Infectious Diseases and Economic Growth

A. Goenka[a], L. Liu[b,], M.H. Nguyen[c,]

[a]*Department of Economics, National University of Singapore, AS2 Level 6, 1 Arts Link, Singapore 117570*
[b]*Department of Economics, Harkness Hall, University of Rochester, Rochester, NY 14627, USA.*
[c] *LERNA-INRA, Toulouse School of Economics, Manufacture des Tabacs, 21 Allée de Brienne, 31000 Toulouse, France.*

Abstract

This paper develops a framework to study the economic impact of infectious diseases by integrating epidemiological dynamics into a neo-classical growth model. There is a two way interaction between the economy and the disease: the incidence of the disease affects labor supply, and investment in health capital can affect the incidence and recuperation from the disease. Thus, both the disease incidence and the income levels are endogenous. The disease dynamics make the control problem non-convex thus usual optimal control results do not apply. We show existence of an optimal solution, continuity of state variables, show directly that the Hamiltonian inequality holds thus establishing optimality of interior paths that satisfy necessary conditions, and of the steady states. There are multiple steady states and the local dynamics of the model are fully characterized. A disease-free steady state always exists, but it could be unstable. A disease-endemic steady state may exist, in which the optimal health expenditure can be positive or zero depending on the parameters of the model. The interaction of the disease and economic variables is non-linear and can be non-monotonic.

Keywords: Epidemiology; Infectious Diseases; Existence of equilibrium, Sufficiency; Health Expenditure; Economic Growth; Bifurcation.

JEL Classification: C61, D51, E13, O41, E32.

---

1. Introduction

This paper develops a theoretical framework to jointly model the determination of income and disease prevalence by integrating epidemiological dynamics into a continuous time neo-classical growth model. It allows us to address the issue of what is the optimal investment in health when there is a two way interaction between the disease transmission and the economy: the incidence of diseases affects the labor force and thus, economic outcomes, while economic choices on investment in health expenditure affect the disease transmission - expenditure in health leads to accumulation of health capital which reduces infectivity to and increases recovery from the disease. In this paper we study what is the best that society can do in controlling the disease transmission by taking into account the externality associated with its spread (see Geoffard and Philipson (1997) and Miguel and Kremer (2004) on externalities of disease transmission). Thus, we look at the social planning problem (see Hall and Jones (2007) which takes a similar approach for non-infectious diseases). We show a steady state with disease prevalence and zero health expenditure could be optimal as it depends on the relative magnitude of marginal product of physical capital investment and health expenditure.

Email addresses: goenka@nus.edu.sg (A. Goenka), lliu18@mail.rochester.edu (L. Liu), mhnguyen@toulouse.inra.fr (M.H. Nguyen)

The key contribution of this paper is that we model both disease dynamics and accumulation of physical and health capital. The existing literature does not simultaneously model these together (see e.g. Bell, et al. (2003), Delfino and Simmons (2000), Geoffard and Philipson (1997), Gersovitz and Hammer (2004), Goenka and Liu (2010), Kremer (1996)). In modeling the interaction between infectious diseases and the macroeconomy, we expect savings behavior to change in response to changes in disease incidence. Thus, it is important to incorporate this into the dynamic model to be able to correctly assess the impact of diseases on capital accumulation and hence, growth and income. As the prevalence of diseases is affected by health expenditure, which is an additional decision to the investment and consumption decision, this has to be modeled as well. Without modeling both physical and health capital accumulation and the evolution of diseases at the same time, it is difficult to understand the optimal response to disease incidence. As the literature does not model both disease dynamics and capital accumulation explicitly, the existing models are like a black-box: the very details of disease transmission and the capital accumulation process that are going to be crucial in understanding their effects and for the formulation of public policy, are obscured. We find that even when the strong assumption of log-linear preferences is made (which is usually invoked to justify fixed savings behavior) there can be non-linear and non-monotonic changes in steady state outcomes.

In order to model the disease transmission explicitly we integrate the epidemiology literature (see Anderson and May (1991), Hethcote (2009)) into dynamic economic analysis. In this paper we examine the effect of the canonical epidemiological structure for recurring diseases - SIS dynamics - on the economy. SIS dynamics characterize diseases where upon recovery from the disease there is no subsequent immunity to the disease. This covers many major infectious diseases such as flu, tuberculosis, malaria, dengue, schistosomiasis, trypanosomiasis (human sleeping sickness), typhoid, meningitis, pneumonia, diarrhoea, acute haemorrhagic conjunctivitis, strep throat and sexually transmitted diseases (STD) such as gonorrhea, syphilis, etc (see Anderson and May (1991)). While this paper concentrates on SIS dynamics, it can be extended to incorporate other epidemiological dynamics. An easy way to understand epidemiology models is that they specify movements of individuals between different states based on some 'matching' functions or laws of motion. Thus, the modeling strategy in the paper can be applied to other contexts such as labor markets with search, diffusion of ideas (Jovanovic and Rob (1989)), etc. In particular, the joint modeling of the non-concave law of motion and capital accumulation in the current paper may be applicable to these models.

As the SIS dynamics are non-concave, care has to be taken in using optimal control techniques. To study optimal solutions there are two sets of problems. First, while the existence of optimal solutions relies on compactness and continuity arguments, this is subtle in continuous time models. We show, under weak assumptions, that the feasible set is weakly compact and state variables are absolutely continuous (Lemma 1). The latter rules out jumps in state and co-state variables in the interior of the feasible set as may happen in non-concave models. We show that convergent sequences are in fact feasible and using concavity of the utility function show that optimal solutions exist (Theorem 1). d'Albis, et al (2008) also have an existence result in an abstract model: our proof is more direct and constructive. Second, to characterize optimal solutions it is usual to study the associated Hamiltonian. However, while the first order conditions (and transversality conditions) of the Hamiltonian are necessary they may not be sufficient. We show directly that for any path where disease are endemic and health expenditures are positive is locally optimal. In particular, the steady states are indeed optimal. This is done by showing that inequality for the maximality of the Hamiltonian holds at the interior paths where the necessary conditions hold, and thus, it also holds at the endemic steady state with positive health expenditures.[1] To check the maximality of the Hamiltonian we can decompose it into two parts: the first depends only on the control variables. As we have concavity in the objective function in control variables, using standard results, the difference between the candidate solution and any other solution is non-negative. The second part depends on the co-state and the state variables. This is helpful as the the non-concavity in the problem arises from the law of evolution of labor only, and we explicitly show this term converges to zero by using a transversality type argument.

In this paper we find a disease-free steady state always exists. It is unique when the birth rate is high. The basic intuition is that healthy individuals enter the economy at a faster rate than they contract the disease so that eventually it dies out even without any intervention. As the birth rate decreases, disease-free steady state undergoes a trans-critical bifurcation and there are multiple steady states. The disease-free steady state still exists but is unstable. An endemic steady state also exists with positive or zero health expenditure depending on the relative magnitude of marginal product of physical capital investment and health expenditure. We show that in an endemic steady state it is socially optimal not to invest in health capital if the discount rate (which indexes longetivity) is sufficiently high or people are very impatient, while there are positive health expenditures if it is low or people are patient. A sufficient condition is provided to guarantee the local saddle-point stability.

This paper sheds light on two strands of recent empirical literature: studies on the relationship between economic variables and disease incidence, and the relationship between income and health expenditure share. The former tries to quantify the impact of infectious diseases on the economy and one important issue is solving the endogeneity of disease prevalence (see Acemoglu and Johnson (2007), Ashraf, et al (2009), Bloom, et al (2009), Young (2005)). Our model, which endogenizes both income and disease incidence, shows that reduced form estimation by assuming a linear relationship is not well justified as non-linearity is an important characteristic of models associated with the disease transmission, and this nonlinearity in disease transmission can become a source of non-linearities in economic outcomes. The latter tries to identify the cause of the changing share of health expenditures. Our findings suggest increase in longevity or decrease in the fertility rate could also generate a positive relationship between income and health expenditure share as observed in the data.

In this paper we abstract away from disease related mortality. This is a significant assumption as it shuts down the demographic interaction. This assumption is made for three reasons. First, several SIS diseases have low mortality so there is no significant loss by making this assumption. Secondly, from an economic modeling point of view, we can use the standard discounted utility framework with a fixed discount rate if there is no disease related

[1]The other two steady states that may exist are essentially neoclassical steady states for which optimality is well known.

mortaility. Thirdly, introducing disease related mortality introduces an additional state variable, population size, and does not permit analysis in per capita terms. In the paper we, however, study the effect of changes in the discount rate on the variables of interest. As discussed in the literature, an increase in longevity reduces discounting, and thus the analysis of varying the discount rate captures some effects of change in mortality.

The paper is organized as follows: Section 2 describes the model and in Section 3 we establish existence of an optimal solution. Section 4 studies the steady state equilibria, Section 5 studies sufficiency conditions and Section 6 contains the stability and bifurcation analysis of the steady states. Section 7 studies the effect on steady states of varying the discount and birth rates, and the last section concludes.

2. The Model

In this paper we study the canonical deterministic SIS model which divides the population into two classes: susceptible (S) and infective (I) (see Figure 1). Individuals are born healthy but susceptible and can contract the disease - becoming infected and capable of transmitting the disease to others, i.e. infective. Upon recovery, individuals do not have any disease conferred immunity, and move back to the class of susceptible individuals. Thus, there is horizontal incidence of the disease so the individuals potentially contract the disease from their peers. This model is applicable to infectious diseases which are absent of immunity or which mutate rapidly so that people will be susceptible to the newly mutated strains of the disease even if they have immunity to the old ones.[2] There is homogeneous mixing so that the likelihood of any individual contracting the disease is the same, irrespective of age. Let S_t be the number of susceptibles at time t, I_t be the number of infectives and N_t the total population size. The fractions of individuals in the susceptible and infected class are $s_t = S_t/N_t$ and $i_t = I_t/N_t$, respectively. Let α be the average number of adequate contacts of a person to catch the disease per unit time or the contact rate. Then, the number of new cases per unit of time is $(\alpha I_t/N_t)S_t$. This is the standard model (also known as frequency dependent) used in the epidemiology literature (Hethcote (2009)). The basic idea is that the pattern of human interaction is relatively stable and what is important is the *fraction of infected people* rather than the total number. If the population increases, the pattern of interaction is invariant. Thus, only the proportion of infectives and not the total size is relevant for the spread of the disease. The parameter α is the key parameter and reflects two different aspects of disease transmission: the biological infectivity of the disease and the pattern of social interaction. Changes in either will change α. The recovery of individuals is governed by the parameter γ and the total number of individuals who recover from the disease at time t is γI_t.

Many epidemiology models assume total population size to be constant when the period of interest is short, i.e. less than a year, or when natural births and deaths and immigration and emigration balance each other. As we are interested in long run effects, we assume that the net population growth rate is non-negative.

Assumption 1. *The birth rate b and death rate d are positive constant scalars with $b \geq d$.*

[2] As there is no disease conferred immunity, there typically do not exist robust vaccines for diseases with SIS dynamics.

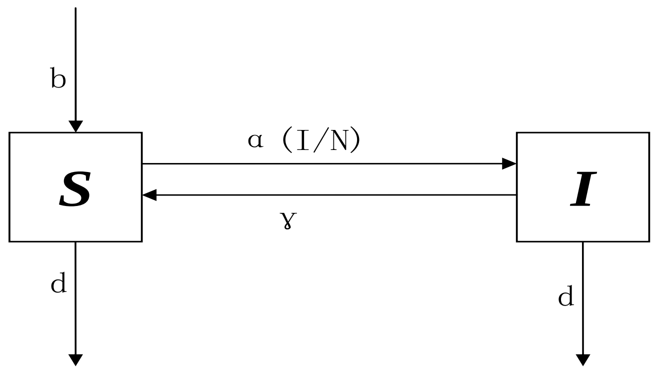

Figure 1: The Transfer Diagram for the SIS Epidemiology Model

Thus, the SIS model is given by the following system of differential equations (Hethcote (2009)):

$$
\begin{aligned}
&dS_t/dt = bN_t - dS_t - \alpha S_t I_t/N_t + \gamma I_t \\
&dI_t/dt = \alpha S_t I_t/N_t - (\gamma + d)I_t \\
&dN_t/dt = (b - d)N_t \\
&S_t, I_t, N_t \geq 0\,\forall t; S_0, I_0, N_0 > 0\,\text{given with } N_0 = S_0 + I_0.
\end{aligned}
$$

Since $N_t = S_t + I_t$ for all t, we can simplify the model in terms of the susceptible fraction s_t:

$$\dot{s_t} = (1 - s_t)(b + \gamma - \alpha s_t) \tag{1}$$

with the total population growing at the rate $b - d$. Note that while it may appear from equation (1) that the dynamics are independent of d, it should be kept in mind that s is the susceptible *fraction* and both the number of susceptibles and the total population depend on d. In this pure epidemiology model, there are two steady states ($\dot{s_t} = 0$) given by: $s_1^* = 1$ and $s_2^* = \frac{b+\gamma}{\alpha}$. We notice s_1^* (the disease-free steady state) exists for all parameter values while s_2^* (the endemic steady state) exists only when $\frac{b+\gamma}{\alpha} < 1$. Linearizing the one-dimensional system around its equilibria, the Jacobians are $Ds|_{s_1^*} = \alpha - \gamma - b$ and $Ds|_{s_2^*} = \gamma + b - \alpha$. Thus, if $b > \alpha - \gamma$ the system has only one disease-free steady state, which is stable; and if $b < \alpha - \gamma$ the system has one stable endemic steady state and one unstable disease-free steady state (See Figure 2). Hence, there is a bifurcation point, i.e. $b = \alpha - \gamma$, where a new steady state emerges and the stability of the disease free steady state changes.[3]

In this paper, we endogenize the parameters α and γ in a two sector growth model. The key idea is that the epidemiology parameters, α, γ, are not immutable constants but are affected by (public) health expenditure. As there is an externality in the transmission of infectious diseases, there may be underspending on private health

[3]Note equation (1) can be solved analytically and these dynamics are global. Since $\dot{s_t} = (1 - s_t)(b + \gamma - \alpha s_t)$, with initial value $s_0 < 1$, is a Bernoulli differential equation, the explicit unique solution is: $s_t = 1 - \frac{e^{[\alpha-(\gamma+b)]t}}{\frac{\alpha}{\alpha-(\gamma+b)}e^{[\alpha-(\gamma+b)]t} + \frac{1}{1-s_0} - \frac{\alpha}{\alpha-\gamma-b}}$ (for $b \neq \alpha - \gamma$) and $s_t = 1 - \frac{1}{\alpha t + \frac{1}{1-s_0}}$ (for $b = \alpha - \gamma$).

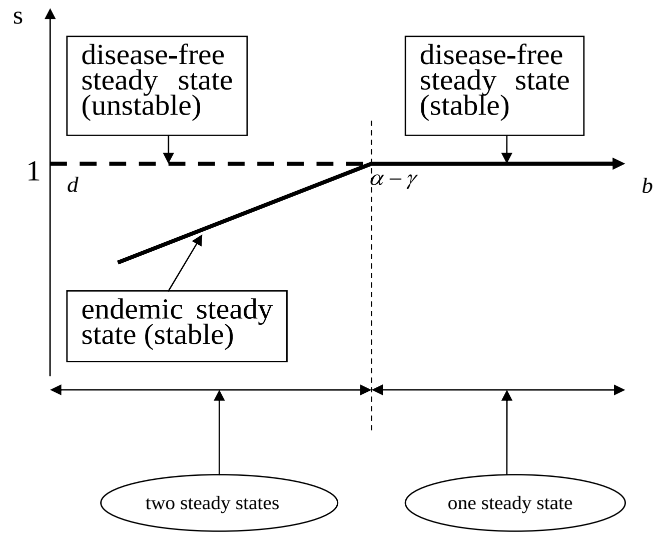

Figure 2: The Steady States, Local Stability and Bifurcation Diagram for SIS Model

expenditure, and due to the contagion effects, private expenditure may not be sufficient to control incidence of the disease.[4] We want to look at the best possible outcome which will increase social welfare. Thus, we study the social planner's problem and concentrate on public health expenditure. In this way, the externalities associated with the transmission of the infectious diseases can be taken into account in the optimal allocation of health expenditure.

We now develop the economic model. There is a population of size N_t growing over time at the rate of $b - d$. Each individual's labor is indivisible: We assume infected people cannot work and labor force consists only of healthy people with labor supplied inelastically.[5] Thus, in time period t the labor supply is $L_t = N_t - I_t = S_t$ and hence, L_t inherits the dynamics of S_t, that is,

$$\dot{l_t} = (1 - l_t)(b + \gamma - \alpha l_t),$$

in terms of the fraction of effective labor $l_t = L_t/N_t$. We allow for health capital to affect the epidemiology parameters, hence, allowing for a two-way interaction between the economy and the infectious diseases. We endogenize them by treating the contact rate and recovery rate as functions of health capital per capita h_t. This takes into account intervention to control the transmission of infectious diseases through their preventive or therapeutic actions. When health capital is higher people are less likely to get infected and more likely to recover from the diseases. We assume that the marginal effect diminishes as health capital increases. We further assume that the marginal effect is finite as health capital approaches zero so that a small public health expenditure will not have a discontinuous effect on disease transmission.

[4]The literature on rational epidemics as in Geoffard and Philipson (1996), Kremer (1996), Philipson (2000) looks at changes in epidemiology parameters due to changes in individual choices. Individual choice is more applicable to disease which transmit by one-to-one contact, such as STDs.

[5]This can be extended to incorporate a partial rather than full loss of productivity due to the illness. Endogenous labor supply could also be introduced and see Goenka and Liu (2012) for details. They show the dynamics are invariant to introduction of endogenous labor supply choice under certain regularity conditions.

