SHOULD THE GROSSMAN MODEL RETAIN ITS ICONIC STATUS IN HEALTH ECONOMICS?

Audrey Laporte


www.canadiancentreforhealtheconomics.ca

February 26, 2014
Should the Grossman model retain its Iconic status in health economics?

JEL Classification: I12

Key words: health economics; grossman model; theoretical framework

Corresponding Author:

Audrey Laporte
Institute of Health Policy, Management, and Evaluation
University of Toronto
155 College St.
Toronto, Canada
Email: audrey.laporte@utoronto.ca
1 Introduction

Ever since 1972, Grossman’s model of investment in health capital has been the cornerstone of the way health economists model health related behavior. Over the years however, a number of criticisms have been aimed at the Grossman model. The major criticisms have recently been set out by Zweifel (2012a), Galama et al. (2012) and Galama and Kapteyn (2011). Zweifel in particular has argued that there are enough sufficiently severe problems with the Grossman model that health economists should look for an alternative starting point. Perhaps the most important of these criticisms are that: 1) the model does not make current health behavior dependent on the past; 2) it does not preclude an individual choosing to live forever; 3) it does not predict that health declines with lower socio-economic status; and 4) the model predicts that there will be a positive relationship between health investment and health status whereas empirically this relationship is typically negative.

In this paper we first outline a simple version of the Grossman model within a dynamic theoretical framework using the tools of optimal control including phase diagram analysis. Phase diagrams allow us to observe the optimal trajectories that the Grossman model predicts individuals will follow over time; they are a qualitative approach that has the advantage over equations alone of making very explicit the dynamics inherent in the Grossman model. This graphical approach will turn out to be very valuable for understanding the criticisms. The alternative approach which consists of finding solutions to the differential equations of the model, while formally correct can obscure the way the individual’s optimal health investment trajectory evolves over time and in

---

1 Zweifel’s editorial prompted a response by Kaestner (2013) to which Zweifel (2013b) responded.
2 Galama and Kapteyn (2011) set the Grossman problem up in an optimal control framework, but work with the present value Hamiltonian. This effectively makes their problem non-autonomous. We work with a current value Hamiltonian, which effectively makes the problem autonomous by removing time as an explicit argument from the problem and therefore allows us to work with phase diagram techniques. Time is obviously still present but has been subsumed into other parts of the problem.
particular can obscure the way the optimal trajectory can change direction as time passes.

Having set up an optimal control representation of a very basic version of the Grossman model, we then turn our attention to addressing the major criticisms leveled at the Grossman model as outlined above. We will show that certain of these criticisms are baseless even in the context of the very simplest version of the model. We will show that the remaining criticisms that have been characterized as fatal structural flaws are in fact merely simplifying assumptions. We discuss alternative ways of relaxing some of those simplifying assumptions in the context of our stripped down model. We also discuss some fruitful directions for future research. In the last section we discuss the implications of the various criticisms we have investigated in terms of the challenge they represent to the validity of the Grossman model and conclude that they do not provide any grounds for removing the Grossman model from its position as a cornerstone of health economics.

2 A bare bones Grossman model

We begin with a very simple version of the Grossman model set within a continuous time optimal control framework\(^3\). The individual’s discounted lifetime utility can be represented by:

\[
\int_0^T U(C, H) e^{-pt} \, dt
\]

The individual’s utility depends on consumption goods C and her stock of health capital H. We assume that her marginal utility with respect to C and H, are positive (i.e. \(U_C, U_H > 0\)) but diminishing, (i.e. \(U_{CC}, U_{HH} < 0\)). We note that the planning horizon starts at \(t=0\) and extends to \(T\). The individual aims to maximize her discounted lifetime utility subject to a set of constraints.

These include the equation of motion for health capital:

\[ \dot{H} = I - \delta H \]  

(2)

\( \dot{H} \) is the change in the individual’s stock of health from period to period. \( H \) is the state variable in the problem because its value over time is governed by the equation of motion (2). We will assume at this point that the production function for health is linear in investment in health, \( I \), and that \( \delta \) the rate of depreciation of the stock of health \( H \) is constant over time. One of the issues which has come up in the literature is whether it is necessary to incorporate decreasing returns to scale into the Grossman model in order to resolve certain criticisms raised by Galama et al. (2012). While it clearly makes sense to allow for decreasing returns to scale in the production of health, we intend to show that certain criticisms that have been tied to the assumption of constant returns are in fact not valid even in the presence of constant returns. We therefore defer the introduction of a non-constant returns production function to a later section of this paper, where we discuss the implications of allowing for a production function for health \( g(I) \) that is non-linear in \( I \), and of allowing \( \delta \) to increase with age.

