-Based Treatment. *Eur. Arch. Psychiatry Clin. Neurosci.* 267, 823–829.

Dakanalis, A., Colmegna, F., Riva, G., Clerici, M., 2017. Validity and Utility of the DSM-5 Severity Specifier for Binge-Eating Disorder. *Int. J. Eat. Disord.* 50, 917–923.

Dockner, E.J, Feichtinger, G., 1993. Cyclical Consumption Patterns and Rational Addiction. *Am. Econ. Rev.* 83, 256–263.

Ferguson, B.S., 2000. Interpreting the Rational Addiction Model. *Health Econ.* 9, 587–598.

Forster, M., 2001. The Meaning of Death: Some Simulations of a Model of Healthy and Unhealthy Consumption, *J. Health Econ.* 20, 613–638.

Field, A.E., Austin, S.B., Taylor, C.B., Malspeis, S., Rosner, B., Rockett, H.R., Gillman, M.W., Colditz, G.A., 2003. Relation Between Dieting and Weight Change Among Preadolescents and Adolescents. *Pediatrics* 112, 900–906.

Fournier, J.C., DeRubeis, R.J., Hollon, S.D., et al., 2010. Antidepressant Drug Effects and Depression Severity: A Patient-Level Meta-Analysis. *JAMA.* 303, 47–53.

Grossman, M., 1972. On the Concept of Health Capital and the Demand for Health. *J. Polit. Econ.* 80, 223–255.

Herpertz-Dahlmann, B., Wewetzer, C., Hennighausen, K., Remschmidt, H., 1996. Outcome, Psychosocial Functioning, and Prognostic Factors in Adolescent Anorexia Nervosa as Determined by Prospective Follow-Up Assessment. *J. Youth Adolescence*. 25, 455–471.

Herrmann, E., Gerss, J., Bierer, S., Köpke, T., Bolenz, C., Hertle, L., Wülfing, C., 2009. Pre-Treatment Global Quality of Health Predicts Progression Free Survival in Metastatic Kidney Cancer Patients Treated With Sorafenib or Sunitinib. *J. Cancer Res. Clin. Oncol.* 135, 61–67.

Lahti-Koski, M., Männistö, S., Pietinen, P., Vartiainen, E., 2005. Prevalence of Weight Cycling and Its Relation to Health Indicators in Finland. *Obesity* 13, 333–341.

Levhari, D., Withagen, C., 1992. Optimal Management of the Growth Potential of Renewable Resources. *J. Econ.* 56, 297–309.

Lydecker, J.A., Ivezaj, V., Grilo, C.M., 2020. Testing the Validity and Clinical Utility of the Severity Specifiers for Binge-Eating Disorder for Predicting Treatment Outcomes. *J. Consult. Clin. Psychol.* 88, 172–178.

Montazeri, A., 2009. Quality of Life Data as Prognostic Indicators of Survival in Cancer Patients: An Overview of the Literature From 1982 to 2008. *Health Qual. Life Outcomes.* 102, 1–21.

Quinten, C., Coens, C., Mauer, M., Comte, S., Sprangers, M.A., Cleeland, C., Osoba, D., Bjordal, K., Bottomley, A., EORTC Clinical Groups, 2009. Baseline Quality of Life as a Prognostic Indicator of Survival: A Meta-Analysis of Individual Patient Data From EORTC Clinical Trials. *Lancet Oncol.* 10, 865–871.

Strulik, H., 2015. A Closed-Form Solution for the Health Capital Model. *J. Demogr. Econ.* 81, 301–316.

Suranovic, S.M., Goldfarb, R.S., 2006. A Behavioral Model of Cyclical Dieting. *The Economics of Obesity*. Vol. 17. Emerald Group Limited, 49–78.

Treadway, A.B., 1970. Adjustment Costs and Variable Inputs in the Theory of the Competitive Firm. *J. Econ. Theory* 2, 329–347.

Voutsadakis, I.A., 2017. Obesity and Diabetes as Prognostic Factors in Patients With Colorectal Cancer. *Diabetes Metab. Syndr.* S109–S114.

Warburton, D.E., Nicol, C.W., Bredin, S.S., 2006. Health Benefits of Physical Activity: The Evidence. *CMAJ.* 174, 801–809.