Assumption 2. *The epidemiology parameter functions* $\alpha(h_t)$, $\Re_+ \to \Re_+$ *and* $\gamma(h_t)$: $\Re_+ \to \Re_+$ *satisfy:*

1. α *is a* C^∞ *function with* $\alpha' \leq 0$, $\alpha'' \geq 0$, $\lim_{h_t \to 0} |\alpha'| < \infty$, $\lim_{h_t \to \infty} \alpha' = 0$ *and* $\alpha \to \overline{\alpha}$ *as* $h_t \to 0$*;*
2. γ *is a* C^∞ *function with* $\gamma' \geq 0$, $\gamma'' \leq 0$, $\lim_{h_t \to 0} \gamma' < \infty$, $\lim_{h_t \to \infty} \gamma' = 0$ *and* $\gamma \to \underline{\gamma}$ *as* $h_t \to 0$. [6]

We assume physical goods and health are generated by different production functions. The output is produced using physical capital and labor, and is either consumed, invested into physical capital or spent in health expenditure. The health capital is produced only by health expenditure.[7] For simplicity, we assume the depreciation rates of two capitals are the same and $\delta \in (0, 1)$. Thus, the physical capital k_t and health capital h_t are accumulated as follows.

$$\dot{k_t} = f(k_t, l_t) - c_t - m_t - \delta k_t - k_t(b - d)$$
$$\dot{h_t} = g(m_t) - \delta h_t - h_t(b - d),$$

where c_t is consumption and m_t is health expenditure.

The physical goods production function $f(k_t, l_t)$ and health capital production function $g(m_t)$ are the usual neo-classical technologies. The health capital production function is increasing in health expenditure but the marginal product is decreasing. The marginal product is finite as health expenditure approaches zero as discussed above.

Assumption 3. *The production function* $f(k_t, l_t) : \Re^2_+ \to \Re_+$*:*

1. $f(\cdot, \cdot)$ *is* C^∞*;*
2. $f_1 > 0, f_{11} < 0$, $f_2 > 0, f_{22} < 0$, $f_{12} = f_{21} > 0$ *and* $f_{11}f_{22} - f_{12}f_{21} > 0$*;*
3. $\lim_{k_t \to 0} f_1 = \infty$, $\lim_{k_t \to \infty} f_1 = 0$ *and* $f(0, l_t) = f(k_t, 0) = 0$.

Assumption 4. *The production function* $g(m_t) : \Re_+ \to \Re_+$ *is* C^∞ *with* $g' > 0$, $g'' < 0$, $\lim_{m_t \to 0} g' < \infty$, $\lim_{m_t \to \infty} g' = 0$, *and* $g(0) = 0$.

It is worth noting that in our model, h can in principle be unbounded. We further assume utility function depends only on current consumption,[8] is additively separable, and discounted at the rate $\theta > 0$.

Assumption 5. *The instantaneous utility function* $u(c_t) : \Re_+ \to \Re_+$ *is* C^∞ *with* $u' > 0$, $u'' < 0$ *and* $\lim_{c_t \to 0} u' = \infty$.

Given concavity of the period utility function, any efficient allocation will involve full insurance. Thus, consumption of each individual is the same irrespective of health status and we do not need to keep track of individual health histories. So we could look at the optimal solution where the social planner maximizes the discounted utility of the representative consumer:[9]

[6] For analysis of the equilibria C^2 is required and for local stability and bifurcation analysis at least C^5 is required. Thus, for simplicity we assume all functions to be smooth functions.

[7] This health capital production function could depend on physical capital as well. If this is the case, there will be an additional first order condition equating marginal product of physical capital in the two sectors and qualitative result of the paper still hold.

[8] We could instead assume utility function depends on both consumption and leisure. As long as we assume it is separable in consumption and leisure, the social planner's problem is well defined. See Goenka and Liu (2012) for details.

[9] Alternatively instead of maximizing the representative agent's welfare we could maximize the total welfare by using $\int_0^\infty e^{-\theta t} e^{(b-d)t} N_0 u(c_t) dt$ (see the discussion in Arrow and Kurz (1970)). It is equivalent to having lower discounting. The qualitative results of this paper still remain although the optimal allocation may vary slightly.

$$\max_{c,m} \int_0^\infty e^{-\theta t} u(c) dt \tag{P}$$

subject to

$$\dot{k} = f(k,l) - c - m - \delta k - k(b-d) \tag{2}$$

$$\dot{h} = g(m) - \delta h - h(b-d) \tag{3}$$

$$\dot{l} = (1-l)(b + \gamma(h) - \alpha(h)l) \tag{4}$$

$$k \geq 0, h \geq 0, c \geq 0, m \geq 0, 0 \leq l \leq 1 \tag{5}$$

$$k_0 > 0, h_0 \geq 0, l_0 > 0 \text{ given.} \tag{6}$$

It is worthwhile noting here that we have irreversible health expenditure as it is unlikely that the resource spent on public health can be recovered. For simplicity, we drop time subscript t when it is self-evident.

3. Existence of an optimal solution

In the problem we study, the law of motion of the labor force (equation (4)) is not concave reflecting the increasing return of controlling diseases so that the Mangasarian conditions do not apply.[10] In addition the maximized Hamiltonian, H^*, may not be concave as it is possible that $\frac{\partial^2 H^*}{\partial^2 l} > 0$. Thus, the Arrow sufficiency conditions may not apply. To obtain a characterization of optimal solutions we first show that there is a solution to the planning problem We then study steady state solutions to the associated Hamiltonian (Section 4) and then show that these are indeed optimal (Section 5) so that they correspond to the solution shown to exist in this section.

The argument for existence of solutions relies on compactness of the feasible set and some form of continuity of objective function. In continuous time models, the relevant variables may be restricted to lie in L^1 but the problem is that a ball in this space is not compact. Thus, we use the weak topology and show weak compactness of the feasible set which is then used to obtain existence of an optimal solution.[11] We first prove the uniform boundedness of the feasible set that deduces the Lebesgue uniformly integrability. Let us denote by $L^1(e^{-\theta t})$ the set of functions f such that $\int_0^\infty |f(t)| e^{-\theta t} dt < \infty$. Recall that $f_i \in L^1(e^{-\theta t})$ weakly converges to $f \in L^1(e^{-\theta t})$ for the topology $\sigma(L^1(e^{-\theta t}), L^\infty)$ (written as $f_i \rightharpoonup f$) if and only if for every $q \in L^\infty$, $\int_0^\infty f_i q e^{-\theta t} dt$ converges to $\int_0^\infty f q e^{-\theta t} dt$ as $i \to \infty$ (written as $\int_0^\infty f_i q e^{-\theta t} dt \longrightarrow \int_0^\infty f q e^{-\theta t} dt$). When writing $f_i \longrightarrow f$, we mean that for every $t \in [0, \infty)$, $\lim_{i \to \infty} f_i(t) = f(t)$, i.e. there is pointwise convergence.

We make the following assumption:

Assumption 6. *There exists $\kappa \geq 0, \kappa \neq \infty$ such that $-\kappa \leq \dot{k}/k$.*

[10]This can be seen from the Hessian: $\begin{pmatrix} 2\alpha & -\gamma' - \alpha' + 2\alpha' l \\ -\gamma' - \alpha' + 2\alpha' l & (1-l)(\gamma'' - \alpha'' l) \end{pmatrix}$.

[11]d'Albis, et al (2008) also establish existence of an optimal solution in an abstract model which can be applied to our problem. We give here a direct, more constructive proof.

This reasonable assumption implies that it is not possible that the growth rate of physical capital converges to $-\infty$ rapidly and is weaker than those used in the literature (see, e.g. Chichilnisky (1981), LeVan and Vailakis (2003), d'Albis et al. (2008)). LeVan and Vailakis (2003) use this assumption in a discrete-time optimal growth model with irreversible investment: $0 \leq (1-\delta)k_t \leq k_{t+1}$ or $-\delta \leq (k_{t+1}-k_t)/k_t$. $\delta > 0$ is the physical depreciation rate in their model, and thus is equivalent to κ. Let us define the net investment : $\iota = \dot{k} + (\delta + b - d)k = f(k,l) - c - m$. $A.6$ then implies there exists $\kappa \geq 0, \kappa \neq \infty$ such that $\iota + [\kappa - (\delta + b - d)]k \geq 0$. If the standard assumption 2 (v) in Chichilnisky (1981) holds (non-negative investment, $\iota \geq 0$) then $A.6$ holds with $\kappa = \delta + b - d$. Therefore, assuming non-negative investment is stronger than $A.6$ in the sense that κ can take any value except for infinity.

We divide the proof into two lemmas. The first lemma proves the relatively weak compactness of the feasible set. For this we show that the relevant variables are uniformly bounded and hence, are uniformly integrable. As a result we are also able to show that the state variables are absolutely continuous which rules out jumps that may arise in a non-concave problem. The continuity property is important in establishing sufficiency conditions for optimality of steady states. Using the Dunford-Pettis Theorem we then have relatively weak compactness of the feasible set.

Lemma 1. *Let us denote by $\mathcal{K} = \{(c,k,h,l,m,\dot{k},\dot{h},\dot{l})\}$ the feasible set satisfying equations(2)-(6). Then*
i) $\mathcal{K}$ is relatively weak compact in $L^1(e^{-\theta t})$.
ii) State variables k, h, l are absolutely continuous.

Proof. i) Since $\lim_{k\to\infty} f_1(k,l) = 0$, for any $\zeta \in (0,\theta)$ there exists a constant A_0 such that $f(k,1) \leq A_0 + \zeta k$. Hence, we have

$$f(k,l) \leq f(k,1) \leq A_0 + \zeta k. \tag{7}$$

Since $\dot{k} = f(k,l) - c - m - k(\delta + b - d)$, it follows that

$$\dot{k} \leq f(k,l) \leq A_0 + \zeta k.$$

Multiplying by $e^{-\zeta\tau}$ we get $e^{-\zeta\tau}\dot{k} - \zeta k e^{-\zeta\tau} \leq A_0 e^{-\zeta\tau}$. Thus,

$$e^{-\zeta t}k = \int_0^t \frac{\partial(e^{-\zeta\tau}k)}{\partial\tau}d\tau + k_0 \leq \int_0^t A_0 e^{-\zeta\tau}d\tau = \frac{-A_0 e^{-\zeta t}}{\zeta} + \frac{A_0}{\zeta} + k_0.$$

This implies $k \leq \dfrac{-A_0}{\zeta} + \dfrac{(A_0 + k_0\zeta)e^{\zeta t}}{\zeta}$. Thus, there exists a constant A_1 such that

$$k \leq A_1 e^{\zeta t}. \tag{8}$$

Therefore, note that $\zeta < \theta$, $\int_0^\infty k e^{-\theta t}dt \leq \int_0^\infty A_1 e^{(\zeta-\theta)t}dt < +\infty$.

Moreover, since $-\dot{k} \leq \kappa k$ and $\dot{k} \leq A_0 + \zeta k \leq A_0 + \zeta A_1 e^{\zeta t}$ there exists a constant A_2 such that $|\dot{k}| \leq A_2 e^{\zeta t}$.

Thus

$$\int_0^\infty |\dot{k}| e^{-\theta t} dt < \int_0^\infty A_2 e^{(\zeta-\theta)t} dt < +\infty.$$

Because $-\dot{k} \leq \kappa k$ and $c = f(k,l) - \dot{k} - m - \delta k - k(b-d)$, it follows from (7) and (8) that

$$\begin{aligned} c &\leq f(k,l) + k(\kappa - \delta - b + d) \\ &\leq A_0 + (\kappa - \delta - b + d + \zeta)k \\ &\leq A_0 + (\kappa - \delta - b + d + \zeta)A_1 e^{\zeta t}. \end{aligned}$$

Thus, we can choose a constant A_3 large enough such that $c \leq A_3 e^{\zeta t}$ which implies

$$0 \leq \int_0^\infty c e^{-\theta t} dt \leq \int_0^\infty A_3 e^{(\zeta-\theta)t} dt < +\infty.$$

Similarly there exists A_4 such that $m \leq A_4 e^{\zeta t}$ and $m \in L^1(e^{-\theta t})$.

Now we prove $|\dot{h}|, h$ belong to the space $L^1(e^{-\theta t})$.

From Assumption 4, there exists a constant B_1 such that $\dot{h} \leq g(m) \leq B_1 e^{\zeta t}$.

Clearly $h = \int_0^t \dot{h} d\tau + h_0 \leq \int_0^t B_1 e^{\zeta \tau} d\tau + h_0 = \frac{B_1}{\zeta} e^{\zeta t} - \frac{B_1}{\zeta} + h_0$ which means there exist B_2 such that $h \leq B_2 e^{\zeta t}$ or $h \in L^1(e^{-\theta t})$. Moreover, $-\dot{h} \leq (\delta + b - d)h$ because $g(m) \geq 0$. Therefore, $-\dot{h} \leq (\delta + b - d)B_2 e^{\zeta t}$. So $|\dot{h}| \leq B_3 e^{\zeta t}$ with $B_3 = \max\{B_1, (\delta + b - d)B_2\}$. Thus, $|\dot{h}| \in L^1(e^{-\theta t})$.

Obviously, $l \in L^\infty$ and $\lim_{t\to\infty} l e^{-\theta t} = 0$. It follows that

$$\int_0^\infty \dot{l} e^{-\theta t} dt = -l_0 + \theta \int_0^\infty l e^{-\theta t} dt \leq -l_0 + \theta \int_0^\infty e^{-\theta t} dt < +\infty.$$

Finally, we will prove that $|\dot{l}| \in L^1(e^{-\theta t})$. Since $0 \leq l \leq 1$ and $\alpha(h)$ is decreasing, we have

$$\begin{aligned} |\dot{l}| &\leq b + |\gamma(h)| + |\alpha(h)| \\ &\leq b + |\gamma(h)| + |\alpha(0)| \\ &= \gamma(h) + b + \alpha(0). \end{aligned}$$

Since $\lim_{h\to\infty} \gamma'(h) \to 0$, there exists a constant B_4 such that $\gamma(h) \leq B_4 + \zeta h \leq B_4 + \zeta B_2 e^{\zeta t}$. Thus, there exists B_5 such that $|\dot{l}| \leq B_5 e^{\zeta t}$. This implies $|\dot{l}| \in L^1(e^{-\theta t})$. We have proven that $\mathcal{K}$ is uniformly bounded on $L^1(e^{-\theta t})$.

Moreover, $\lim_{a\to\infty} \int_a^\infty k e^{-\theta t} dt \leq \lim_{a\to\infty} \int_a^\infty A_1 e^{(\zeta-\theta)t} dt = 0$. This property is true for other variables in $\mathcal{K}$. Therefore $\mathcal{K}$ satisfies Dunford-Pettis theorem and it is relatively compact in the weak topology $\sigma(L^1(e^{-\theta t}), L^\infty)$.

ii) We have shown that every state variable and their derivatives lie in $L^1(e^{-\theta t})$. Thus, they belong to the Sobolev space, $W_1^1(e^{-\theta t})$. Functions which lie in a Sobolev space are continuous. (Proposition 1, Ashkenazy and Le Van (1999), and Theorem 1, Maz'ja (1985)). A function is absolutely continuous if its derivative is intergrable.

As we have shown that the time derivatives of the state variables lie in $L^1(e^{-\theta t})$, the state variables are absolutely continuous. ■

We know from above that as the feasible set is weakly compact, the control variables and derivatives of state variables weakly converge. The following Lemma shows that in fact, the state variables converge pointwise. In addition, the limit of the sequence of the time derivatives of the state variables, is the time derivative of the limit point of the state variables (Second part of Lemma 2.1). The second part of the Lemma (i.e. 2.2) notes that as we are considering a feasible sequence, the weight $\omega_{i(n)}$ is the *same* for all the variables in the sequence. This fact becomes important in the main existence proof. The continuity of the state variables plays an important role in subsequent properties of the Hamiltonian and is used to characterize optimal solutions including the showing optimality (Section 5). However, this does not play an immediate role in Theorem 1 which shows existence of an optimal solution.

Lemma 2. 1. *For any state variable and its derivative, i.e.* $(x_i, \dot{x}_i) \in \mathcal{K}$*, suppose that* $(x_i, \dot{x}_i) \rightharpoonup (x^*, y)$*. Then* $x_i \longrightarrow x^*$ *as* $i \to \infty$ *and* $y = \dot{x}^*$*.*

2. *There exists a function* $\mathcal{N} : N \to N$ *and a sequence of sets of real numbers* $\{\omega_{i(n)} \mid i = n, ..., \mathcal{N}(n)\}$ *such that* $\omega_{i(n)} \geq 0$ *and* $\sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} = 1$ *such that for any variable* $f_i \in \mathcal{K}$*, the sequence* v_n *defined by* $v_n = \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} f_i \to v^*$ *as* $n \to \infty$*.*

Proof. 1) Since $\mathcal{K}$ is relatively compact in the weak topology $\sigma(L^1(e^{-\theta t}), L^\infty)$, a sequence $(x_i, \dot{x}_i)$ in $\mathcal{K}$ has a subsequence (denoted again by x_i for simplicity of notation) which weakly converges to some limit point (x^*, y) in $L^1(e^{-\theta t})$ as $i \to \infty$.

For any $x_i \in \mathcal{K}$ by hypothesis, $x_i \rightharpoonup x^*$. We first claim that, for $t \in [0, \infty)$, $\int_0^t x_i ds \to \int_0^t x^* ds$. Note that $x_i \rightharpoonup x^*$ for the topology $\sigma(L^1(e^{-\theta t}), L^\infty)$ if and only if for every $q \in L^\infty$, $\int_0^\infty x_i q e^{-\theta t} dt \to \int_0^\infty x^* q e^{-\theta t} dt$.