The equation of motion for health capital (2) makes clear that the individual must take active measures in the form of investment in health activities \( I \) in order to mitigate the effect of \( \delta \) on health which will decline at a proportional rate \( \frac{\dot{H}}{H} = \delta \) when \( I=0 \). The individual’s initial stock of health is \( H(t_0) = H_0 \) and \( H(T)=H_{MIN} \) is the terminal value of the individual’s health stock.

We are treating this problem as an optimal control problem with a fixed, finite horizon \( T \) and a fixed endpoint \( H_{MIN} \). One of the points of contention about the Grossman model in the literature is the question of whether given the assumptions made in Grossman’s original paper, an individual...
could in fact live forever. Grossman defined $H_{MIN}$ as the death level of health capital and the question then is whether by investing enough to keep $H$ above $H_{MIN}$ the individual could postpone death indefinitely. In optimal control terms, that would make the problem an infinite horizon problem. The endpoint of an optimal control problem is determined by a terminal transversality condition. As Ehrlich and Chuma (1990) note, there are conditions that could be invoked to make it optimal for an individual to choose a finite life when the option of an infinite life is open to her.

At this point in our discussion we opt for the much simpler solution of assuming that the individual knows that her life is finite and defining $H_{MIN}$ not as the death level of health but rather as a level of health below which the individual does not wish her health to drop while she is alive. Assuming that individuals treat life as finite does not seem too much of a stretch. We will return to the issue of the implications of assuming a finite $T$ for the individual’s health investment path in a later section of this paper. We also note that the assumption that the end of the horizon is known has been criticized as unrealistic: we will discuss approaches to making the Grossman model stochastic in a later section. We can delay issues pertaining to the endpoint since these simplifying assumptions do not affect the first set of criticisms of the Grossman model that we will address below.

We assume that the individual allocates her income ($Y$) at any point in time between spending on health investment goods, $I$, and spending on consumption goods, $C$. Some formulations of Grossman’s model include an asset accumulation equation, but this is not essential to the workings of the model so we do not include one. The individual’s instantaneous budget constraint is:

$$Y = p_I I + p_C C$$

(3)

We will be working with the consumption version of Grossman’s model so $Y$ will not depend on $H$. 

5
$p_I$ and $p_C$ are the prices of C and I respectively. For simplicity we normalize $p_C$ to equal 1. We can re-write (3) as $C = Y - p_I I$, and substitute this constraint into the utility function.

We will also assume, consistent with Grossman (1972) that I is non-negative:

$$I \geq 0$$

(4)

The current value Hamiltonian for this problem is:

$$\mathcal{H} = U(Y - p_I I, H) + \Psi[I - \delta H] + \lambda I$$

(5)

At each point in time, the individual will choose the level of I, the control variable in the model, so as to maximize (5). $\Psi$ is the co-state variable in the problem, interpreted as the shadow price of health capital, and reflecting the change in maximized lifetime utility when the individual is given an extra unit of health capital. $\lambda$ is the Lagrange multiplier attached to the non-negativity constraint on investment. When $I > 0$, $\lambda = 0$ and when $I = 0$, $\lambda > 0$. We initially assume that $\lambda = 0$ so we have an interior solution and turn to the case where $\lambda > 0$ in a later section.

The Pontryagin necessary condition with respect to I is:

$$\mathcal{H}_1 : -p_I U_C(Y - p_I I, H) + \Psi = 0$$

(6)

We can re-write (6) as $\Psi = p_I U_C(Y - p_I I, H)$. This condition identifies the optimum value of I, and therefore, given the budget constraint, C at each instant. It says that the marginal benefit of an extra unit of health, $\Psi$ must be equal to its marginal cost which is the utility value of the forgone consumption of C type goods $p_I U_C(Y - p_I I, H)$. 
The second of Pontryagin’s necessary conditions is:

\[ \dot{\Psi} = \rho \psi - H \tag{7} \]

or

\[ \dot{\Psi} = \psi [\rho + \delta] - U_H \tag{7a} \]

Where \( \rho \) is the subjective discount rate.

Equation (7) tells us how the shadow price of health evolves along the individual’s optimal trajectory. The evolution of the shadow price of health is clearly a key element in the individual’s health investment decision, but for practical purposes we are much more interested in finding an expression for the evolution of \( I \) itself over time.