Warburton, D.E.R., Bredin, S.S.D., 2017. Health Benefits of Physical Activity: A Systematic Review of Current Systematic Reviews. *Curr. Opin. Cardiol.* 32, 541–556.

Zerwas, S., Lund, B.C., Von Holle, A., Thornton, L.M., Berrettini, W.H., Brandt, H., Crawford, S., Fichter, M.M., Halmi, K.A., Johnson, C., Kaplan, A.S., La Via, M., Mitchell, J., Rotondo, A., Strober, M., Woodside, D.B., Kaye. W.H., Bulik C.M., 2013. Factors Associated With Recovery From Anorexia Nervosa. *J. Psychiatr. Res.* 47, 972–979.

Zweifel, P., 2012. The Grossman Model After 40 Years. *Eur. J. Health Econ.* 13, 677–682.

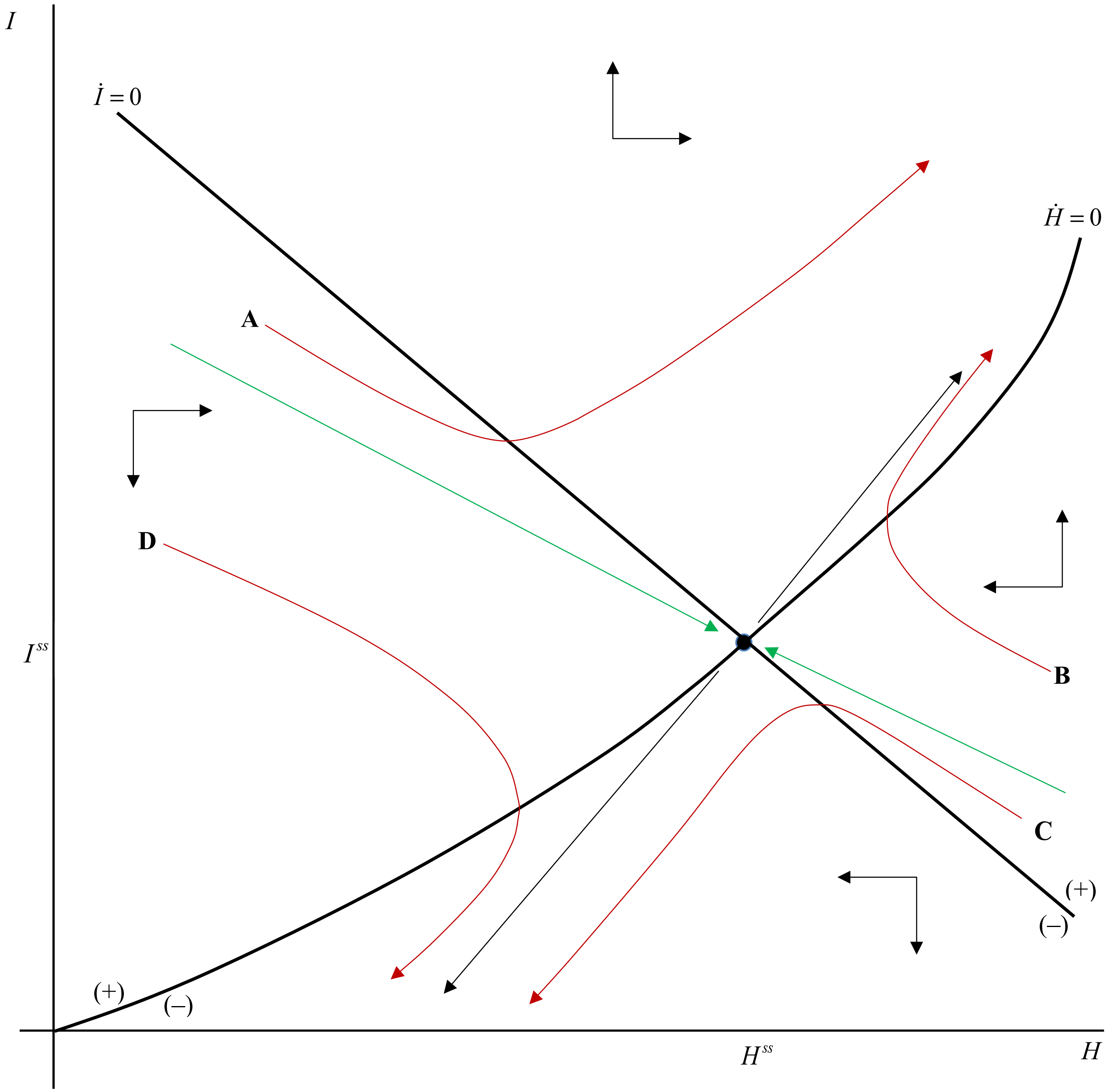

Figure 1

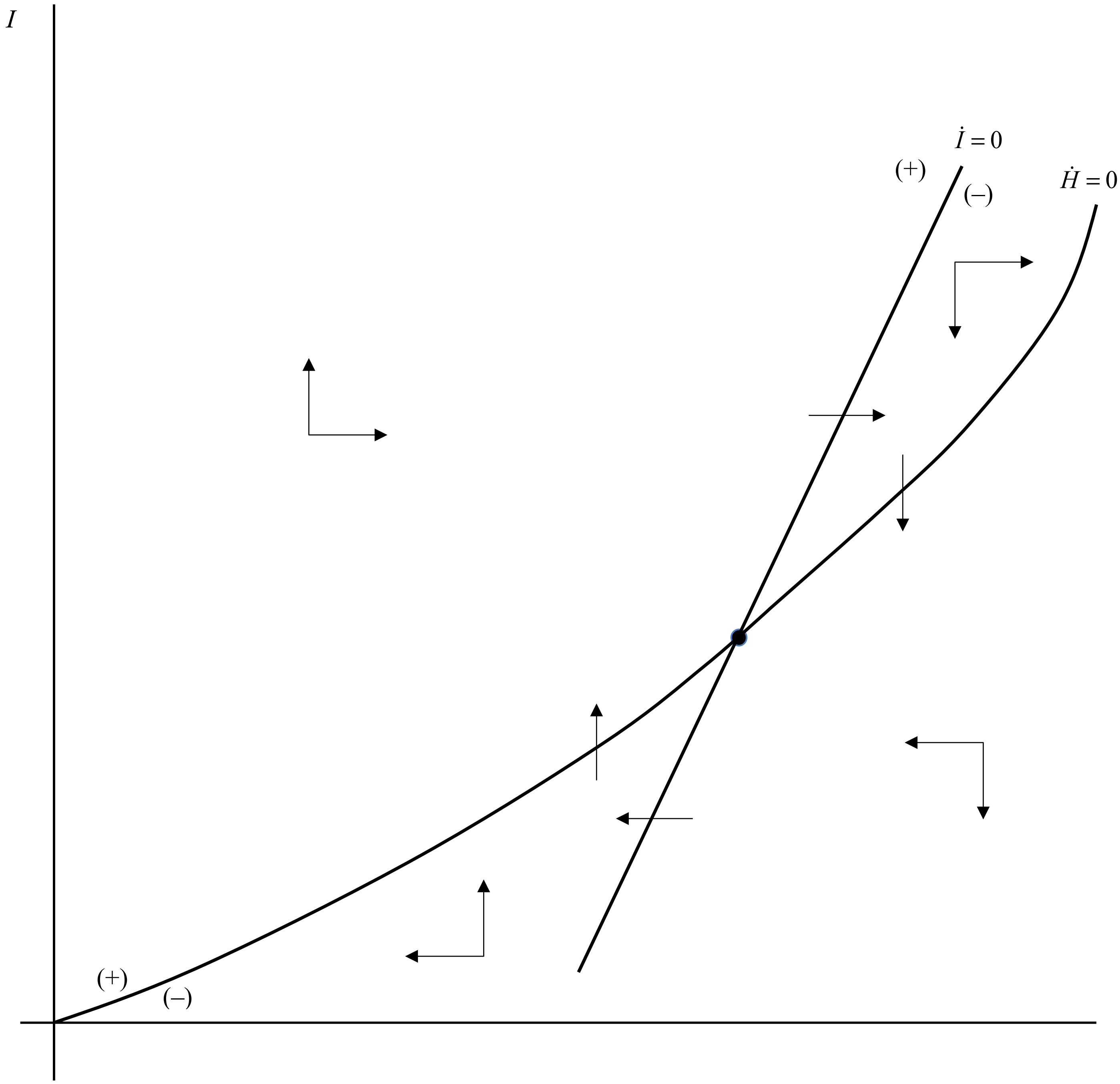

Figure 2