Pick any t in $[0, \infty)$ and let

$$q(s) = \begin{cases} \frac{1}{e^{-\theta s}} \text{ if } s \in [0, t] \\ 0 \text{ if } s > t. \end{cases}$$

Therefore, $q \in L^\infty$ and we get $\int_0^t x_i ds = \int_0^\infty x_i q e^{-\theta s} ds \to \int_0^\infty x^* q e^{-\theta s} ds = \int_0^t x^* ds$.

Now, given that $x_i \rightharpoonup x^*$ and $\dot{x}_i \rightharpoonup y$. By the claim, for all $t \in [0, \infty)$ we have $\int_0^t \dot{x}_i ds \to \int_0^t y ds$. This implies, for any t, $x_i \to \int_0^t y ds + z_0$. Thus $\int_0^t y ds + x_0 = x^*$. Therefore, $x_i \to x^*$ and $y = \dot{x}^*$.

2) A direct application of Mazur's Lemma. ■

We are now in a position to prove the existence of solution to the social planner's problem. The difficulty is that while we have convergence of the feasible sequences, as only the closure of the feasible set is closed, and not the feasible set, we do not know that the limit point is in fact feasible. This is shown below.

Theorem 1. *Under Assumptions A.1-A.6, there exists a solution to the social planner's problem.*

Proof. Since u is concave, for any $\bar{c} > 0$, $u(c) - u(\bar{c}) \leq u'(\bar{c})(c - \bar{c})$. Thus, if $c \in L^1(e^{-\theta t})$ then $\int_0^\infty u(c) e^{-\theta t} dt$ is well defined as

$$\int_0^\infty u(c) e^{-\theta t} dt \leq \int_0^\infty [u(\bar{c}) - u'(\bar{c})\bar{c}] e^{-\theta t} dt + u'(\bar{c}) \int_0^\infty c e^{-\theta t} dt < +\infty.$$

Let us define $S :\stackrel{def}{=} \sup_{c\in\mathcal{K}} \int_0^\infty u(c)e^{-\theta t}dt$. Assume that $S > -\infty$ (otherwise the proof is trivial). Let $c_i \in \mathcal{K}$ be the maximizing sequence of $\int_0^\infty u(c)e^{-\theta t}dt$ so $\lim_{i\to\infty} \int_0^\infty u(c_i)e^{-\theta t}dt = S$.

Since $\mathcal{K}$ is relatively weak compact, suppose that $c_i \rightharpoonup c^*$ for some c^* in $L^1(e^{-\theta t})$. By Lemma 2, there is a sequence of convex combinations

$$v_n = \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} c_{i(n)} \to c^*, \text{ with } \omega_{i(n)} \geq 0 \text{ and } \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} = 1.$$

Because u is concave, we have

$$\begin{aligned}\limsup_{n\to\infty} u(v_n) &= \limsup_{n\to\infty} u\left(\sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} c_{i(n)}\right) \\ &\leq \limsup_{n\to\infty} \left[u(c^*) + u'(c^*)\left(\sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} c_{i(n)} - c^*\right)\right] = u(c^*).\end{aligned}$$

Since this holds for almost all t, integrate w.r.t $e^{-\theta t}dt$ to get

$$\int_0^\infty \limsup_{n\to\infty} u(v_n)e^{-\theta t}dt \leq \int_0^\infty u(c^*)e^{-\theta t}dt.$$

Using Fatou's lemma we have

$$\limsup_{n\to\infty} \int_0^\infty u(v_n)e^{-\theta t}dt \leq \int_0^\infty \limsup_{n\to\infty} u(v_n)e^{-\theta t}dt \leq \int_0^\infty u(c^*)e^{-\theta t}dt. \tag{9}$$

Moreover, by Jensen's inequality we get

$$\limsup_{n\to\infty} \int_0^\infty u(v_n)e^{-\theta t}dt \geq \limsup_{n\to\infty} \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} \int_0^\infty u(c_{i(n)})e^{-\theta t}dt. \tag{10}$$

But since $\int_0^\infty u(c_{i(n)})e^{-\theta t}dt \to S$, (9) and (10) imply $\int_0^\infty u(c^*)e^{-\theta t}dt \geq S$.

So it remains to show that c^* is feasible (because $\mathcal{K}$ is only relatively weak compact, it is not straightforward that $c^* \in \mathcal{K}$). The task is now to show that there exists some (k^*, l^*, h^*, m^*) in $\mathcal{K}$ such that $(c^*, k^*, l^*, h^*, m^*)$ satisfy (2)-(6).

Consider a feasible sequence $(k_{i(n)}, l_{i(n)}, h_{i(n)}, m_{i(n)})$ in $\mathcal{K}$ associated with $c_{i(n)}$ we have

$$\begin{aligned} c^* &= \lim_{n\to\infty} v_n = \lim_{n\to\infty} \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} c_{i(n)} \\ &= \lim_{n\to\infty} \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} [f(k_{i(n)}, l_{i(n)}) - m_{i(n)} - k_{i(n)}(\delta + b - d) - \dot{k}_{i(n)}] \\ &= \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} [f(\lim_{n\to\infty} k_{i(n)}, \lim_{n\to\infty} l_{i(n)}) - (\delta + b - d) \lim_{n\to\infty} k_{i(n)}] \\ & \quad - \lim_{n\to\infty} \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} \dot{k}_{i(n)} - \lim_{n\to\infty} \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} m_{i(n)}. \end{aligned}$$

According to Lemma 2, there exists k^*, l^* such that $\lim_{n\to\infty} k_{i(n)} = k^*, \lim_{n\to\infty} l_{i(n)} = l^*$.

By Lemma 2, $\dot{k}_{i(n)} \rightharpoonup \dot{k}^*$ and since $m_{i(n)}$ in $\mathcal{K}$, there exists m^* such that $m_{i(n)} \rightharpoonup m^*$. Thus it follows from Lemma 2 that

$$\lim_{n\to\infty} \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} \dot{k}_{i(n)} \to \dot{k}^*, \lim_{n\to\infty} \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} m_{i(n)} \to m^*.$$

Therefore,

$$c^* = f(k^*, l^*) - \dot{k}^* - m^* - \delta k^* - k^*(b - d).$$

Since $\dot{l}_i \rightharpoonup \dot{l}^*$, by Lemma 2, there exists $v_n = \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} \dot{l}_{i(n)} \to \dot{l}^*$ as $n \to \infty$. Thus,

$$\dot{l}^* = \lim_{n\to\infty} \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} \dot{l}_{i(n)} = \lim_{n\to\infty} \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} [(1 - l_{i(n)})(b + \gamma(h_{i(n)}) - \alpha(h_{i(n)}) l_{i(n)})].$$

In view of Lemma 2 , $h_{i(n)} \longrightarrow h^*, l_{i(n)} \longrightarrow l^*$ as $n \to \infty$ and $\gamma(h_{i(n)}), \alpha(h_{i(n)})$ are continuous, we get

$$\begin{aligned} \dot{l}^* &= \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} [(1 - l^*)(b + \gamma(h^*) - \alpha(h^*) l^*)] \\ &= (1 - l^*)(b + \gamma(h^*) - \alpha(h^*) l^*). \end{aligned}$$

Applying a similar argument and using Jensen's inequality yields

$$\begin{aligned} \dot{h}^* &= \lim_{n\to\infty} \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} \dot{h}_{i(n)} = \lim_{n\to\infty} \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} [g(m_{i(n)}) - \delta h_{i(n)} - h_{i(n)}(b - d)] \\ &\leq g(\lim_{n\to\infty} \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} m_{i(n)}) - \lim_{n\to\infty} \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} (\delta + b - d) h_{i(n)} \\ &= g(m^*) - \delta h^* - h^*(b - d). \end{aligned}$$

Thus

$$g(m^*) \geq \delta h^* + h^*(b - d) + \dot{h}^*. \tag{11}$$

Because u is increasing, $c^* = f(k^*, l^*) - \dot{k}^* - m^* - \delta k^* - k^*(b-d)$ should be the maximal value which implies, at the optimum, m^* should be the minimal value. Therefore the constraint (11) should be binding at the optimum since $g(m)$ is increasing.

The proof is done. ■

We have shown that the control variables c, m and derivatives of state variables weakly converge in the weak topology $\sigma(L^1(e^{-\theta t}), L^\infty)$, while the state variables converge pointwise (Lemma 2). The problem is that even if we have a weakly convergent sequence, the limit point may not be feasible. For pointwise convergent sequences, the continuity is all that is necessary to prove the feasibility. Therefore, concavity is not needed for state variables. Theorem 1 shows that the limit point is indeed optimal in the original problem. For weakly convergent sequence, Mazur's Lemma is used to change into pointwise convergence. Jensen's inequality is used to eliminate the convex-combination-coefficients to prove the feasibility. Thus, concavity with respect to control variables is crucial. Our proof is adapted from work of Chichilnisky (1981), Romer (1986) and d'Albis, et al. (2008) to the SIS dynamic model with less stringent assumptions and a nonconvex technology. Chichilnisky (1981) used the theory of Sobolev weighted space and imposed a Caratheodory condition on utility function, Romer (1986) made assumptions that utility function has an integrable upper bound, satisfies a growth condition and d'Albis et al (2008) assumed feasible paths are uniformly bounded and the technology is convex with respect to the control variables.

4. Characterization of Steady States

To analyze the solution to the planner's problem, we look at first order conditions to the planning problem. We know that an optimal solution exists, and the first order conditions of the associated Hamiltonian are necessary (see for example, Caputo (2005), Theorem 14.5). As discussed earlier the usual sufficiency conditions based on some type of concavity may not hold. In this section we study the first order conditions to the associated Hamiltonian and study the steady state solutions. We show for some parameters there is a unique (steady state) solution to the first order conditions. For others, there are multiple steady state solutions. In the next section we establish that the steady state solutions satisfy appropriate sufficiency conditions for optimality.

From the Inada conditions we can rule out $k = 0$, and the constraint $l \geq 0$ is not binding since $\dot{l} = b + \gamma > 0$ whenever $l = 0$. In fact, l is bounded from 0 since when l is small enough, $\dot{l} > 0$. The constraint $h \geq 0$ can be inferred from $m \geq 0$, and hence, can be ignored. Now consider the social planner's maximization problem with irreversible health expenditure $m \geq 0$ and inequality constraint $l \leq 1$. The current value Lagrangian for the optimization problem above is:

$$\begin{aligned}\mathcal{L} &= u(c) + \lambda_1[f(k,l) - c - m - \delta k - k(b-d)] + \lambda_2[g(m) - \\ &\quad - \delta h - h(b-d)] + \lambda_3(1-l)(b + \gamma(h) - \alpha(h)l) + \mu_1(1-l) + \mu_2 m\end{aligned}$$

where λ_1, λ_2, λ_3 are costate variables, and μ_1, μ_2 are Lagrange multipliers. The Kuhn-Tucker conditions and transversality conditions are given by

$$
\begin{aligned}
c: &\quad u'(c) = \lambda_1, && (12)\\
m: &\quad m(\lambda_1 - \lambda_2 g') = 0, \quad m \geq 0, \quad \lambda_1 - \lambda_2 g' \geq 0, && (13)\\
k: &\quad \dot{\lambda_1} = -\lambda_1(f_1 - \delta - \theta - b + d), && (14)\\
h: &\quad \dot{\lambda_2} = \lambda_2(\delta + \theta + b - d) - \lambda_3(1-l)(\gamma' - \alpha' l), && (15)\\
l: &\quad \dot{\lambda_3} = -\lambda_1 f_2 + \lambda_3(\theta + b + \gamma + \alpha - 2\alpha l) + \mu_1, && (16)\\
&\quad \mu_1 \geq 0, \quad 1 - l \geq 0, \quad \mu_1(1-l) = 0, && (17)\\
&\quad \lim_{t\to\infty} e^{-\theta t}\lambda_1 k = 0, \quad \lim_{t\to\infty} e^{-\theta t}\lambda_2 h = 0, \quad \lim_{t\to\infty} e^{-\theta t}\lambda_3 l = 0. && (18)
\end{aligned}
$$

The system dynamics are given by equations (2)-(6) and (12)-(18). If x is a variable, we use $\tilde{x}$ to denote its steady state value.[12] We characterize steady states in terms of exogenous parameters b and θ (See figure 3). Define $\underline{l} := \min\{\frac{b+\gamma}{\overline{\alpha}}, 1\}$, $\underline{k}$ such that $f_1(\underline{k}, \underline{l}) = \delta + b - d + \theta$ and $\overline{k}$ such that $f_1(\overline{k}, 1) = \delta + b - d + \theta$. Clearly $\overline{k} \geq \underline{k}$.

Proposition 1. *Under Assumptions $A.1 - A.6$,*

1. *There always exists a unique disease-free steady state with $\tilde{l} = 1$, $\tilde{m} = 0$, $\tilde{h} = 0$, and $\tilde{k} = \overline{k}$;*
2. *There exists an endemic steady state ($\tilde{l} < 1$) if and only if $b < \overline{\alpha} - \underline{\gamma}$ and there is a solution $(\tilde{l}, \tilde{k}, \tilde{m}, \tilde{h})$ to the following system of equations:*

$$
\begin{aligned}
& l = \frac{\gamma(h) + b}{\alpha(h)} && (19)\\
& f_1(k, l) = \delta + \theta + b - d && (20)\\
& g(m) = (\delta + b - d)h && (21)\\
& m(f_1(k,l) - f_2(k,l) l'_\theta(h) g'(m)) = 0 && (22)\\
& m \geq 0 && (23)\\
& f_1(k,l) \geq f_2(k,l) l'_\theta(h) g'(m), && (24)
\end{aligned}
$$

where we define $l'_\theta(h) := \dfrac{(1-l)(\gamma'(h) - \alpha'(h)l)}{\theta + \alpha(h) - b - \gamma(h)}$.

Proof. From $\dot{l} = 0$ we have either $\tilde{l} = 1$ (disease-free case) or $\tilde{l} = \dfrac{\gamma(\tilde{h}) + b}{\alpha(\tilde{h})} < 1$ (endemic case).

Case 1: $\tilde{l} = 1$. Since $\dot{\lambda_2} = \lambda_2(\delta + b - d + \theta) = 0$, $\tilde{\lambda}_2 = 0$. As g' is finite by assumption, $\tilde{\lambda}_1 - \tilde{\lambda}_2 g' = \tilde{u}' > 0$, which implies $\tilde{m} = 0$ by equation (13). Since $g(0) = 0$, $\tilde{h} = 0$ from equation (3). From $\dot{\lambda_1} = 0$, $\tilde{k} = \overline{k}$. So the model degenerates to the neo-classical growth model. Moreover $\tilde{l} = 1$ exists for all parameter values.

Case 2: $\tilde{l} < 1$. This steady state exists if and only if there exists $\tilde{h} \geq 0$ such that $\tilde{l} = \dfrac{\gamma(\tilde{h}) + b}{\alpha(\tilde{h})} < 1$ and $(\tilde{l}, \tilde{k}, \tilde{m}, \tilde{m})$ is a steady state solution to the dynamical system. For the former, by assumption $A.2$, $\dfrac{\gamma(h) + b}{\alpha(h)}$ is increasing in h. So if $\dfrac{b + \underline{\gamma}}{\overline{\alpha}} < 1$, that is, $b < \overline{\alpha} - \underline{\gamma}$, we can find $\tilde{h} \geq 0$ such that $\tilde{l} < 1$. For the latter, since $\tilde{l} < 1$,

[12] This is to distinguish steady states from optimal paths which are indexed by superscript $*$.

$\mu_1 = 0$. From $\dot{\lambda_2} = 0$ and $\dot{\lambda_3} = 0$, we have:

$$\tilde{\lambda}_2 = \frac{u'(\tilde{c})f_2(\tilde{k},\tilde{l})}{f_1(\tilde{k},\tilde{l})}\frac{(1-\tilde{l})((\tilde{\gamma}'-\tilde{\alpha}')\tilde{l})}{\theta+\alpha(\tilde{h})-b-\gamma(\tilde{h})}$$

So equation (13) could be written as equations (22)-(24). Moreover by letting $\dot{h} = 0$, $\dot{\lambda_1} = 0$ and $\dot{l} = 0$ we have equations (19)-(21). ■

Therefore, the economy has a unique disease-free steady state in which the disease is completely eradicated and there is no need for any health expenditure. In this case, the model reduces to the standard neo-classical growth model. Note that the disease-free steady state always exists. Furthermore, when the birth rate is smaller than $\overline{\alpha}-\underline{\gamma}$, in addition to the disease-free steady state, there exists an endemic steady state in which the disease is prevalent and there is non-negative health expenditure. The L.H.S. of equation (24) is the marginal benefit of physical capital investment while the R.H.S. is marginal benefit of health expenditure. To see this, on the R.H.S. the last term $g'(m)$ is the marginal productivity of health expenditure, the middle term $l'_\theta(h)$ can be interpreted as the marginal contribution of health capital on effective labor supply and the first term $f_2(k,l)$ is the marginal productivity of labor. Essentially we can think there is an intermediate production function which transforms one unit of health expenditure into labor supply through the effect on endogenous disease dynamics. Equations (22)-(24) say that if the marginal benefit of physical capital investment is higher than the marginal benefit of health expenditure, there will be no health expenditure. In summary, either there is a unique disease free steady state (when b is large enough); or if b is small enough, two steady states, one where the diseases is eradicated or where the disease is endemic. In the latter steady state there are positive or zero health expenditures depending on the parameters of the model. All three types of steady states cannot co-exist at the same time in this model.[13]

Next we characterize endemic steady states further.

Assumption 7. $\alpha(\alpha''(\gamma+b)-\gamma''\alpha) > 2\alpha'(\alpha'(\gamma+b)-\gamma'\alpha)$.