We note that equation (6), implicitly defines \( I \) as a function of \( \Psi \) and \( H \). Since equation (6) has to hold for all \( t \), we can use equation (6) along with equation (7a) and equation (2), the equation of motion for \( H \) to derive an equation of motion for \( I \). We do this by differentiating equation (6) with respect to time to give equation (8):

\[ \dot{\Psi} = p_I (-p_I U_{CC} + U_{CH} \dot{H}) = p_I (-p_I U_{CC} \dot{I} + U_{CH} [I - \delta H]) \tag{8} \]

We now have two \( \dot{\Psi} \) equations ((7a) and (8)) that emerge from the first order conditions, so we can equate them and use the definition of \( \Psi \) to obtain the equation of motion for \( I \):

\[ \dot{I} = \frac{p_I U_{CH} [I - \delta H] - p_I U_C [\rho + \delta] + U_H}{p_I^2 U_{CC}} \tag{9} \]

Since we used equation (6) to derive equation (9) the necessary condition for \( I \) will be satisfied along any trajectory that involves equation (9).
One of the advantages of using optimal control as opposed to dynamic programming or calculus of variations is the fact that phase diagram techniques for optimal control theory are well developed. In this paper, we will be working in state control space, meaning that we will represent the evolution of \( H \) and \( I \) over time in a diagram with \( I \) on the vertical and \( H \) on the horizontal axis. The equations of motion for \( H \) and \( I \), will be used to determine how the individual’s optimal trajectory evolves over time. The starting point for a phase diagram is the definition of stationary loci for each of \( I \) and \( H \). The stationary locus for \( I \) has the property that, when we are on it there is no intrinsic tendency for \( I \) to change. In other words, on the stationary locus for \( I \), \( \dot{I} = 0 \). Similarly, on the stationary locus for \( H \), \( \dot{H} = 0 \). Given the stationary loci we can divide \((I,H)\) space into regions where \( I \) and \( H \) are increasing or decreasing. We do this by finding the phase arrows for \( I \) and \( H \) (see Ferguson and Lim (1998) for an introduction to the technique).

Figure 1 is the basic phase diagram corresponding to our Grossman model, showing the stationary loci for \( H \) and \( I \), the phase arrows in the four regions defined by the stationary loci and illustrative trajectories in each region which are consistent with the phase arrows. Obviously we are only interested in considering trajectories that have the potential to reflect the individual’s lifetime utility maximization decision. This means that the Pontryagin necessary conditions must be satisfied at every point along any trajectory that is worth our considering. Because the phase diagram is derived using the first-order conditions, every trajectory in it is a potential optimal trajectory. The trajectory that is actually chosen by the individual depends on the level of health that she starts with and where she wants to end up.

Like virtually all optimal control problems, ours displays saddle-point dynamics. When we have a saddle-point equilibrium (\( E \) in Figure 1), defined as the point of intersection of the stationary loci (because at that point there is no intrinsic tendency for either \( I \) or \( H \) to change), there are
only two trajectories, referred to as the stable branches, that converge to the equilibrium. There are also two trajectories referred to as the unstable branches, which point directly away from the equilibrium. Every other trajectory that could be drawn on the diagram, would initially move toward the equilibrium but eventually turn around and diverge from it. That is why saddle-point dynamics are generally taken to imply the uniqueness of solution trajectories.

We have referred to the intersection of the stationary loci as the equilibrium for the problem. Normally in economic modeling, we take it for granted that the system will either be at or be converging to its equilibrium. In optimal control problems this is certainly true for most macroeconomic applications and for models of economic growth. It is not however true for most microeconomic problems. The reason for this is that it takes an infinite amount of time to reach the equilibrium. Optimal control problems include among their necessary conditions, what are referred to as terminal transversality conditions. For an infinite horizon problem the transversality conditions tell us that whatever the initial value of our state variable, we must pick the control variable so that we are on the stable branch to the equilibrium. For a finite horizon problem, like ours, the transversality condition gives us a different endpoint, so the equilibrium point of the system will not be part of the individual’s optimal trajectory. It makes no sense for us to assume that our optimizing individual is fully informed and forward looking with the one minor exception that she assumes that she will live forever. Economies may live forever, individuals within them do not.

For a finite horizon problem then, we have a number of options for the terminal transversality condition. One is that the stock of the state variable at T equals zero meaning that the stock of the state variable has been used up at the end of the horizon. This condition is often used in models of accumulation of financial assets in the absence of a bequest motive. In some cases it is not possible for the state variable to reach zero in finite time. In those cases the transversality condition is $\Psi$.
at $T$ equals zero, meaning that there is no value to having another unit of the state variable at the end of the problem. The third case is what is known as a fixed endpoint problem in which a specific target value is chosen for the state variable at $T$. This would be the case in a model for example, of financial asset accumulation when the individual has a target bequest motive. To begin with, in what follows we will assume that $H_{MIN}$ is a fixed endpoint target. We will discuss the implications of relaxing this assumption in a later section.

We should note that the term equilibrium has been used in some of the Grossman literature in a sense different from the way we have just defined it, where we have defined it as a steady state. Wagstaff (1986) for example, refers in his equation (4a) to the equilibrium condition for health capital. Wagstaff is working with the investment version of the Grossman model, whereas we are
working with the consumption version but we can derive a counterpart to Wagstaff’s equation (4a).