By $A.7$ we can show

$$\begin{aligned} l''_\theta(h) &= \frac{\partial l'_\theta(h)}{\partial h} \\ &= -\frac{(\alpha-\gamma-b+\theta)(\alpha-\gamma-b)[\alpha(\alpha''(\gamma+b)-\gamma''\alpha)-2\alpha'(\alpha'(\gamma+b)-\gamma'\alpha)]+\alpha\theta(\alpha'(\gamma+b)-\gamma'\alpha)(\alpha'-\gamma')}{\alpha^3(\alpha-\gamma-b+\theta)^2} \\ &< 0 \end{aligned}$$

From equations (19)-(21), we could write $(\tilde{l},\tilde{k},\tilde{m})$ as a function of h. We have $\tilde{l}(h)$ given by equation (19) with $l'(h) := \dfrac{\partial \tilde{l}(h)}{\partial h} = \dfrac{\gamma'\alpha-(\gamma+b)\alpha'}{\alpha^2} > 0$. $\tilde{m}(h) > 0$ is given by equation (21) with $\dfrac{\partial \tilde{m}(h)}{\partial h} = \dfrac{\delta+b-d}{g'(m)} > 0$. $\tilde{k}(h)$ is determined by equation (20), that is, at the steady state marginal productivity of physical capital equals to the marginal cost. Since f_1 is strictly decreasing and lies in $(0,+\infty)$ for each $\tilde{l}(h)$, we can always find a unique

[13] In an earlier version of the paper under different specifications the two types of disease endemic steady states do co-exist.

$\tilde{k}(h)$ and $\dfrac{\partial \tilde{k}(h)}{\partial h} = -f_{12}\dfrac{\partial \tilde{l}(h)}{\partial h}/f_{11} > 0$. Since $\dfrac{\partial f_2(\tilde{k}(h),\tilde{l}(h))}{\partial h} = \dfrac{f_{11}f_{22}-f_{12}f_{21}}{f_{11}}\dfrac{\partial \tilde{l}(h)}{\partial h} < 0$, $l''_\theta(h) < 0$ and $\dfrac{\partial \tilde{g}'(h))}{\partial h} = g''\dfrac{\partial \tilde{m}(h)}{\partial h} < 0$, the R.H.S. of equation (24) decreases as h increases. That is, we have diminishing marginal product of health capital under $A.7$, which guarantees the uniqueness of the endemic steady state.

From equation (23), there are two cases: $\tilde{m} = 0$ and $\tilde{m} > 0$. The first is termed as the endemic steady state without health expenditure and the second the endemic steady state with health expenditure. For the endemic steady state without health expenditure, $\dot{h} = 0$ implies $\tilde{h} = 0$, and

$$f_1(\underline{k},\underline{l}) \geq f_2(\underline{k},\underline{l})l'_\theta(0)g'(0), \tag{25}$$

where $l'_\theta(0) := \dfrac{(1-\underline{l})(\gamma'(0)-\alpha'(0)\underline{l})}{\theta+\overline{\alpha}-b-\underline{\gamma}}$. Due to the diminishing marginal product of health capital mentioned above, a unique endemic steady state without health expenditure exists if and only if inequality (25) is satisfied. Otherwise an endemic steady state with health expenditure exists.

Proposition 2. *Under $A.1 - A.7$, for each fixed $b \in [d, \overline{\alpha}-\underline{\gamma})$ there exists a unique $\hat{\theta}(b)$ such that:*

1. *If $\theta \geq \hat{\theta}(b)$, there exists a unique endemic steady state without health expenditure with $\tilde{l} = \underline{l}, \tilde{m} = 0, \tilde{h} = 0$, $\tilde{k} = \underline{k}$ and $\tilde{c} = f(\underline{k},\underline{l}) - \delta\underline{k} - \underline{k}(b-d)$.*
2. *If $\theta < \hat{\theta}(b)$, there exists a unique endemic steady state with health expenditure with $(\tilde{l},\tilde{m},\tilde{h},\tilde{k},\tilde{c})$ determined by:*

$$\begin{aligned}
&l = \frac{\gamma(h)+b}{\alpha(h)}\\
&f_1(k,l) = \delta+b-d+\theta\\
&f_2(k,l)l'_\theta(h)g'(m) = \delta+b-d+\theta\\
&g(m) = (\delta+b-d)h\\
&c = f(k,l)-m-\delta k-k(b-d).
\end{aligned}$$

Proof. An endemic steady state without health expenditure exists if and only if equation (25) is satisfied. Fix any $b \in [d,\overline{\alpha}-\underline{\gamma})$, L.H.S. of equation (25) is increasing in θ as $f_1(\underline{k},\underline{l}) = \delta+b-d+\theta$ while the R.H.S. of equation (25) is decreasing in θ. So for each b there exists a unique $\hat{\theta}(b)$ such that $f_1(\underline{k},\underline{l}) = f_2(\underline{k},\underline{l})l'_\theta(0)g'(0)$. Note $\hat{\theta}(b)$ could be non-positive.

Case 1: $\hat{\theta}(b)$ is positive. If $\theta \geq \hat{\theta}(b)$, equation (25) is satisfied and an endemic steady state without health expenditure exists. Otherwise an endemic steady state with health expenditure exists.

Case 2: $\hat{\theta}(b)$ is non-positive. Then equation (25) is satisfied for all $\theta > 0$ and only an endemic steady state without health expenditure exists. With abuse of notation, we define $\hat{\theta}(b) = \max\{\hat{\theta}(b), 0\}$.

If $\theta \geq \hat{\theta}(b)$, $\tilde{m} = 0$ and from equations(19)-(21) we have $\tilde{h} = 0$, $\tilde{l} = \underline{l}$, $\tilde{k} = \underline{k}$ and $\tilde{c} = f(\underline{k},\underline{l}) - \delta\underline{k} - \underline{k}(b-d)$.

If $\theta < \hat{\theta}(b)$, $\tilde{m} > 0$ and equation (24) holds at equality. It implies marginal product of physical capital investment equals the marginal product of health expenditure, and equals to the marginal cost. As $\tilde{l},\tilde{k},\tilde{m}$ could be written as functions of h, we only need to show there always exists a unique solution $\tilde{h}$ to the following equation:

$$f_2(\tilde{k}(h), \tilde{l}(h)) l'_\theta(h) g'(\tilde{m}(h)) = \delta + b - d + \theta$$

Since R.H.S. of the above equation decreases as h increases, $\lim_{h\to\infty} f_2 l'_\theta(h) g'(m) = 0$ and $\lim_{h\to 0} f_2 l'_\theta(h) g'(m) = f_2(\underline{k}, \underline{l}) l'_\theta(0) g'(0) > f_1(\underline{k}, \underline{l}) = \delta + b - d + \theta$, the above equation always has a unique solution. And $\tilde{c} = f(\tilde{k}, \tilde{l}) - \tilde{m} - \delta\tilde{k} - \tilde{k}(b-d) > 0$ due to equation (12) and Inada conditions. That is, under $A.1$-$A.7$ there exists an endemic steady state with health expenditure if $\theta < \hat{\theta}(b)$. ■

So far we have characterized the steady state equilibria in terms of exogenous parameters b and θ. Lets summarize the results here (See Figure 3). Figure 3 resembles Figure 2 but now with more economic meaning built in. First, we look at the birth rate b. If the birth rate is very high and greater than the critical value $\overline{\alpha} - \underline{\gamma}$, there is only one disease-free steady state. This is guaranteed by the pure epidemiology model and diseases are eradicated even without any intervention. The basic intuition is that healthy individuals enter the economy at a faster rate than they contract the disease so that eventually it dies out even without any health expenditure. When the birth rate is low and lies in the range $[d, \overline{\alpha} - \underline{\gamma})$, there are two steady states: disease-free steady state and endemic steady state. In an endemic steady state, diseases are prevalent and there is an option of intervention. Depending on the relative magnitude of marginal product of physical capital investment and health expenditure, investment in health could be either positive or zero. If we fix the parameter birth rate b, which lies in the range where the disease could be prevalent, we find if the discount rate θ lies below the curve $\hat{\theta}(b)$ there is positive health expenditure in controlling diseases, otherwise health expenditure is zero. That is, people with a lower discount rate or who are patient are more likely to invest in health than people with high discount rate or who are impatient. A sufficient condition is provided below to show the curve $\hat{\theta}(b)$ is indeed downward sloping, which is also consistent with the findings in the numerical analysis.

Hence, an endemic steady state without health expenditure is well justified and exists when marginal product of physical capital investment is no less than marginal product of health expenditure. In other words, despite the prevalence of the disease, if marginal product of physical capital investment is greater than marginal product of health expenditure, there will be no investment in health. Thus, the prevalence of the disease is not sufficient (from purely an economic point of view) to require health expenditures. It is conceivable that in labor abundant economies with low physical capital this holds, and thus, we may observe no expenditure on controlling an infectious disease while in other richer economies there are public health expenditures to control it.

The welfare analysis of the steady state equilibria is relatively straightforward. The disease-free steady state is no doubt always better than the endemic steady state as there is full employment and no health expenditure is needed to control the prevalence of infectious diseases. For the two endemic steady states, it does not make too much sense to compare them as they do not coexist.

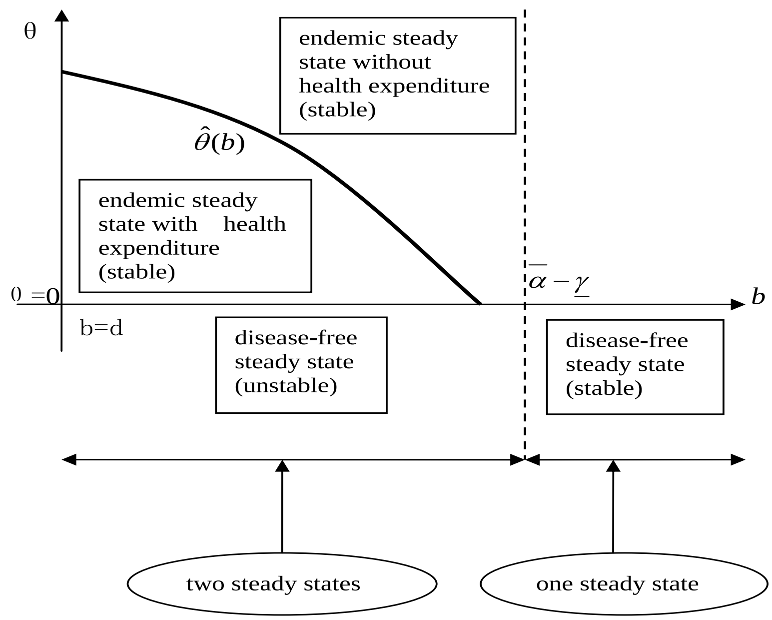

Figure 3: The Steady States, Local Stability and Bifurcation Diagram

4.1. Sufficient Condition for Characterization of Endemic Steady State

This section studies the properties of the function $\hat{\theta}(b)$ for $b \in [d, \overline{\alpha} - \underline{\gamma})$.

Assumption 8. *Elasticity of marginal contribution of health capital on labor supply with respect to birth rate is small, that is,* $\frac{\partial l'_\theta(0)/\partial b}{l'_\theta(0)/b} < b\left[\frac{1}{f_1} - \frac{f_{21}}{f_{11}f_2} - \frac{f_{22}f_{11}-f_{21}f_{12}}{\overline{\alpha}f_{11}f_2}\right]$.[14]

Lemma 3. *$\hat{\theta}(b)$ is decreasing in b. As $b \to \overline{\alpha} - \underline{\gamma}$, $\hat{\theta}(b)$ approaches a non-positive number.*

Proof. Since $\underline{k}$ is given by $f_1(\underline{k}, \underline{l}) = \delta + b - d + \theta$, we have

$$\frac{\partial \underline{k}}{\partial \theta} = \frac{1}{f_{11}} \quad \text{and} \quad \frac{\partial \underline{k}}{\partial b} = \frac{1}{f_{11}} - \frac{f_{12}}{\overline{\alpha} f_{11}}.$$

Moreover, function $\hat{\theta}(b)$ is determined by

$$H = 1 - \frac{f_2(\underline{k}, \underline{l})}{f_1(\underline{k}, \underline{l})} l'_\theta(0) g'(0) = 0.$$

By the implicit function theorem, $\hat{\theta}(b)$ is continuous and

$$\frac{\partial H}{\partial \hat{\theta}} = -\frac{f_{21}f_1 - f_{11}f_2}{f_1^2}\frac{\partial \underline{k}}{\partial \theta} l'_\theta(0) g'(0) - \frac{f_2}{f_1}\frac{\partial l'_\theta(0)}{\partial \theta} g'(0) > 0,$$

[14]For a Cobb-Douglas production function $f(k,l) = Ak^a l^{1-a}$, the assumption reduces to $\frac{\partial l'_\theta(0)/\partial b}{l'_\theta(0)/b} < \frac{b}{(1-a)(\delta+b-d+\theta)}$. As $\frac{\partial l'_\theta(0)}{\partial b} = -\frac{l'_\theta(0)}{\overline{\alpha}(1-\underline{l})} - \frac{(1-\underline{l})\alpha'(0)}{\overline{\alpha}(\theta+\overline{\alpha}-b-\underline{\gamma})} + \frac{l'_\theta(0)}{\theta+\overline{\alpha}-b-\underline{\gamma}}$, the assumption is then given by $-\frac{1}{\overline{\alpha}(1-\underline{l})} - \frac{\alpha'(0)}{\overline{\alpha}(\gamma'(0)-\alpha'(0)\underline{l})} + \frac{1}{\theta+\overline{\alpha}-b-\underline{\gamma}} < \frac{1}{(1-a)(\delta+b-d+\theta)}$, which is satisfied for a wide range of parameter values.

and

$$\frac{\partial H}{\partial b} = -\left(\frac{\overline{\alpha} f_{21} + f_{22} f_{11} - f_{21} f_{12}}{\overline{\alpha} f_{11} f_1} - \frac{f_2}{f_1^2}\right) l'_\theta(0) g'(0) - \frac{f_2}{f_1}\frac{\partial l'_\theta(0)}{\partial b} g'(0) > 0$$

under $A.8$. Thus, we have $\partial\hat{\theta}/\partial b < 0$, that is $\hat{\theta}(b)$ is decreasing in b.

Let $b \to \overline{\alpha} - \underline{\gamma}$. For any $\theta > 0$, $\underline{l} \to 1$, $l'_\theta(0) \to 0$ and R.H.S. of equation (25) goes to 0. However, the L.H.S. of equation (25) equals to $\delta + b - d + \theta$, which is strictly positive as b approaches $\overline{\alpha} - \underline{\gamma}$. So as $b \to \overline{\alpha} - \underline{\gamma}$, equation (25) is satisfied for all $\theta > 0$, which means $\hat{\theta}(b)$ goes to some non-positive number as $b \to \overline{\alpha} - \underline{\gamma}$. ■

From Figure 3, it is easy to see the graph $\hat{\theta}(b)$ intersects the horizontal axis at the point which lies on the left side of $b = \overline{\alpha} - \underline{\gamma}$. Let us denote $\hat{\theta}(d)$ as the intersection point of both the function $\hat{\theta}(b)$ and vertical axis $b = d$. As the function $\hat{\theta}(b)$ is a one-to-one mapping, we could write its inverse mapping as $\hat{b}(\theta)$ for $\theta \in (0, \hat{\theta}(d)]$ and define $\hat{b}(\theta) = d$ for $\theta > \hat{\theta}(d)$.

Proposition 3. *Under Assumption $A.1 - A.8$, for each $\theta > 0$*

1. *If $\hat{b}(\theta) \leq b < \overline{\alpha} - \underline{\gamma}$, a unique endemic steady state without health expenditure exists;*
2. *If $d \leq b < \hat{b}(\theta)$, a unique endemic steady state with health expenditure exists.*

Proof. Omitted. ■

This proposition is easily seen in Figures 5 and 8.

Remark 1. *We have shown in Lemma 1 (ii) that state variables k, h, l are continuous. Following Proposition 3 in Ashkenazy and Le Van (1999), we can show that the control variables c, m are continuous, and following Le Van et al, (2007), in the interior, co-state variables are also continuous. So the problem of jumps at "junction points" does not arise at interior steady states.*

If jumps do take place, then this can happen only at boundary solutions, specifically when $l = 1$. The steady state where the disease is eradicated, $\tilde{l} = 1, \tilde{m} = 0$, $\tilde{h} = 0$, and $\tilde{k} = \overline{k}$ can be shown to be the optimal steady state as it corresponds to the neoclassical steady state with convex technology.

5. Sufficient conditions

We have shown that a solution to the social planner's problem exists, and we know that the first order conditions are necessary. We can have three types steady states. Optimality of the disease free steady state is not in question as it is the neoclassical steady state. When the disease is endemic there can either be a steady state where there are no health expenditure or one where are positive health expenditures. Only one of these exists for any set of parameters (Proposition 2). Optimality of the first kind of steady state is also not an issue as in the neighborhood of the steady state, $m = 0, h = 0$. Hence, it is just a neo-classical steady state with a smaller effective labor force, $\tilde{l} < 1$, and locally the Arrow conditions hold. The steady sate of most interest, i.e. where $\tilde{l} < 1, \tilde{m} > 0, \tilde{h} > 0$, is the problematic one. Thus, we show directly that it satisfies the inequality for the maximality of the Hamiltonian at the endemic steady state with positive health expenditures, and hence the steady state is optimal. First, we prove for *any* interior path which satisfies the first order and transversality condition, it is locally optimal. As we are considering any interior path we need the following assumption.

Assumption 9. *For all $l \in (0,1)$, $f_1(\hat{k}, l) > (b - d + \delta)$, where $\hat{k}$ is the maximum sustainable capital stock for the given l, i.e. $\hat{k}$ solves $f(\hat{k}, l) = (b - d + \delta)\hat{k}$.*

Assumption 3 implies existence of the maximum sustainable capital stock, $\hat{k}$.

Proposition 4. *Consider an interior path with endemic diseases with positive health expenditure: Let denote $x_t^* = (k_t^*, h_t^*, l_t^*)$ where $x_0^* = (k_0, h_0, l_0)$ and $z_t^* = (c_t^*, m_t^*)$. If there exist $\lambda_t = (\lambda_{1,t}, \lambda_{2,t}, \lambda_{3,t}) : [0, \infty) \rightarrow R^n$ that are absolutely continuous and satisfy conditions (2)-(6) and (12)-(18) then (x^*, z^*) is a locally optimal solution of (P).*

Proof.

Consider the current value Hamiltonian

$$
\begin{aligned}
H(x_t, u_t, \lambda_t) \quad = \quad & u(c_t) + \lambda_{1,t}[f(k_t, l_t) - c_t - m_t - \delta k_t - k_t(b-d)] + \lambda_{2,t}[g(m_t) - \\
& - \delta h_t - h_t(b-d)] + \lambda_{3,t}(1-l_t)(b + \gamma(h_t) - \alpha(h_t) l_t)
\end{aligned}
$$

where the first-order and transversality conditions satisfied at $(x_t^*, z_t^*, \lambda_t)$

$$u'(c_t) = \lambda_{1,t} = \lambda_{2,t} g', \tag{26}$$

$$\dot{\lambda_{1,t}} = -\lambda_{1,t}(f_1 - \delta - \theta - b + d), \tag{27}$$

$$\dot{\lambda_{2,t}} = \lambda_{2,t}(\delta + \theta + b - d) - \lambda_{3,t}(1 - l_t)(\gamma' - \alpha' l_t), \tag{28}$$

$$\dot{\lambda_{3,t}} = -\lambda_{1,t} f_2 + \lambda_{3,t}(\theta + b + \gamma + \alpha - 2\alpha l_t), \tag{29}$$

$$\lim_{t\to\infty} e^{-\theta t} \lambda_{1,t} k_t^* = 0, \ \lim_{t\to\infty} e^{-\theta t} \lambda_{2,t} h_t^* = 0, \ \lim_{t\to\infty} e^{-\theta t} \lambda_{3,t} l_t^* = 0.$$

Consider any feasible path (x_t, z_t) with the same initial condition x_0^*. First, we show that $\lim_{t\to\infty} e^{-\theta t} \langle \lambda_t, x_t^* - x_t \rangle = 0$.

From (27) we get $\lambda_{1,t} = \lambda_{1,0}.e^{-(f_1 - \delta - \theta - b + d)t}$ and

$$\lim_{t\to\infty} \lambda_{1,t} e^{-\theta t} = \lim_{t\to\infty} \lambda_{1,0} e^{-(f_1 - \delta - \theta - b + d)t} e^{-\theta t} = \lim_{t\to\infty} \lambda_{1,0} e^{-(f_1 - \delta - b + d)t} = 0$$

by Assumption 9. Since $k_t \leq \max\{k_0, \hat{k}\}$, $\lim_{t\to\infty} e^{-\theta t} \lambda_{1,t}(k_t^* - k_t) = 0$.

By Assumption 6, $c_t \leq f(k_t, l_t) + \kappa k_t \leq f(\hat{k}, 1) + \kappa \hat{k}$ and hence, $m_t \leq f(\hat{k}, 1) + \kappa \hat{k} = \hat{m}$, $\lambda_{2,t} = \frac{\lambda_{1,t}}{g'(m_t)} \leq \frac{\lambda_{1,t}}{g'(\hat{m})}$. Since $\dot{h}_t \leq g(\hat{m})$ we have

$$h_t \leq \int_0^t g(\hat{m}) ds + h_0 = g(\hat{m})t + h_0.$$

Then, using l'Hôpital's rule we get

$$\begin{aligned} 0 &\le \lim_{t\to\infty} e^{-\theta t}\lambda_{2,t}h_t \le \lim_{t\to\infty} e^{-\theta t}\frac{\lambda_{1,t}}{g'(\hat{m})}(g(\hat{m})t + h_0) \\ &= \lim_{t\to\infty} e^{-\theta t}\frac{\lambda_{1,t}g(\hat{m})t}{g'(\hat{m})} = \lim_{t\to\infty}\frac{g(\hat{m})t}{g'(\hat{m})\lambda_{1,0}e^{(f_1-\delta-b+d)t}} \\ &= \lim_{t\to\infty}\frac{g(\hat{m})}{g'(\hat{m})\lambda_{1,0}(f_1-\delta-b+d)e^{(f_1-\delta-b+d)t}} = 0. \end{aligned}$$

Therefore, $\lim_{t\to\infty} e^{-\theta t}\lambda_{2,t}(h_t^* - h_t) = 0$.

Since l_t is bounded away from 0 (from the law of motion), let the lower bound be denoted by $\hat{l}$. Then $0 \le \lim_{t\to\infty} e^{-\theta t}\lambda_{3,t}\hat{l} \le \lim_{t\to\infty} e^{-\theta t}\lambda_{3,t}l_t^* = 0$, the last inequality is the transversality condition. This implies $\lim_{t\to\infty} e^{-\theta t}\lambda_{3,t} = 0$. Furthermore, for any feasible l_t, $0 \le \lim_{t\to\infty} e^{-\theta t}\lambda_{3,t}l_t \le \lim_{t\to\infty} e^{-\theta t}\lambda_{3,t} = 0$ (since $l_t \le 1$). Thus, $\lim_{t\to\infty} e^{-\theta t}\lambda_{3,t}(l_t^* - l_t) = 0$.

Now, for any (x_t, z_t) feasible with $x_0 = x_0^*$, we have

$$\begin{aligned} &\int_0^\infty e^{-\theta t}[H(x_t^*, z_t^*, \lambda_t) - H(x_t, z_t, \lambda_t)] + \langle\dot{\lambda}_t - \theta\lambda_t, x_t^* - x_t\rangle]dt = \\ = &\int_0^\infty e^{-\theta t}[H(x_t^*, z_t^*, \lambda_t) - H(x_t, z_t, \lambda_t) + \langle\dot{\lambda}_t - \theta\lambda_t, x_t^* - x_t\rangle]dt = \\ = &\int_0^\infty e^{-\theta t}[H(x_t^*, z_t^*, \lambda_t) - H(x_t^*, z_t, \lambda_t)]dt \\ &+ \int_0^\infty e^{-\theta t}[H(x_t^*, z_t, \lambda_t) - H(x_t, z_t, \lambda_t) + \langle\dot{\lambda}_t - \theta\lambda_t, x_t^* - x_t\rangle]dt \end{aligned}$$

Using $u'(c_t^*) = \lambda_{1,t} = \lambda_{2,t}g'(m_t^*)$ and concavity of u, g we have

$$\begin{aligned} &H(x_t^*, z_t^*, \lambda_t) - H(x_t^*, z_t, \lambda_t) \\ = &\ u(c_t^*) - u(c_t) - \lambda_{1,t}(c_t^* - c_t) - \lambda_{1,t}(m_t^* - m_t) + \lambda_{2,t}(g(m_t^*) - g(m_t)) \\ \ge &\ u'(c_t^*)(c_t^* - c_t) - \lambda_{1,t}(c_t^* - c_t) - \lambda_{1,t}(m_t^* - m_t) + \lambda_{2,t}g'(m_t^*)(m_t^* - m_t) \\ = &\ 0. \end{aligned}$$

Thus

$$\int_0^\infty e^{-\theta t}[H(x_t^*, z_t^*, \lambda_t) - H(x_t^*, z_t, \lambda_t)]dt \ge 0. \tag{30}$$

Consider the next term:

$$\begin{aligned}
& \int_0^\infty e^{-\theta t}[H(x_t^*, z_t, \lambda_t) - H(x_t, z_t, \lambda_t)]dt + \int_0^\infty e^{-\theta t}\langle \dot{\lambda}_t - \theta\lambda_t, x_t^* - x_t\rangle dt \\
= & \int_0^\infty e^{-\theta t}\langle \lambda_t, \dot{x}_t^* - \dot{x}_t\rangle + \int_0^\infty e^{-\theta t}\langle \dot{\lambda}_t - \theta\lambda_t, x_t^* - x_t\rangle dt \\
= & \int_0^\infty e^{-\theta t}[\lambda_{1,t}(\dot{k}_t^* - \dot{k}_t) + \lambda_{2,t}(\dot{h}_t^* - \dot{h}_t) + \lambda_{3,t}(\dot{l}_t^* - \dot{l}_t)]dt + \\
& + \int_0^\infty [e^{-\theta t}(\dot{\lambda}_{1,t} - \theta\lambda_{1,t})(k_t^* - k_t) + e^{-\theta t}(\dot{\lambda}_{2,t} - \theta\lambda_{2,t})(h_t^* - h_t\rangle + e^{-\theta t}(\dot{\lambda}_{3,t} - \theta\lambda_{3,t})(l_t^* - l_t)]dt
\end{aligned}$$

Since $e^{-\theta t}(\dot{\lambda}_t - \theta\lambda_t) = \frac{d(e^{-\theta t}\lambda_t)}{dt}$ we have

$$\begin{aligned}
& \int_0^\infty [e^{-\theta t}\lambda_{1,t}(\dot{k}_t^* - \dot{k}_t) + e^{-\theta t}(\dot{\lambda}_{1,t} - \theta\lambda_{1,t})(k_t^* - k_t)]dt \\
= & \int_0^\infty [\frac{d(e^{-\theta t}\lambda_{1,t})(k_t^* - k_t)}{dt}]dt = \lim_{t\to\infty} e^{-\theta t}\lambda_{1,t}(k_t^* - k_t) - \lambda_{1,0}(k_0^* - k_0) \\
= & \lim_{t\to\infty} e^{-\theta t}\lambda_{1,t}(k_t^* - k_t).
\end{aligned}$$

By same reasoning we have

$$\begin{aligned}
& \int_0^\infty e^{-\theta t}[H(x_t^*, z_t, \lambda_t) - H(x_t, z_t, \lambda_t)]dt + \int_0^\infty e^{-\theta t}\langle \dot{\lambda}_t - \theta\lambda_t, x_t^* - x_t\rangle dt \\
= & \lim_{t\to\infty} e^{-\theta t}\lambda_{1,t}(k_t^* - k_t) + \lim_{t\to\infty} e^{-\theta t}\lambda_{2,t}(h_t^* - h_t) + \lim_{t\to\infty} e^{-\theta t}\lambda_{3,t}(l_t^* - l_t) = 0. \qquad (31)
\end{aligned}$$

It follows from (30) and (31)

$$\int_0^\infty e^{-\theta t}[H(x_t^*, z_t^*, \lambda_t) - H(x_t, z_t, \lambda_t) + < \dot{\lambda}_t - \theta\lambda_t, x_t^* - x_t >]dt \geq 0$$

which is equivalent to

$$\int_0^\infty e^{-\theta t}(u(c_t^*) - u(c_t))dt + \lim_{t\to\infty} e^{-\theta t}\lambda_{1,t}(k_t^* - k_t) + \lim_{t\to\infty} e^{-\theta t}\lambda_{2,t}(h_t^* - h_t) + \lim_{t\to\infty} e^{-\theta t}\lambda_{3,t}(l_t^* - l_t) \geq 0$$

or

$$\int_0^\infty e^{-\theta t}(u(c_t^*) - u(c_t))dt \geq 0.$$

■

Corollary 1. *The disease endemic steady state with health expenditures is locally optimal.*

Remark 2. *At the disease endemic steady state with health expenditures, since* $f_1 > b - d + \delta$*, Assumption 9 is not needed.*

As the endemic steady state with positive health expenditure satisfies the necessary conditions, we have shown that it is indeed optimal. This is true for the other two steady states as well. Thus, we have the following result.

Theorem 2. *All the steady states are locally optimal.*

Using Assumption 9 we show that $\lim_{t\to\infty} e^{-\theta t}\langle \lambda_t, x_t^* - x_t\rangle = 0$. This assumption is needed to check (local) optimality of a path that satisfies the necessary conditions. As this condition holds at a steady state, it is not need for optimality of a steady state. This is crucial as when we check the maximality of the Hamiltonian we can decompose it into two parts: the first just relies on the control variables and we have concavity in the objective function in control variables, and thus, using standard results the difference between the candidate solution and any other solution is non-negative; and a term that depends on the co-state and the state variables as given above. Recall, the non-concavity in the problem arises from the law of evolution of l only. As indicated, we show this term converges to zero, and we are able to obtain sufficiency of the first order conditions. Thus, three things turn out to be important in our problem: the boundedness of the state and control and hence, co-state variables; concavity of objective in control variables; and the continuity of the control and state/co-state variables.

6. Local Stability and Bifurcation

The dynamical system is given by equations (2)-(6), (12)-(18) and there are three possible steady states. In order to examine their stabilities we linearize the system around each of these. To simplify the exposition we make the following assumption.

Assumption 10. *The instantaneous utility function* $u(c) = \log(c)$.

Substituting $\lambda_1 = u'(c) = 1/c$ into equation (14), we get

$$\dot{c} = c(f_1 - \delta - \theta - b + d).$$

6.1. Disease-Free Case

At the disease-free steady state, $\lambda_1 > \lambda_2 g'$. Since all the functions in this model are smooth functions, by continuity there exists a neighborhood of the steady state such that the above inequality still holds. Thus, from equation(13) we have $m = 0$ in this neighborhood. Intuitively, around the steady state the net marginal benefit of health investment is negative: the disease is eradicated and health investment only serves to reduce physical capital accumulation and hence, lower levels of consumption, and thus no resources are spent on eradicating diseases. As $m = 0$ in the neighborhood of the steady state, we have a maximization problem with only one choice variable -

consumption - and the dynamical system reduces to:

$$\begin{aligned}
\dot{k} &= f(k,l) - c - \delta k - k(b-d) \\
\dot{h} &= -\delta h - h(b-d) \\
\dot{l} &= (1-l)(b - \alpha(h)l + \gamma(h)) \\
\dot{c} &= c(f_1 - \delta - \theta - b + d).
\end{aligned}$$

By linearizing the system around the steady state, we have[15]:

$$\mathcal{J}_1 = \begin{pmatrix} \theta & 0 & f_2^* & -1 \\ 0 & -\delta-(b-d) & 0 & 0 \\ 0 & 0 & \overline{\alpha}-(\underline{\gamma}+b) & 0 \\ c^* f_{11}^* & 0 & c^* f_{12}^* & 0 \end{pmatrix}.$$

The eigenvalues are $\Lambda_1 = -\delta - (b-d) < 0$, $\Lambda_2 = \frac{\theta - \sqrt{\theta^2 - 4c^* f_{11}^*}}{2} < 0$, $\Lambda_3 = \frac{\theta + \sqrt{\theta^2 - 4c^* f_{11}^*}}{2} > 0$, and $\Lambda_4 = \overline{\alpha} - (\underline{\gamma} + b)$. The sign of Λ_4 depends on b. We notice if $b = \overline{\alpha} - \underline{\gamma}$, $\mathcal{J}_1$ has a single zero eigenvalue. Thus, we have a non-hyperbolic steady state and a bifurcation may arise. In other words, the disease-free steady state possesses a 2-dimensional local invariant stable manifold, a 1-dimensional local invariant unstable manifold and 1-dimensional local invariant center manifold. In general, however, the behavior of trajectories in center manifold cannot be inferred from the behavior of trajectories in the space of eigenvectors corresponding to the zero eigenvalue. Thus, we shall take a close look at the flow in the center manifold. As the zero eigenvalue comes from dynamics of l, and the dynamics of l and h are independent from the rest of the system, we could just focus on the dynamics of l and h. By taking b as bifurcation parameter (see Kribs-Zaleta (2003) and Wiggins (2002)) the dynamics on the center manifold is given by (See the Appendix for details):

$$\dot{\xi} = \overline{\alpha}\xi \left(\xi - \frac{1}{\overline{\alpha}}(b - (\overline{\alpha} - \underline{\gamma})) \right).$$

The fixed points of the above equation are given by $\xi = 0$ and $\xi = \frac{1}{\overline{\alpha}}(b - (\overline{\alpha} - \underline{\gamma}))$, and plotted in figure 4. We can see the dynamics on the center manifold exhibits a transcritical bifurcation at $b = \overline{\alpha} - \underline{\gamma}$. Hence, for $b < \overline{\alpha} - \underline{\gamma}$, there are two fixed points; $\xi = 0$ is unstable and $\xi = \frac{1}{\overline{\alpha}}(b - \overline{\alpha} - \underline{\gamma})$ is stable. These two fixed points coalesce at $b = \overline{\alpha} - \underline{\gamma}$, and for $b > \overline{\alpha} - \underline{\gamma}$, $\xi = 0$ is stable and $\xi = \frac{1}{\overline{\alpha}}(b = \overline{\alpha} - \underline{\gamma})$ is unstable. Thus, an exchange of stability occurs at $b = \overline{\alpha} - \underline{\gamma}$.