To do this, recall our equation (7):

\[ \dot{\Psi} = \rho\psi - H \]

(7)

\[ \dot{\Psi} = \rho\psi - [U_H - \psi\delta] \]

(7b)

\[ \dot{\Psi} = [\rho + \delta]\psi - U_H \]

(7c)

Rearranging gives us:

\[ [\rho + \delta]\psi = \dot{\Psi} + U_H \]

(7d)

\[ [\rho + \delta]\psi - \dot{\Psi} = U_H \]

(7e)

\[ \psi[\rho + \delta - \frac{\dot{\Psi}}{\psi}] = U_H \]

(7f)

This is not an equilibrium condition in the sense in which we are using the term since it does not involve a steady state value of H. As we have already noted, the individual will not go to the steady state point in a finite horizon problem. There will however be an optimal level of health at each value of t, where optimal is defined as a value satisfying the necessary conditions conditional on what has happened in the past.\(^4\)

3 Tackling the criticisms

Having set up the basic model we are now in a position to deal with two arguments that have been made which purport to expose fatal weaknesses in the Grossman model. First, Galama et al. (2012) argue that solutions to the current health investment decision lack history in that they do not take account of initial health and Zweifel (2012a) argues that the Grossman model does not...
reflect the behavior of an individual who has suffered a major illness. Second, Galama et al. (2012) argue that the Grossman model does not predict that health will decline faster for individuals of lower socio-economic status.

Lack of path dependence

The issue of whether the solution to the Grossman model takes account of initial health ties into the claim which is made in some of the literature (see overview in Galama et al. 2012) to the effect that the model has a bang-bang solution meaning that the individual can jump instantaneously to her optimal level of health and stay there indefinitely. As we have already noted, a finite horizon problem, which arguably is the only one that makes sense for the individual investment in health problem, does not have a unique permanent optimal level of health. Rather the individual has an optimal trajectory and this optimal trajectory is conditional on her initial inherited stock of health. In other words, the individual cannot choose her initial stock of health. All she can choose is an optimal trajectory conditional on that initial stock. Meaning that, in terms of the model, she takes her initial stock of health as given and chooses the level of health investment I, which puts her on the trajectory which is optimal conditional on her initial stock of health.

In Figure 2 we show the optimal trajectory for an individual born with a high level of initial health \((H_{HIGH})\), a finite life and a target endpoint stock of health \(H_{MIN}\). Given \(H_{HIGH}\), the individual wants to find a trajectory that satisfies the necessary conditions for optimization and that will take her from \(H_{HIGH}\) to \(H_{MIN}\) over an elapsed time of exactly \(T\). She cannot choose her initial value of \(H\)-that is given at the beginning of the problem (e.g. at birth). All she can do is choose I taking account of her utility function and her budget constraint. In Figure 2 because she starts from a high initial level of health, her initial level of I can be relatively low. She then adjusts I throughout...
her planning horizon to control the rate of change of $H$ that is given by equation (2). In Figure 2 she starts with a relatively low value of $I$ after which $I$ increases until her trajectory reaches the stationary locus for $I$, at which point $I$ starts to decrease. $H$ declines continuously in the case that we have illustrated in Figure 2 but at a varying rate. As $I$ increases the rate of decline of $H$ decreases. Given equation (2), the initial value of $H$, and the terminal value of $H$, we see that only one trajectory satisfies the optimality conditions and fits the time horizon.

In this context the argument that has been made in the literature, that an individual could have been born with more than her optimal stock of $H$ makes no sense - a high initial $H$ cannot reduce utility. It would make sense, as we shall see later, to say that, given that her initial stock of health is very high, any non-zero level of health investment would be higher than optimal.
In Figure 3, we add a second individual who is identical to the first with the sole exception that she is born with a much lower initial stock of health ($H_{LOW}$). We see that compared to the first individual, according to the phase arrows, that she will tend to start with a high value of I (point A) causing her stock of health to increase, and as her health increases she can then reduce her level of I. Her stock of H continues to increase until her trajectory reaches the $\dot{H} = 0$ locus at which point both H and I will decrease until she reaches the end of the horizon. We see that the Grossman model predicts that it is optimal for the individual born with a low initial stock of health to front-load her health investment. Contrary to the assertion made by Galama et al. (2012) then, in a standard Grossman model with constant returns to scale, current health status at any value of t is indeed a function of the individual’s initial level of health and her history of prior health investments made.

Figure 3: Low versus high initial health
In Figure 4 we again consider an individual who is born with a high stock of health. She is initially following the trajectory illustrated in Figure 1 but part way through the trajectory she is struck by a significant unanticipated health shock which takes her from health level $H_{HIGH}$ to $H_{LOW}$. At $H_{LOW}$ it is clear that the original level of health investment will no longer be optimal. In terms of a control theory problem a shock like this to the state variable necessitates re-planning. Re-optimizing over the remaining horizon (which may or may not have been affected by the shock), taking as the initial stock of health for the new plan, that stock with which she now finds herself to be endowed at the instant after the illness struck. In this Figure the individual responds to this new lower health state by dramatically increasing her level of investment to point C and following a new optimal trajectory for the remainder of the planning horizon. The amount the upward jump will depend on how long that remaining horizon is and in particular whether the illness has also shortened the individual’s life expectancy. The shorter the remaining horizon, the smaller the expected jump.

Zweifel (2012b) says that the Grossman models predicts “...total investment in health should decrease at least in the case of a serious illness when time to death suddenly becomes short, reducing the present value of returns to investment. Neither prediction is borne out by the data” (pg. 361); meaning that the model predicts a downward jump because of the shortened horizon. We see from Figure 4 that the model can indeed predict an upward jump in investment, $I$, the size of which will depend on the length of the individual’s remaining horizon and also on the magnitude of the cross-partial between health and consumption in the individual’s utility function since that will play a key role in determining the opportunity cost of a jump in $I$.

We see that even in the context of the bare bones version of the Grossman model that individuals respond to their particular health history. In this figure we have assumed that the shock was
completely unanticipated, elsewhere this problem has been set up as a stochastic control problem making use of the Ito derivative for a Poisson process (Laporte and Ferguson (2007)). In that paper, it was shown that the Grossman model can be modified to allow the individual to take account of the probability of a major health shock when she is making her initial health investment plans. Even in that case however, once the shock has occurred the individual re-optimizes.

Socio-economic gradient in health

Galama et al. (2012) argue that the Grossman model is not capable of predicting a socio-economic gradient in health, specifically that it does not allow health to decline faster for individuals with lower socio-economic status.

---

5 One reason that it might appear that the model presented here does not make current decisions depend on past decisions is that the solutions to control theory problems are expressed as open loop rather than as feedback solutions. As we have seen however, from the necessary conditions, the optimal value of I at any time does depend on H and in the event of a shock to H the individual will re-plan taking account of how her realized H differs from the value she had originally anticipated.
lower socio-economic status (SES). Here we take SES to refer to income and consider the case of two individuals born with the same high level of initial health \((H_{HIGH})\), one of which has a higher income at each value of \(t\) than the other. As before, we assume that for each individual, income is constant over time, so we are not dealing with the case where health status can affect income. The two individuals also have the same equation of motion for health so income differences do not mean differences in intrinsic productivity when it comes to producing health. We show these two individuals in Figure 5. Both have the same stationary locus for \(H\) but the individual with a higher income faces a lower opportunity of investment in health in terms of the value of the consumption utility given up for each additional unit of \(I\) purchased. In terms of the phase diagram, the stationary locus for \(I\) for the higher income individual is thus shifted up relative to that of the lower income individual (dashed \(\dot{I} = 0\) locus) and the higher income individual will have an optimal trajectory involving higher initial values of \(I\).

Both individuals start from the same initial health stock and both have the same equation of motion for \(H\), but the fact that the higher income individual will have a higher initial value of \(I\) means that her stock of health will decline more slowly than will that of the lower income individual. As a result, even though the two individuals are born with the same initial health stock we see that the Grossman model predicts that health will decline faster for the individual with lower SES. To see this more clearly, we can rearrange the equation of motion for health (2), to yield:

\[
\frac{\dot{H}}{H} = \frac{I}{H} - \frac{\delta H}{H} = \frac{I}{H} - \delta
\]

Where \(\frac{\dot{H}}{H}\) is the instantaneous rate of change in the stock of health. The two individuals face the same \(\delta\) and given that both individuals start at the same level of \(H\), we note that if \(I = 0\) that
\[
\frac{dH}{dI} = -\delta 
\]
and that if \( I > 0 \) then \( -\delta \) will be a smaller negative number. We see that the individual with higher SES (will have higher \( I \)) and thus will experience a slower rate of decline in her health.

We see that even if two people start with the same level of health an SES gradient emerges from even the simplest Grossman model because of the different choices the individuals will make about \( I \) and the gradient emerges even though we did not assume \( Y(H) \). Contrary then, to the assertion made by Galama et al. (2012) we see that the Grossman model does predict that a high SES individual will have a higher level of \( H \) than the low SES individual.

**The health threshold**

Galama and Kapteyn (2011) say that one criticism of the Grossman model that has not been satisfactorily addressed is the fact that in empirical work \( I \) and \( H \) are generally found to be negatively
related whereas it is said the model predicts a positive relation. They propose modifying the basic
Grossman model to allow for a corner solution in \( I \) i.e. \( I=0 \). They refer to this as Grossman’s
missing health threshold. In this section, we consider both the threshold model and the claim that
the Grossman model predicts a positive relationship between health and investment.