Therefore, for the original dynamical system if $b > \overline{\alpha} - \underline{\gamma}$, there is a 3-dimensional stable manifold and a 1-dimensional unstable manifold, and if $b < \overline{\alpha} - \underline{\gamma}$, there is a 2-dimensional stable manifold and 2-dimensional unstable

[15]For brevity for a function $\phi(x^*)$ we write ϕ^*

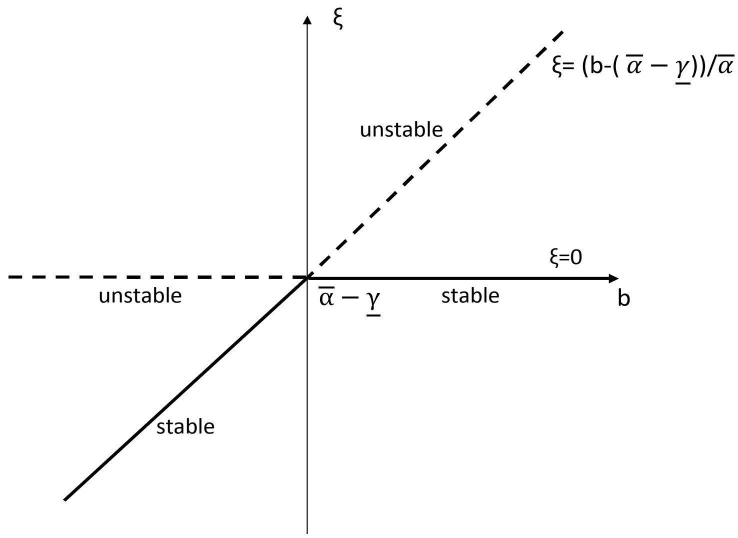

Figure 4: The Transcritical Bifurcation Diagram

manifold, that is if $b > \overline{\alpha} - \underline{\gamma}$, the disease-free steady state is locally saddle stable and has a unique stable path, and if $b < \overline{\alpha} - \underline{\gamma}$, the disease-free steady state is locally unstable.

6.2. Disease-Endemic Case

6.2.1. Disease-endemic case without health expenditure

For the endemic steady state without health expenditure, $\lambda_1 > \lambda_2 g'$ and $m^* = 0$. By continuity, this also holds in a small neighborhood of the steady state. Thus, it is similar to the disease-free case except that $l^* < 1$. Linearizing the system around the steady state:

$$\mathcal{J}_2 = \begin{pmatrix} \theta & 0 & f_2^* & -1 \\ 0 & -\delta - (b-d) & 0 & 0 \\ 0 & (1-l^*)(\gamma'^* - \alpha'^* l^*) & \overline{\alpha} - (\underline{\gamma} + b) & 0 \\ c^* f_{11}^* & 0 & c^* f_{12}^* & 0 \end{pmatrix}.$$

The eigenvalues are $\Lambda_1 = -\delta-(b-d) < 0$, $\Lambda_2 = \frac{\theta - \sqrt{\theta^2 - 4c^* f_{11}^*}}{2} < 0$, $\Lambda_3 = \frac{\theta + \sqrt{\theta^2 - 4c^* f_{11}^*}}{2} > 0$, and $\Lambda_4 = (\underline{\gamma}+b) - \overline{\alpha} < 0$. Thus, it has a 3-dimensional stable manifold and 1-dimensional unstable manifold, that is the endemic steady state without health expenditure is locally saddle stable and has a unique stable path.

6.2.2. Disease-endemic case with health expenditure

For the endemic case with health expenditure, the dynamical system is given by equations (2)-(6), (12)-(18) with $\lambda_1 = \lambda_2 g'$, $m^* > 0$ and $l^* < 1$. Simplifying, the system reduces to :

$$\begin{aligned}
\dot{k} &= f(k,l) - c - m - \delta k - k(b-d) \qquad (32)\\
\dot{h} &= g(m) - \delta h - h(b-d)\\
\dot{l} &= (1-l)(b+\gamma(h) - \alpha(h)l)\\
\dot{c} &= c(f_1 - \delta - \theta - b + d)\\
\dot{m} &= (c\lambda_3 g'(m)(1-l)(\gamma' - \alpha' l) - f_1)\frac{g'(m)}{g''(m)}\\
\dot{\lambda_3} &= -\frac{1}{c}f_2 + \lambda_3\theta - \lambda_3(2\alpha(h)l - b - \gamma(h) - \alpha(h)).
\end{aligned}$$

We now have a higher dimensional system than the earlier two cases as $m^* > 0$ and $h^* > 0$. Linearizing around the steady state the Jacobian is given by:

$$\mathcal{J}_3 = \begin{bmatrix}
\theta & 0 & f_2^* & -1 & -1 & 0\\
0 & -\delta - (b-d) & 0 & 0 & g'^* & 0\\
0 & (1-l^*)(\gamma'^* - \alpha'^* l^*) & b + \gamma^* - \alpha^* & 0 & 0 & 0\\
c^* f_{11}^* & 0 & c^* f_{12}^* & 0 & 0 & 0\\
-f_{11}^* \frac{g'^*}{g''^*} & \frac{f_1^*(\gamma''^* - \alpha''^* l^*)}{\gamma'^* - \alpha'^* l^*}\frac{g'^*}{g''^*} & \left(\frac{f_1^*(2\alpha'^* l^* - \alpha'^* - \gamma'^*)}{(1-l^*)(\gamma'^* - \alpha'^* l^*)} - f_{12}^*\right)\frac{g'^*}{g''^*} & \frac{f_1^*}{c^*}\frac{g'^*}{g''^*} & f_1^* & \frac{f_1^*}{\lambda_3^*}\frac{g'^*}{g''^*}\\
-\frac{f_{12}^*}{c^*} & -\lambda_3^*(2\alpha'^* l^* - \gamma'^* - \alpha'^*) & -\frac{f_{22}^*}{c^*} - 2\lambda_3^*\alpha^* & \frac{f_2^*}{c^{*2}} & 0 & \frac{f_2^*}{c^*\lambda_3^*}.
\end{bmatrix}.$$

Let us denote $\mathcal{J}_3$ as a matrix $(a_{ij})_{6\times 6}$ with the signs of a_{ij} given as follows:

$$\begin{bmatrix}
a_{11}(+) & 0 & a_{13}(+) & -1 & -1 & 0\\
0 & a_{22}(-) & 0 & 0 & a_{25}(+) & 0\\
0 & a_{32}(+) & a_{33}(-) & 0 & 0 & 0\\
a_{41}(-) & 0 & a_{43}(+) & 0 & 0 & 0\\
a_{51}(-) & a_{52}(+) & a_{53} & a_{54}(-) & a_{55}(+) & a_{56}(-)\\
a_{61}(-) & a_{62} & a_{63} & a_{64}(+) & 0 & a_{66}(+)
\end{bmatrix}$$

Note that as $l^* = \frac{\gamma^* + b}{\alpha^*} < 1$, at the steady state $a_{33} = b + \gamma^* - \alpha^* < 0$. $\dot{\lambda}_3 = 0$ implies

$$\lambda_3^* = \frac{f_2^*}{c^*(\theta - 2\alpha^* l^* + b + \gamma^* + \alpha^*)} = \frac{f_2^*}{c^*(\theta + \alpha^* - b - \gamma^*)} > 0.$$

Thus, only a_{53}, a_{62} and a_{63} remain to be signed. The characteristic equation, $|\Lambda I - \mathcal{J}_3| = 0$, can be written as a polynomial of Λ:

$$P(\Lambda) = \Lambda^6 - D_1\Lambda^5 + D_2\Lambda^4 - D_3\Lambda^3 + D_4\Lambda^2 - D_5\Lambda + D_6 = 0 \qquad (33)$$

where the D_i are the sum of the i-th order minors about the principal diagonal of $\mathcal{J}_3$ which are explicitly defined (See Appendix C). Let Λ_i $(i = 1, \dots, 6)$ denote the solutions of the characteristic equation. By Vietae's formula we have

$$\Lambda_1 + \Lambda_2 + \Lambda_3 + \Lambda_4 + \Lambda_5 + \Lambda_6 = D_1 = 3\theta > 0,$$

which implies there exists at least one positive root.[16]

We now prove that, under the following assumption, the steady state is locally saddle stable, that is there are exactly three negative roots and three positive roots of the above characteristic equation.

Assumption 11. 1. $(\alpha^* - b - \gamma^*)(\gamma'^* - \alpha'^*) < (\theta + \alpha^* - b - \gamma^*)\left(\gamma'^* + \alpha'^* - \frac{2\alpha'^*(b+\gamma^*)}{\alpha^*}\right)$

2. $\theta < \frac{\varpi_1 + \varpi_2 + \sqrt{(\varpi_1+\varpi_2)^2 + 32(\varpi_1^2 + \varpi_2^2)}}{16}$, *where* $\varpi_1 = \delta + b - d, \varpi_2 = \alpha^* - \gamma^* - b$.

We can see that *A.11 (1)* holds if α'^*, i.e. the marginal effect of health capital on the contact rate at steady state, is very small. *A.11(2)* says that θ is small enough. It should be kept in mind that these are sufficient conditions for local saddle-point stability and in some problems of interest may not hold giving rise to richer dynamics. It follows from *A.11(1)* that

$$2\alpha'^* l^* - \alpha'^* - \gamma'^* = -\left(\gamma'^* + \alpha'^* - \frac{2\alpha'^*(b+\gamma^*)}{\alpha^*}\right) < 0.$$

Hence, $a_{53} > 0, a_{62} > 0$. With this assumption, every sign of a_{ij} is defined except for a_{63}.

Lemma 4. *Under A.1 - A.11 (1),* $\det \mathcal{J}_3 = D_6 < 0$ *and there exists at least one negative root.*

Proof. See the Appendix for the proof. ∎

Lemma 5. *Under A.1 - A.11, we have* $D_1 D_2 - D_3 < 0$, $D_2 < 0$ *and* $D_3 < 0$.

Proof. See the Appendix for the proof. ∎

Proposition 5. *Under A.1-A.11, if* $D_1 D_4 - D_5 \geq 0$ *or if* $D_1 D_4 - D_5 < 0$ *and* $(D_1 D_2 - D_3) D_5 < D_1^2 D_6$, *the endemic steady state with health expenditure is locally saddle stable.*

Proof. The number of negative roots of $P(\Lambda)$ is exactly the number of positive roots of

$$P(-\Lambda) = \Lambda^6 + D_1 \Lambda^5 + D_2 \Lambda^4 + D_3 \Lambda^3 + D_4 \Lambda^2 + D_5 \Lambda + D_6 = 0. \quad (34)$$

We use the Routh's stability criterion which states that the number of positive roots of equation (34) is equal to

[16]By positive (negative) root, we mean either real positive (negative) root or imaginary root with positive (negative) real part.

the number of changes in sign of the coefficients in the first column of the Routh's table as shown below:

$$\begin{bmatrix} 1 & D_2 & D_4 & D_6 & 0 \\ D_1 & D_3 & D_5 & 0 & 0 \\ a_1 & a_2 & D_6 & 0 & 0 \\ b_1 & b_2 & 0 & 0 & 0 \\ c_1 & D_6 & 0 & 0 & 0 \\ d_1 & 0 & 0 & 0 & 0 \\ D_6 & 0 & 0 & 0 & 0 \end{bmatrix}$$

where

$$a_1 = \frac{D_1 D_2 - D_3}{D_1},\ a_2 = \frac{D_1 D_4 - D_5}{D_1},\ b_1 = \frac{a_1 D_3 - a_2 D_1}{a_1},$$
$$b_2 = \frac{a_1 D_5 - D_6 D_1}{a_1},\ c_1 = \frac{b_1 a_2 - a_1 b_2}{b_1},\ d_1 = \frac{c_1 b_2 - b_1 D_6}{c_1}.$$

Recall that we have $D_1 > 0, D_2 < 0, D_3 < 0, D_6 < 0$ and $D_1 D_2 - D_3 < 0$. So $a_1 < 0$ and the sign of the first column in the Routh's table is given as:

$$\begin{matrix} 1 & D_1 & a_1 & b_1 & c_1 & d_1 & D_6 \\ + & + & - & \pm & \pm & \pm & - \end{matrix}$$

As the signs of b_1, c_1, d_1 are indeterminate, we check all possible 8 cases. Among all the cases, 6 cases have exactly 3 times change of signs, which implies equation (34) has exactly 3 positive roots, or equation (33) has exactly 3 negative roots and the steady state is saddle point stable. However, for the other two cases, the steady state is either a sink or unstable, which we shall rule out.

Case 1: Suppose $b_1 > 0$, $c_1 < 0$ and $d_1 > 0$, which implies the steady state is a sink. Since

$$d_1 > 0 \Rightarrow c_1 b_2 < b_1 D_6 \Rightarrow b_2 > 0$$
$$b_1 > 0 \Rightarrow a_1 D_3 < a_2 D_1 \Rightarrow a_2 > 0$$
$$c_1 < 0 \Rightarrow b_1 a_2 < a_1 b_2 \Rightarrow a_2 < 0,$$

we reach a contradiction. So we cannot have the case that there are 5 times change of signs, that is, there cannot be 5 positive roots in equation (34) or 5 negative roots in the equation (33). Thus, the steady state cannot be a sink.

Case 2: Suppose $b_1 < 0$, $c_1 < 0$ and $d_1 < 0$, which implies the steady state is unstable.

If $D_1D_4 - D_5 \geq 0$, that is $a_2 > 0$, we have

$$c_1 < 0 \Rightarrow b_1a_2 > a_1b_2 \Rightarrow b_2 > 0$$
$$d_1 < 0 \Rightarrow c_1b_2 > b_1D_6 \Rightarrow b_2 < 0.$$

we reach a contradiction.

If $D_1D_4 - D_5 < 0$ and $(D_1D_2 - D_3)D_5 < D_1^2D_6$, $a_2 < 0$ and

$$b_2 = \frac{a_1D_5 - D_6D_1}{a_1} = \frac{(D_1D_2 - D_3)D_5 - D_1^2D_6}{D_1a_1} > 0.$$

It contradicts to $d_1 < 0$ which implies $b_2 < 0$.

So if $D_1D_4 - D_5 \geq 0$ or if $D_1D_4 - D_5 < 0$ and $(D_1D_2 - D_3)D_5 < D_1^2D_6$, we cannot have the case that there are only 1 time change in sign, that is, there can not be only 1 positive root in equation (34) or only 1 negative root in equation (33). The steady state can not be unstable. ■

The local stability and bifurcation of the dynamical system are summarized in Figure 3. When the birth rate b is greater than $\overline{\alpha} - \underline{\gamma}$, there is only a disease-free steady state which is locally stable. When b decreases to exactly $\overline{\alpha} - \underline{\gamma}$, the stable disease-free equilibrium goes through a transcritical bifurcation to two equilibria: one is the unstable disease-free steady state and the other is the stable endemic steady state with or without health expenditure.

7. Effects of Varying Discount and Birth Rates

With the results on existence and local stability, we are now ready to explore how the steady state properties of the model change as the parameters vary. The results of comparative statics in this section improve our understanding on two important empirical issues. First, we show as parameters vary, there is a nonlinearity in steady state changes due to the switches among the steady states and the role played by the endogenous changes in health expenditure. The non-linearities in equilibrium outcomes, which are often assumed away, may be very important in understanding aggregate behavior. While we are unable to study global dynamics as it is difficult in the system to derive policy functions and thus, are unable to study the full range of dynamics, the results point out that even steady states may change in a non-linear way. So the reduced formed estimation on examining the effect of diseases on the economy (e.g. Acemoglu and Johnson (2007), Ashraf, et al. (2009), Bell, et et al. (2003), Bloom, et al. (2009), and Young (2005)) by assuming a linear relationship may not be well justified as non-linearity is an important characteristic of models associated with the disease transmission, and this nonlinearity in disease transmission can become a source of non-linearities in economic outcomes. Second, we study the endogenous relationship between health expenditure (as percentage of output) and output. This can help us understand the changing share of health expenditures over decades in many countries. There are many factors which are thought to be the cause of positive relationship between income and health expenditure share in the literature, including technological development,

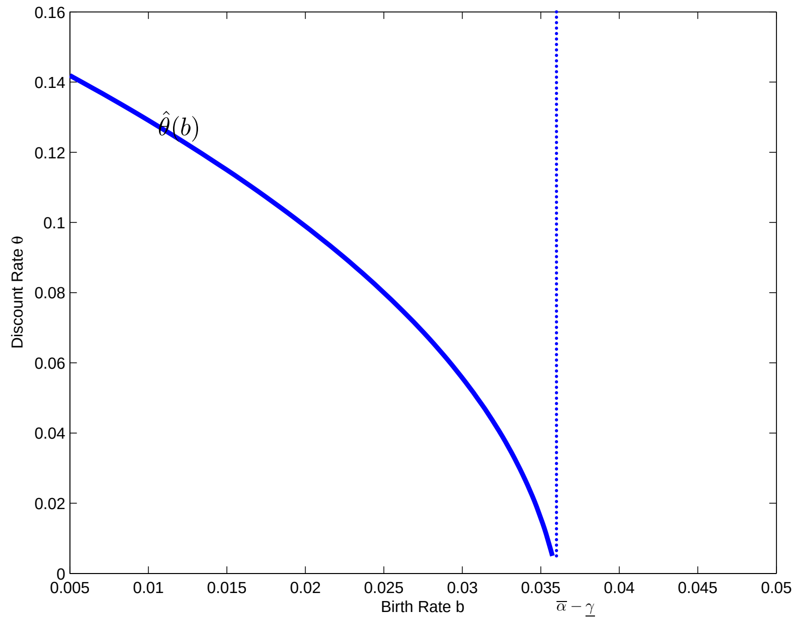

Figure 5: The Steady States in the Parameter Space (θ, b)

institutional change, health as a luxury good, etc. However, our results suggest maybe we should pay close look at more fundamental factors such as change in longevity and fertility rate. It should be emphasized that while we are looking at only public health expenditure on infectious diseases, this methodology can be extended to incorporate non-infectious diseases. Moreover, as we only model one type of infectious diseases here, the comparative statics results need to be interpreted with caution when comparing them with the empirical facts.