Galama and Kapteyn (2011) set the problem up as we do in the form of an optimal control
problem. They incorporate explicitly a non-negativity constraint on \( I \). It is worth noting that in
his original 1972 paper, Grossman discussed such a non-negativity constraint in some detail but his
calculus of variations based approach did not lend itself to a formal demonstration of the effects
of this assumption. We allowed for the possibility of a non-negativity constraint on \( I \) in the basic
version of the Grossman model above, where we introduced the term \( \lambda I \). When \( I \) is positive, \( \lambda \) will
be equal to 0, so we have to this point been assuming an interior solution. We can see the effect of
allowing the non-negativity constraint on \( I \) to be binding if we modify equation (6) as follows:

\[
-pIUC + \Psi + \lambda = 0
\]  

(6a)

and then rearrange to obtain:

\[
\Psi = pIUC - \lambda
\]  

(6b)

This necessary condition can be interpreted as marginal benefit equals marginal cost where \( \Psi \) (the
shadow price of an additional unit of \( H \)) is the marginal benefit term. When \( I \) is positive and \( \lambda \)
is zero, the right hand side of (6b) is the marginal cost of the investment necessary to yield an
additional unit of \( H \). When \( I=0 \), we note two things about the right-hand side of (6b). One is that
\( C \) will now be equal to income \( Y \), and the other is that \( \lambda \) will be positive. The fact that we have
to subtract \( \lambda \) from the RHS for the equality to hold means that even when \( I=0 \), the LHS, which
is marginal benefit of an additional unit of \( H \), is less than the marginal cost. In other words, \( I=0 \)
when the marginal benefit of another unit of health is so low that we cannot satisfy the first order condition with equality without subtracting the $\lambda$ term from the RHS. Given the definition of $\Psi$, we can actually see two cases where this might arise. One at the beginning of the planning horizon, for someone who is born with what one might call perfect health, so that the benefit from investing in health is virtually zero, which is the case Galama and Kapteyn(2011) appear to have in mind. The other, is the case where $\Psi$ is very low because the remaining time horizon is short so that the payoff period to investing in health is too short to make it worth doing. This would occur at the end of the planning horizon. Thus, it is quite possible that the phase diagram for an individual could look like Figure 6.

Figure 6: The health threshold effect
The relationship between I and H

Galama and Kapteyn (2011) suggest that this threshold effect deals with the criticism raised by Zweifel (2012a) to the effect that whereas the Grossman model predicts a positive relationship between I and H, a great many empirical studies find a negative relationship. However, looking across the range of phase diagrams that we have presented to this point, we can see that in some parts of the optimal trajectory I and H are moving in the same direction and in other parts they move in opposite directions. Zweifel’s argument that the Grossman model predicts a positive relationship between I and H rests on the fact that it includes a production function in which an increase in I leads to increases in H. What this criticism overlooks is that the individual’s optimal choice of I is determined by equation (6) (we are again assuming an interior solution for I). Equation (6) defines I implicitly as a function of H and $\Psi$. If we differentiate equation (6) we find that the partial derivative of I with respect to H conditional on (6) being satisfied is in fact negative. The partial derivative of I with respect to $\Psi$ is positive. The actual trajectory of I will as we have seen, emerge when the equations of motion for H and $\Psi$ are integrated with equation (6). Since the necessary condition for I tells us how I adjusts to changes in H, the assertion that the Grossman model predicts a positive relationship between an individual’s H and her optimal choice of I at any value of t is simply incorrect. Those articles in the literature which find that higher H goes along with lower I (i.e. that healthier people invest less in their health at that point in time) are in fact quite consistent with the optimizing conditions of the model.

Up to this point we have been working with a production function for H that consisted of just I. Galama et al. (2012) suggest it is necessary to replace this with some function $g(I)$ to allow for decreasing returns to scale. In this paper we have argued that the theoretical criticisms which they propose to tackle by this substitution are in fact, erroneous. On the other hand, from an empirical
point of view it makes perfect sense to replace the simple linear function with something that looks
more like a traditional production function. When we do this we should also note, as Grossman
implicitly did, one fundamental difference between the health production function and a traditional
production function. Typically a traditional production function even with decreasing returns to
scale is increasing in inputs. As Grossman notes however, in his discussion of the investment
version of his model, it is not unreasonable to think in terms of the existence of something which
we can call perfect health, a level of health beyond which no increase in inputs can take us. (We
deliberately ignore the question of whether perfect health has been increasing over the centuries.) An
empirical health production function then would ideally take account of the fact that the marginal
productivity of I, depends not just on the level of I but on the level of H. In empirical work where
we are using discrete time structures we might write the health production function as \( g(I_t, H_{t-1}) \)
or perhaps \( g(I_t, H (\text{perfect health}) - H_{t-1}) \) allowing the marginal productivity of I to decrease as H
gets closer to its upper limit. This however, is an empirical issue, not a theoretical one.