For the numerical analysis, we specify the following functional forms: output $y = f(k, l) = Ak^a l^{1-a}$, health production function $g(m) = \phi_3(m + \phi_1)^{\phi_2} - \phi_3\phi_1^{\phi_2}$, contact rate function $\alpha(h) = \alpha_1 + \alpha_2 e^{-\alpha_3 h}$, recovery rate function $\gamma(h) = \gamma_1 - \gamma_2 e^{-\gamma_3 h}$. By convention we choose $A = 1, a = 0.36, \delta = 0.05$ and $d = 0.5\%$. Since there are no counterparts for health related functions in the economic literature, we choose $\phi_1 = 2, \phi_2 = 0.1, \phi_3 = 1, \alpha_1 = \alpha_2 = 0.023, \alpha_3 = 1, \gamma_1 = 1.01, \gamma_2 = \gamma_3 = 1$ such that Assumptions *A.1-A.7* are satisfied. Sufficient conditions for stability (*A.11*) may not be satisfied as the parameters are varied, but we check that the stability properties continue to hold in the parameter range of interest.

We have $\overline{\alpha} = 0.046$, $\underline{\gamma} = 0.01$, and the function $\hat{\theta}(b)$ is shown in Figure 5. As we discuss above, if $b > 0.036$, only a stable disease-free steady state exists. As the birth rate decreases across 0.036, the stable disease-free steady state goes through a transcritical bifurcation to two steady states: one is the unstable disease-free steady state and the other is the stable endemic steady state. Below the curve $\hat{\theta}(b)$, endemic steady state with positive health expenditure exists, and above the curve, endemic steady state without health expenditure exists. Two experiments are conducted here. First, we take a vertical slice at birth rate $b = 2\%$ and vary discount rate θ. The disease-free steady state always exists, and from Figure 5 we see if $\theta < 0.1$ an endemic steady state with health expenditure exists, and if $\theta \geq 0.1$ an endemic steady state without health expenditure exists. Second, we take a horizontal slice at discount rate $\theta = 0.05$, and vary birth rate b. From Figure 5, we see if $b > 3.6\%$ only a disease-free steady state exists, if $b \in (3.1\%, 3.6\%)$ both the disease-free steady state and endemic steady state without health expenditure exist, and if $b \in (0.5\%, 3.1\%)$ both the disease-free steady state and endemic steady state with health expenditure exist. Note the analytical results of comparative statics on disease-free steady state are not given below as they are exactly the same as in the standard neo-classical growth model, but the simulation results are included in the figures.

7.1. Discount rate θ

This comparative statics can be interpreted as studying the effect of increasing longevity as a decrease in θ is often interpreted as an increase in longevity (Hall and Jones (2007)). As θ is varied, in the endemic steady state without health expenditure,

$$\frac{dk^*}{d\theta} = \frac{1}{f_{11}} < 0, \quad \text{and} \quad \frac{dc^*}{d\theta} = \frac{\theta}{f_{11}} < 0.$$

The disease prevalence $l^* = \frac{\gamma+b}{\bar{\alpha}}$ remains unchanged.

In the endemic steady state with health expenditure, we have $\frac{\partial m}{\partial h} = \frac{\delta+(b-d)}{g'} > 0$ and $\frac{\partial l'_\theta(h)}{\partial \theta} = \frac{-l'_\theta(h)}{\alpha(h)-(\gamma+b)+\theta} < 0$. Let $\Psi = g'l'l'_\theta(f_{11}f_{22} - f_{12}f_{21}) + f_{11}(f_2 g' l''_\theta + f_2 g'' \frac{\partial m}{\partial h} l'_\theta) > 0$. By the multi-dimensional implicit function theorem , we have:

$$\begin{aligned}
\frac{dk^*}{d\theta} &= \frac{1}{\Psi}\left(f_{22}g'l'l'_\theta + f_2 g' l''_\theta + f_2 g'' \frac{\partial m}{\partial h} l'_\theta - f_{12}l'\left(1 - f_2 g' \frac{\partial l'_\theta}{\partial \theta}\right)\right) < 0, \\
\frac{dh^*}{d\theta} &= \frac{1}{\Psi}\left(f_{11}\left(1 - f_2 g' \frac{\partial l'_\theta}{\partial \theta}\right) - f_{21}g'l'_\theta\right) < 0, \\
\text{and, thus,} \quad \frac{dl^*}{d\theta} &= l' \frac{dh^*}{d\theta} < 0, \\
\frac{dc^*}{d\theta} &= (f_1 - \delta - (b-d))\frac{dk^*}{d\theta} + (f_2 l' - \delta - (b-d))\frac{dh^*}{d\theta} < 0.
\end{aligned}$$

Therefore, from the analytical comparative statics results, we see in the endemic steady state without health expenditure variations in the discount rate have no effect on the spread of infectious diseases, since without health expenditures the mechanism of disease spread is independent of society's behavior. The smaller discount rate only leads to higher physical capital and consumption in exactly the same way as in the neo-classical model. In the endemic steady state with health expenditure, as the discount rate decreases, that is as the people become more patient, they spend more resources in prevention of infections or getting better treatment. The rise in health capital leads to a larger labor force, and both physical capital and consumption will increase.

This is also seen from the simulation in Figure 6 with solid line denoting the disease-free steady state and dash line denoting the endemic steady state. For the disease-free steady state, there is full employment (panel (1)) and both health expenditure (panel (2)) and health capital (panel (3)) are zero. As θ decreases or people become more patient, physical capital (panel (4)), consumption (panel (5)) and output (panel (6)) increase following the exact same mechanism of standard neo-classical economy. For the endemic steady state with $\theta > 0.1$, a change in θ

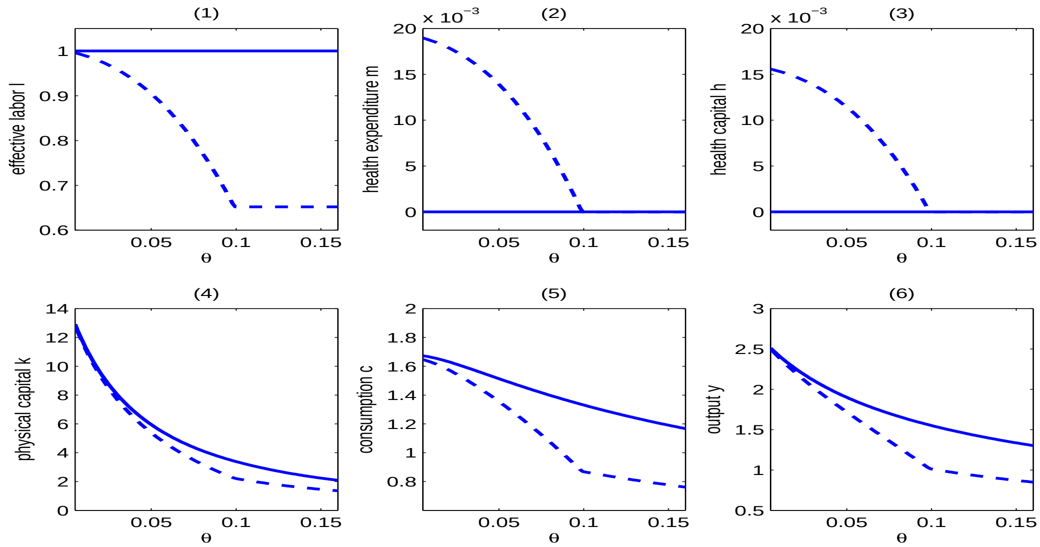

Figure 6: Change in Economic Variables as Discount Date θ Varies(Solid Line - Disease-free Case; Dash Line - Disease-endemic Case)

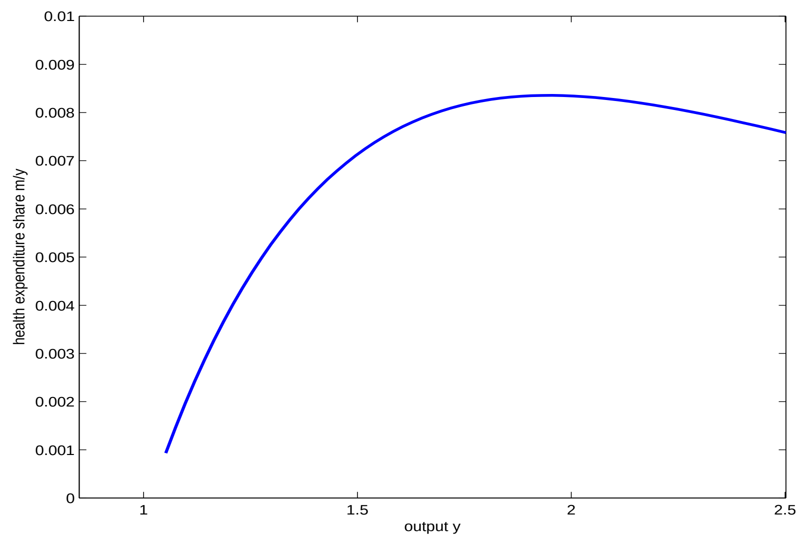

Figure 7: Change in Health Expenditure Share and Output as Discount Rate θ Decreases

doesn't have any effect on the effective labor force, and both health expenditure and health capital remain zero. But as θ decreases, physical capital, consumption and output increase. When $\theta < 0.1$, both health expenditure and health capital are positive, and further decreases in θ cause all economic variables increase. Physical capital, consumption and output increase at a faster rate than in the endemic steady state without health expenditure.

In Figure 7 if we only focus on the endemic steady state with positive health expenditure part, as θ decreases, both output y and health expenditure m increases, while the share of health expenditure m/y first increases and then decreases. We can see from panel (4) and (3) in Figure 6 that the rate of investment in physical capital (slope of the curve) is increasing while that of health capital (slope of the curve) is decreasing as θ decreases. This leads to an initial increase in the share of health expenditure in output and then an eventual decrease. The intuition is that as people become more patient, they spend more on health. This has two effects. First, as the incidence of diseases is controlled the increase in the effective labor force increases the marginal product of capital which leads to the increasing rate of physical capital investment. Second, as the incidence of diseases decreases, due to the externality in disease transmission the fraction of infectives decreases. This decreases the rate of investment in health expenditures. This leads to a non-monotonicity in the share of health expenditure. The *initial* positive relationship between income and health expenditure is similar to the finding of Hall and Jones (2007). However, unlike their model we do not have to introduce a taste for health. They need to assume that the marginal utility of life extension does not decline as rapidly as that of consumption declines as income increases, i.e. there is a more rapid satiation of consumption than life extension.

7.2. Birth rate b

In the endemic steady state without health expenditure, we have

$$\frac{dl^*}{db} = \frac{1}{\bar{\alpha}} > 0, \quad \frac{dk^*}{db} = \underbrace{\frac{1}{f_{11}}}_{-} + \underbrace{\frac{f_{12}}{-\bar{\alpha} f_{11}}}_{+}, \quad \text{and} \quad \frac{dc^*}{db} = \underbrace{\frac{\theta - k f_{11}}{f_{11}}}_{-} + \underbrace{\frac{\theta f_{12} - f_2 f_{11}}{-\bar{\alpha} f_{11}}}_{+}.$$

A decrease in birth rate causes effective labor force to decrease due to fewer healthy newborns. However, the effect of the birth rate decrease on other economic variables is ambiguous due to two offsetting aspects. First, it has a positive effect (the minus sign) as the marginal cost of physical capital decreases which leads to higher physical capital and consumption. Second, there is a negative effect (the positive sign): The proportion of healthy people decreases due to fewer healthy newborns, and thus, the smaller labor force leads to lower physical capital and consumption.

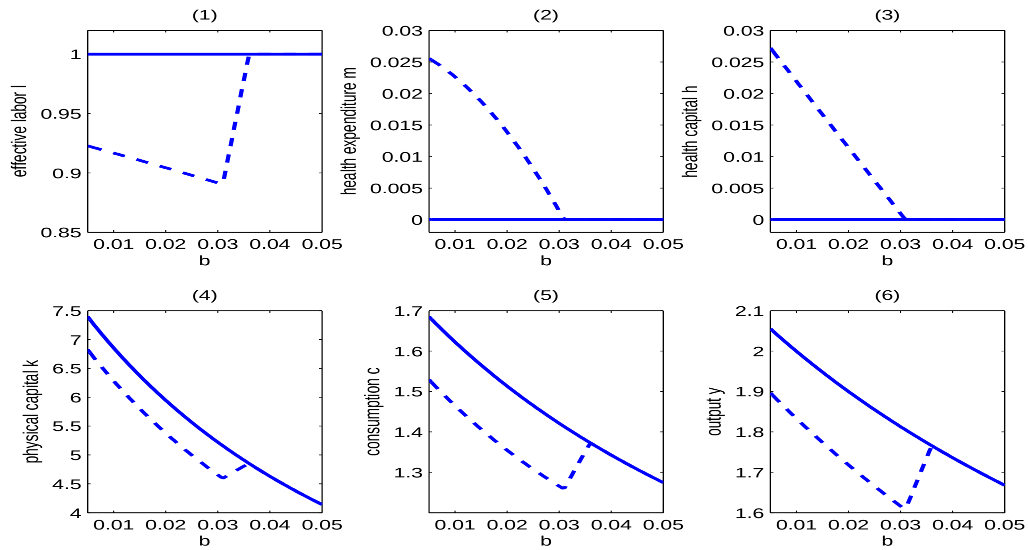

Figure 8: Change in Economic Variables as Birth Date b Varies(Solid Line - Disease-free Case; Dash Line - Disease-endemic Case)

In the endemic case with health expenditure, by the implicit function theorem we have:

$$\frac{dk^*}{db} = \underbrace{\frac{1}{\Psi}(f_{22}g'l'l'_\theta + f_2g'l''_\theta + f_2g''\frac{\partial m}{\partial h}l'_\theta - f_{12}l')}_{-} \underbrace{-\frac{1}{\Psi}f_2f_{12}g'\frac{1}{\alpha}l''_\theta}_{+} + \underbrace{\frac{1}{\Psi}f_2f_{12}g'l'\frac{\partial l'_\theta}{\partial b}}_{?}$$

$$\frac{dh^*}{db} = \underbrace{\frac{1}{\Psi}(f_{11} - f_{21}g'l'_\theta)}_{-} + \underbrace{\frac{1}{\Psi}\frac{1}{\alpha}g'l'_\theta(f_{21}f_{12} - f_{11}f_{22})}_{-} + \underbrace{\frac{1}{\Psi}(-f_{11}f_2g'\frac{\partial l'_\theta}{\partial b})}_{?}$$

and then $\quad \dfrac{dl^*}{db} = \dfrac{1}{\alpha} + l'(h)\dfrac{dh^*}{db}$

where $\frac{\partial l'_\theta}{\partial b} = -\frac{\alpha'}{\alpha^2} + \frac{\theta(\theta\alpha' + \alpha(\alpha' - \gamma'))}{\alpha^2(\alpha - (\gamma + b) + \theta)^2}$.

Therefore, the effect of birth rate decrease is ambiguous. The basic reasoning is similar to the endemic case without health expenditure above, but here it becomes more complex by involving changes in health capital, and hence effective labor supply may increase rather than decrease. First, as above there is a positive effect: the marginal cost of physical capital and health capital will decrease which lead to higher physical capital and health capital. Second, if effective labor force increases (decreases), there is a positive (negative) effect, as the marginal productivity of physical capital increases (decreases) physical capital increases (decreases). Third, effect of changing birth rate on the marginal product of health capital on labor supply, that is $\partial l'_\theta/\partial b$, is unclear.

To see the effects more clearly we consider the parametrized economy. We vary birth rate b from 0.5% to 5%. As we already know, the disease-free steady state always exists (shown in solid line in Figure 8). There is full employment, and both health expenditure and health capital are zero. As b decreases, physical capital, consumption and output increase. The endemic steady state is shown in dash line. For the endemic steady state with $b \in (3.1\%, 3.6\%)$, a decrease in b causes effective labor force to drop as fewer healthy people are born, and both

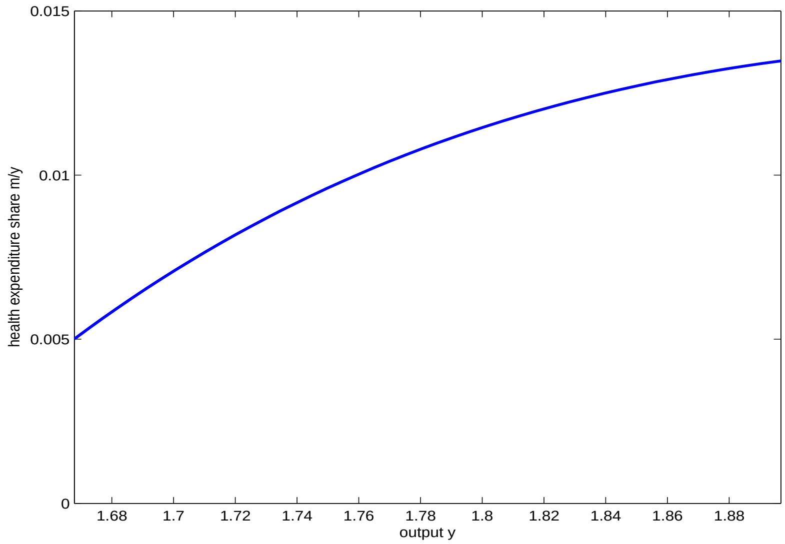

Figure 9: Change in Health Expenditure Share and Output as Birth Rate b Decreases

health expenditure and health capital remain zero. As b decreases further, physical capital, consumption and output decrease as the negative effect dominates due to decreasing effective labor supply. When $b \in (0.5\%, 3.1\%)$, both health expenditure and health capital are positive, and a decrease in b causes all economic variables to increase. It shows that the positive effect dominates. The intuition is that as the birth rate falls the cost of the marginal worker falling ill becomes higher and this leads to an increase in health expenditure and hence health capital. This leads to a larger effective labor force, and then higher physical capital, consumption and output. This is consistent with the empirical finding that low birth rates are associated with higher per capita income (see Brander and Dowrick (1994)).