Similarly, an empirical rather than a theoretical issue is Zweifel’s criticism that the Grossman
model assumes “a fixed ratio between individuals health care expenditure and the cost of their own
health enhancing efforts regardless of their state of health” (pg. 677 of his editorial). To tackle
this issue we could extend the production function to \( g(I, S, H) \) where I is medical care and S
represents the individual’s other inputs into the production of health. This version of the model
adds another control variable S, and another Pontryagin necessary condition for S and if we were
to phase diagram it we would have to work in state-co-state space rather than state-control space.
However, it is easy enough to show that as in any production problem -at the optimum the ratio of
the marginal products of I and S must equal their price ratios. Zweifel’s criticism is that Grossman
assumed that the production function was constant returns to scale in I and S. The isocost map

22
for a constant returns to scale production function in two inputs is homothetic so that the ratio of I to S will be constant so long as their price ratio remains unchanged regardless of the level of output. Zweifel’s criticism then is that the homotheticity of this isoquant map is unaffected by changes in the individual’s level of H. Again, this is an empirical rather than a theoretical issue. Constant returns to scale is a common simplifying assumption but is not a theoretical prediction of the Grossman model. It seems perfectly sensible that the relative productivities of I and S will vary with the individual’s level of health so that the isoquant map is not homothetic as in the suggestion by Galama et al. (2012) that decreasing returns might make more sense than constant returns. It would seem sensible here to experiment with flexible functional forms for the health production function. Thus we could replace I in the equation of motion for health with g(I), g'(I)>0 and g''(I)<0, so that the first order conditions become:

\[ H_1 : -p_I U_C(Y - p_I H) + \Psi g_I = 0 \]  

Or

\[ \Psi = \frac{p_I U_C}{g_I} \]  

If we were to derive the phase diagrams for this problem we would find that they were qualitatively unchanged from the diagrams presented.

This last point also relates indirectly to comments made by Grossman (2000) about the difficulty of implementing the health capital model empirically. The assumption of constant returns to scale is a simplifying assumption which has the advantage of economizing on observations in that the CRS assumption lets us reduce the number of parameters which must be estimated. Grossman (2000) refers to difficulties associated with estimating forward looking higher order difference equations but one gets the sense that he maybe conflating the individual’s optimization problem
with the econometricians problem of estimating the intrinsic dynamics of health capital and health
investment. Set up in discrete time terms, the Grossman model yields a second order system (see
for example, Jones et al. (2014)). This system can be written either in forward looking or backward
looking terms. If the model has been properly specified, either version should yield the same story
regarding the dynamics about health investment and health capital. The individual will make her
decisions on the basis of her expectations about the future but as we have seen will have to re-plan
if she suffers a major shock. The econometrician estimating a backward looking equation clearly
has to allow for shocks in estimating a difference equation for individual behavior. Grossman refers
to the need for three observations giving the impression that this was a fairly high hurdle when
Grossman (2000) was being written. Since then, as a result of the increased availability of longer
individual level datasets we would seem to have the degrees of freedom necessary to tackle a number
of the empirical issues which have been raised in the Grossman literature. These include moving
beyond more restrictive forms of production e.g. from Cobb-Douglas to translog and the additional

**Finite versus infinite horizon**

One key issue in the literature is whether the Grossman model would actually allow an individual
to choose to live forever. If it does, this would have to be regarded as a major theoretical weakness
since nobody as yet has successfully made that choice. This is an issue which goes all the way back
to Grossman’s original 1972 paper and which he said in Grossman (2000) had been of concern to
him at the time. Grossman’s solution in the 1972 paper was to have the rate of depreciation of
health capital increase at an increasing rate so that at a certain point even with constant returns
to scale in the production function, it became impossible to maintain H above \( H_{MIN} \). As we have
written our stripped down Grossman model, $\delta$ has been constant. Grossman's approach is not at all unreasonable if we assume that the human body simply breaks down at some point at a rate faster than it is possible to repair it\(^6\). That would seem to suggest that technological improvement in the health production function would still open the possibility of immortality. However, if we are going to use the Grossman model as the basis for empirical investigation of human health behavior it does not seem unreasonable to do what we have done in this paper and assume the individual knows that their life is finite. In other words, to work with a fixed finite horizon form of the problem. Zweifel has argued that the assumption that the length of life is known is unreasonable. It is certainly the case that people die at different ages. One possible approach is to define a maximum possible length of life and allow people to choose their length of life up to that upper limit. The transversality condition for an endogenous horizon is that the Hamiltonian equals zero at the optimal end of life. This can clearly be a result of different choices made by different individuals, conditional on their endowments. We can also consider a model in which there is an absolute upper limit to $T$ and the actual age at death is stochastic with the probability of death being a function of the individual’s stock of health capital. This moves us into the realm of stochastic control theory and Itô’s Lemma, techniques which have been little used in health economics despite the fact that health behaviours reflect decisions about inter-temporal optimization under uncertainty. As noted above, one example is Laporte and Ferguson (2007), in which Itô’s Lemma applied to a Poisson process was used to model the effect of the probability of an individual experiencing a major illness in her life on the individual’s health investment decision\(^7\). In that paper, the illness was major but not fatal but it would be possible to extend that analysis to take into account the likelihood of a fatal illness. The key issue associated with the way we make life finite is the implication of the method chosen for end

\(^6\)For a different take on the issue of why death happens, see Robson and Hillard (2007).

\(^7\)Cropper (1977) presents a version of the Grossman model with uncertainty in relation to health where the health shocks are characterized as being minor and having no effect on the stock of health capital (pg. 1277)
of life consumption of health care. This is a case however, where we need to separate the effects of two assumptions, one pertaining to the finite nature of the horizon and the other to the implications of the assumption that is made about the rate of depreciation of health capital. The most obvious combination of these assumptions would appear to be the case where there is an upper bound to the length of life and the depreciation rate also increases at an increasing rate late in life.

**Variable versus fixed delta**

One of the simplifying assumptions typically made in the Grossman literature (although we should note, not by Grossman himself), is that the rate of depreciation of health capital $\delta$, is constant. We have already noted that we can interpret the role of $I$ in our equation of motion for health, in the case of an individual who is born very healthy, as one of controlling the rate at which she allows her health capital to decline over time. That rate of decline depends on $\delta$. If $\delta$ increases late in life, then we would expect $I$ to have to increase in order to prevent $H$ from declining too rapidly. Thus the observation in Zweifel (2013) that health care expenditure is “found to just about explode toward the end of life”, is consistent with a rapidly increasing $\delta$ of the sort assumed by Grossman in his 1972 paper toward the end of life.

From a purely technical point of view making $\delta$ increase with age, given that age increases at exactly the same rate as time passes, turns our optimal control problem from a one state variable to a two state variable problem. To this point we have been using phase diagram analysis to illustrate the arguments we have been making about the workings of the Grossman model, unfortunately, it is extremely difficult, and in many cases impossible to draw a phase diagram for a two-state variable problem. The nearest we could come would be to use a two-stage optimal control problem in which delta was low during the first stage and high during the second stage. This would not
result in a discrete jump in I at the point where we pass from the first to the second stage since the transversality conditions for a two-stage problem would make such a jump sub-optimal. In terms of the phase diagram it is clear that an increase in $\delta$ would rotate the stationary locus for H upward. That by itself would be relatively easy to illustrate. Unfortunately, a change in delta will also shift the stationary locus for I. The result of these two effects is that the overall effect on the optimal trajectory is ambiguous. The most productive approach to introducing age-dependent depreciation rates would appear to be theoretical simulation. This approach is discussed by Koka, Laporte and Ferguson (2014).

4 Conclusion

In this paper we have been dealing with a one-state variable version of the Grossman model rather than a two-state variable version that includes a financial asset accumulation equation. We do this for a couple of reasons. The less important one is that nobody ever actually seems to do much with an asset equation even if they write it into the formal set up of the model. The more important one is that a one-state variable version of the problem allows for the use of the phase diagram technique which is very valuable for revealing the depths of the dynamics present even in a very stripped down version of the Grossman model. It would obviously be of interest to develop a model in which an individual can invest in health, education and financial assets as part of a single lifetime plan, and we leave this to future work. It is however very important to understand the depths of the dynamics present in the simple model before moving on to more complicated ones.

So what does all this mean for the Grossman model? We have argued that the criticisms enumerated by Galama et al. (2012), Galama and Kapteyn (2011), and Zweifel (2012a, 2012b) do...
not in fact constitute a serious indictment of the theoretical structure of Grossman’s 1972 model. In fact, most of the criticisms of the Grossman theoretical structure seem to come down to having looked at an intrinsically dynamic model through static eyes. On the other hand, we see that there is still room to extend the original theoretical structure in a number of directions including in the direction of stochastic control theory. The phase diagrams presented highlight the fact that an individual’s I and H follow an actual trajectory. The implication for empirical work is that we need to use explicit dynamic modeling techniques such as using a system of inter-related difference equations rather than trying to use static equations to estimate what the Grossman model reveals to be a fundamentally dynamic structure. It is our contention that when we apply techniques of dynamic economic analysis that are standard in other areas of economics to the Grossman model we can clearly see that its status as the workhorse of modeling individual health related behaviours is well justified.

5 References


decreasing returns to scale in a health capital model?”, *Health Economics*, 21(9): 1080- 1100.