In Figure 9, if we only focus on endemic steady state with positive health expenditure, we get the endogenous positive relationship between output and the share of health expenditures as birth rate falls. The reason is that decreases in the birth rate increases the marginal cost of an additional worker falling ill. The optimal response is to raise health expenditure, i.e. a more aggressive strategy to control the incidence of the disease. This interacts with the rising per capita capital stock and the increasing marginal product of capital which cause the output to rise as well.

8. The Conclusion

In a recent paper, Goenka and Liu (2012) examine a discrete time formulation of a similar model. In that paper, however, there is only a one way interaction between the disease and the economy. The disease affects the labor force as in this model, but the labor supply by healthy individuals is endogenous and the epidemiology parameters are treated as biological constants. Under the simplifying assumption of a one-way interaction, the dynamics become two-dimensional and the global dynamics are analyzed. The key result is that as the disease becomes more infective, cycles and then eventually chaos emerges. Here, we endogenize the epidemiology parameters. Thus, it is a framework to study optimal health policy. However, the dynamical system becomes six dimensional and we have to restrict our analysis to local analysis of the steady state.

This paper develops a framework to study the interaction of infectious diseases and economic growth by establishing a link between the economic growth model and epidemiology model. There is a problem in modeling disease dynamics as they are non-concave and thus, the usual methods in the literature are not applicable. One of the contributions of the paper is to show that in fact the Hamiltonian approach can still be used as despite this, the problem is sufficiently well-behaved. However, the non-concavity can give rise to more interesting economic possibilities. We find that there are multiple steady states. Furthermore by examining the local stability we explore how the equilibrium properties of the model change as the parameters are varied. Although the model we present here is elementary, it provides a fundamental framework for considering more complicated models. It is important to understand the basic relationship between disease prevalence and economic growth before we go even further to consider more general models. The model also points out the link between the health expenditures and income - both of which are endogenous - may be driven by fundamental factors - drop in the fertility rate or increase in the longevity. In related work, Goenka and Liu (2013) extend the framework to the case of endogenous growth where there is an additional choice on how much human capital to accumulate.

Acknowledgements

We would like to thank Murali Agastya, Michele Boldrin, Russell Cooper, Boyan Jovanovic, Takashi Kamihigashi, M. Ali Khan, Cuong Le Van, Francois Salanie, Karl Shell and seminar participants at the 2007 Asian General Equilibrium Theory Workshop, Singapore; 2008 European General Equilibrium Workshop, Paestum; FEMES 2008; NUS Macro Brown Bag Workshop; University of Paris I; University of Cagliari; City University London; SWIM Auckland 2010; Kyoto Institute of Economic Research; 2010 Econometric Society World Congress; 2010 Cornell-PSU Macro Workshop; CUNY; and 2012 Royal Economic Society meetings for helpful comments and suggestions. We would especially like to thank the anonymous referee and Editor, Atsushi Kajii, whose comments spurred us to extend the results in the paper. Aditya Goenka would like to acknowledge research support from FASS Research Grant R-122-000-131-112. Manh Hung Nguyen would like to thank the Isaac Manasseh Meyer Fellowship for research support. The usual disclaimer applies.

[1] Acemoglu, D. and Johnson, S. (2007) Disease and development: The effect of life expectancy on economic growth, *Journal of Political Economy*, 115: 925-985.

[2] Anderson, R.M. and May, R.M. (1991) *Infectious Diseases of Humans*, Oxford: Oxford Science Publications.

[3] Arrow, K.J. and Kurz, M. (1970) *Public investment, the rate of return, and optimal fiscal policy*, Baltimore: The Johns Hopkins Press.

[4] Askenazy, P. and Le Van, C. (1999) A model of optimal growth strategy, *Journal of Economic Theory* 85: 24-51.

[5] Ashraf, Q.H., Lester, A., and Weil, D.N. (2009) When does improving health raise GDP? In D. Acemoglu, K. Rogoff, and M. Woodford (Eds.) *NBER Macroeconomics Annual 2008*, Chicago: The University of Chicago Press.

[6] Bell, C., Devarajan, S. and Gersbach, H., (2003) The long-run economic costs of AIDS: theory and an application to South Africa, *Working paper*, The World Bank.

[7] Bleakley, H. (2007) Disease and development: Evidence from hookworm eradication in the American South, *Quarterly Journal of Economics*, 122: 73-117.

[8] Bloom, D.E., Canning, D. and Fink, G. (2009) Diseases and development revisited, NBER Working Paper No. 15137.

[9] Brander, J.A. and Dowrick, S. (1994) The Role of fertility and population in economic growth: empirical results: From aggregate cross-national data, *Journal of Population Economics*, 7: 1-25.

[10] Chichilnisky, G. (1981) Existence and characterization of optimal growth paths including models with non-convexities in utilities and technologies, *Review of Economic Studies*, 48, 51-61.

[11] d'Albis, H., Gourdel, P. and Le Van, C. (2008) Existence of solutions in continuous-time optimal growth models, *Economic Theory*, 37(2): 321-333.

[12] Delfino, D. and Simmons, P.J. (2000) Positive and normative issues of economic growth with infectious diseases, Discussion Papers in Economics 2000/48, University of York.

[13] Geoffard, P.-Y. and Philipson, T. (1997) Rational epidemics and their rational control, *International Economic Review*, 37: 603-624.

[14] Gersovitz, M. and Hammer, J.S. (2004) The economical control of infectious diseases, *The Economic Journal*: 1-27.

[15] Goenka, A., and Liu, L. (2012) Infectious diseases and endogenous fluctuations, *Economic Theory*, 50(1): 125-149, 2012.

[16] Goenka, A. and Liu, L. (2013) Infectious diseases, human capital and economic growth. Mimeo. NUS.

[17] Hall, R.E. and Jones, C. (2007) The value of life and the rise in health spending, *Quarterly Journal of Economics*, 122 (1): 39-72.

[18] Hethcote, H.W. (2009) The basic epidemiology models, in S. Ma and Y. Xia (Eds.) *Mathematical understanding of infectious disease dynamics*, Singapore: World Scientific.

[19] Jovanovic, B., and Rob, R. (1989) The growth and diffusion of knowledge, *Review of Economic Studies* 56: 569-582.

[20] Kribs-Zaleta, C.M. (2002) Center manifolds and normal forms in epidemic models, in C. Castillo-Chavez, et al. (Eds.) *Mathematical Approaches for Emerging and Reemerging Infectious Diseases: An Introduction* : 269-86.

[21] Kremer, M. (1996) Integrating behavioral choice into epidemiological models of the AIDS epidemic, *Quarterly Journal of Economics*, 111: 549-573.

[22] Le Van, C., and Vailakis, Y. (2003) Existence of a competitive equilibrium in a one sector growth model with irreversible investment, *Economic Theory*, 22: 743-771.

[23] Le Van,C., R. Boucekkine and C. Saglam (2007), Optimal control in infinite horizon problems: a Sobolev space approach, *Economic Theory*, 497-509.

[24] Maz'ja, V.G. (1985) *Sobolev Spaces* Heidelburg: Springer Verlag.

[25] Miguel, E. and Kremer, M. (2004) Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities, *Econometrica* 72, 159-217.

[26] Philipson, T. (2000) Economic epidemiology and infectious diseases, in J. Newhouse and T. Culyer (Eds.) *Handbook of Health Economics*, New York: North-Holland.

[27] Renardy, M and Rogers, R.C. (2004) *An introduction to partial differential equations*. Texts in Applied Mathematics 13 (Second edition ed.). New York: Springer-Verlag.

[28] Romer, P. (1986) Cake eating, chattering, and jumps: existence results for variational problems, *Econometrica*, 54, 897-908.

[29] Weil, D. (2007) Accounting for the effect of health on economic growth, *Quarterly Journal of Economics*, 122.

[30] Wiggins, S. (2003) *Introduction to applied nonlinear dynamical systems and chaos*, New York: *Springer-Verlag.*

[31] Young, A. (2005) The gift of the dying: The tragedy of AIDS and the welfare of future African generations, *Quarterly Journal of Economics* 120: 243-266.

Appendix A. Existence of Optimal Solution

We recall Dunford-Pettis Theorem, Mazur's Lemma (Renardy and Rogers (2004)) and the reverse Fatou's Lemma.

Let F be a family of scalar measurable functions on a finite measure space (Ω, Σ, μ), F is called uniformly integrable if $\{\int_E |f|\, d\mu,\ f \in F\}$ converges uniformly to *zero* when $\mu(E) \to 0$.

Dunford-Pettis Theorem: *Denote $L^1(\mu)$ the set of functions f such that $\int_\Omega |f|\, d\mu < \infty$ and K be a subset of $L^1(\mu)$. Then K is relatively weak compact if and only if K is uniformly integrable.*

When applying Fatou's Lemma to the non-negative sequence given by $g - f_n$, we get the following reverse Fatou's Lemma .

Fatou's Lemma: *Let f_n be a sequence of extended real-valued measurable functions defined on a measure space (Ω, Σ, μ). If there exists an integrable functiong on Ω such that $f_n \leq g$ for all n, then* $\limsup_{n\to\infty} \int_\Omega f_n d\mu \leq \int_\Omega \limsup_{n\to\infty} f_n d\mu$.

Mazur's lemma shows that any weakly convergent sequence in a normed linear space has a sequence of convex combinations of its members that converges strongly to the same limit. Because strong convergence is stronger than pointwise convergence, it is used in our proof for the state variables to converge pointwise to the limit obtained from weak convergence.

Mazur's Lemma: *Let $(M, ||\ ||)$ be a normed linear space and let $(f_n)_{n\in N}$ be a sequence in M that converges weakly to some f^* in M. Then there exists a function $\mathcal{N} : N \to N$ and a sequence of sets of real numbers $\{\omega_{i(n)} \mid i = n, \ldots, \mathcal{N}(n)\}$ such that $\omega_{i(n)} \geq 0$ and $\sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} = 1$ such that the sequence $(v_n)_{n\in N}$ defined by the convex combination $v_n = \sum_{i=n}^{\mathcal{N}(n)} \omega_{i(n)} f_i$ converges strongly in M to f^**, i.e., $||v_n - f^*|| \to 0$ as $n \to \infty$.

Appendix B. Local Stability Analysis

Appendix B.1. Center manifold calculation for disease-free case

Here, we introduce the procedure of calculating center manifold instead of the calculation part itself. We use $\dot{x} = G(x, b)$ to denote the dynamic system, where $x = (k, h, l, c)^T \in \Re^4_+$, and $G : \Re_+ \times \Re^4_+ \to \Re^4_+$ is the vector field. Moreover, we use x^* to denote its equilibrium point, and so $G(x^*, b) = 0$. Bifurcation occurs when $b^* = \overline{\alpha} - \underline{\gamma}$. We assume $G(x, b)$ to be at least C^5. We follow the procedure given by Kribs-Zaleta (2002) and Wiggins (2003):

1. Defining $\check{x} = x - x^*$ and $\check{b} = b - b^*$, we transform the dynamical system into $\dot{\check{x}} = G(\check{x} + x^*, \check{b} + b^*)$ with the equilibrium point $\check{x}^* = 0$ and bifurcation point $\check{b}^* = 0$. Then we linearize the system at point 0 to get $\dot{\check{x}} = D_x G(x^*, b^*)\check{x} + D_b G(x^*, b^*)\check{b} + R(\check{x}, \check{b})$, where $R(\check{x}, \check{b})$ is the high order term;
2. Let $A = D_x G(x^*, b^*)$, $B = D_b G(x^*, b^*)$ and calculate matrix A's eigenvalues, corresponding eigenvectors matrix TA (placing the eigenvector corresponding to zero eigenvalue first) and its inverse TA^{-1}. By transforming $\check{x} = TA \cdot y$, we get $\dot{y} = TA^{-1} \cdot A \cdot TA \cdot y + TA^{-1} \cdot B \cdot \check{b} + TA^{-1} \cdot R(TA \cdot y, \check{b})$, where $TA^{-1} \cdot A \cdot TA$ is its Jordan canonical form;
3. We separate y into two vectors y_1, the first term, and y_2, the rest terms, and then we can rewrite the system as:

$$\begin{aligned} \dot{y}_1 &= \Gamma_1 y_1 + \check{R}_1(TA \cdot y, \check{b}) \\ \dot{y}_2 &= \Gamma_2 y_2 + \check{R}_2(TA \cdot y, \check{b}); \end{aligned}$$

   Since $TA^{-1} \cdot B \neq 0$, we separate it into two vectors Δ_1 with only one element, and Δ_2 with the rest, and form a system as:

$$\begin{pmatrix} \dot{y}_1 \\ \dot{\check{b}} \\ \dot{y}_2 \end{pmatrix} = \underbrace{\begin{pmatrix} \Gamma_1 & \Delta_1 & 0_{1\times3} \\ 0 & 0 & 0_{1\times3} \\ 0_{3\times1} & \Delta_2 & \Gamma_2 \end{pmatrix}}_{C} \underbrace{\begin{pmatrix} y_1 \\ \check{b} \\ y_2 \end{pmatrix}}_{y_b} + \underbrace{\begin{pmatrix} \check{R}_1(TA \cdot y, \check{b}) \\ 0 \\ \check{R}_2(TA \cdot y, \check{b}) \end{pmatrix}}_{\check{R}_b(TA \cdot y, \check{b})};$$

4. In order to put matrix C into Jordan canonical form, we make another linear transformation $y_b = TC \cdot \xi$, and get $\dot{\xi} = TC^{-1} \cdot C \cdot TC \cdot \xi + TC^{-1} \cdot \check{R}_b(TA \cdot TC \cdot \xi, \check{b})$, where $\xi = (\xi_1, \check{b}, \xi_2, \xi_3, \xi_4)$. Therefore, we can now write

the system as:

$$
\begin{aligned}
\dot{\xi}_1 &= \Pi_1\xi_1 + \check{R}_1(\xi_1,\xi_2,\xi_3,\xi_4,\check{b}) \\
\dot{\xi}_2 &= \Pi_2\xi_2 + \check{R}_2(\xi_1,\xi_2,\xi_3,\xi_4,\check{b}) \\
\dot{\xi}_3 &= \Pi_3\xi_3 + \check{R}_3(\xi_1,\xi_2,\xi_3,\xi_4,\check{b}) \\
\dot{\xi}_4 &= \Pi_4\xi_4 + \check{R}_4(\xi_1,\xi_2,\xi_3,\xi_4,\check{b}) \\
\dot{\check{b}} &= 0;
\end{aligned}
$$

5. Take $\xi_i = \Upsilon_i(\xi_1,\check{b}) \quad (i = 2,3,4)$ as a polynomial approximation to the center manifold, and differentiate both sides w.r.t. t :

$$\Pi_i\xi_i + \check{R}_i(\xi_1,\Upsilon_2,\Upsilon_3,\Upsilon_4,\check{b}) = D_{\xi_1}\Upsilon_i(\xi_1,\check{b})[\Pi_1\xi_1 + \check{R}_1(\xi_1,\Upsilon_2,\Upsilon_3,\Upsilon_4,\check{b})].$$

   And then solve for the center manifold by equating the coefficient of each order;

6. Finally, we write the differential equation for the dynamical system on the center manifold by substituting $\Upsilon_i(\xi_1,\check{b})$ in $\check{R}_1(\xi_1,\xi_2,\xi_3,\xi_4,\check{b})$, and get the system:

$$
\begin{aligned}
\dot{\xi}_1 &= \Pi_1\xi_1 + \check{R}_1(\xi_1,\Upsilon_2(\xi_1,\check{b}),\Upsilon_3(\xi_1,\check{b}),\Upsilon_4(\xi_1,\check{b}),\check{b}) \\
\dot{\check{b}} &= 0.
\end{aligned}
$$

So in our model the dynamics on the center manifold is given by:

$$\dot{\xi}_1 = \overline{\alpha}\xi_1\left(\xi_1 - \frac{1}{\overline{\alpha}}(b - (\overline{\alpha} - \underline{\gamma}))\right).$$

Appendix B.2. Stability analysis for disease-endemic case

For the Jacobian matrix $\mathcal{J}_3$, we have:

$$
\begin{aligned}
&a_{11} = \theta, a_{13} = f_2^*, a_{14} = a_{15} = -1, a_{22} = -\delta - (b-d), a_{25} = g'^*, a_{32} = (1-l^*)(\gamma'^* - \alpha'^* l^*) \\
&a_{33} = b + \gamma^* - \alpha^*, a_{41} = c^* f_{11}^*, a_{43} = c^* f_{12}^*, a_{51} = -f_{11}^*\frac{g'^*}{g''^*}, a_{52} = \frac{f_1^*(\gamma''^* - \alpha''^* l^*)}{\gamma'^* - \alpha'^* l^*}\frac{g'^*}{g''^*} \\
&a_{53} = \left(\frac{f_1^*(2\alpha'^* l^* - \alpha'^* - \gamma'^*)}{(1-l^*)(\gamma'^* - \alpha'^* l^*)} - f_{12}^*\right)\frac{g'^*}{g''^*}, a_{54} = \frac{f_1^*}{c^*}\frac{g'^*}{g''^*}, a_{55} = f_1^*, a_{56} = \frac{f_1^*}{\lambda_3^*}\frac{g'^*}{g''^*} \\
&a_{61} = -\frac{f_{12}^*}{c^*}, a_{62} = -\lambda_3^*(2\alpha'^* l^* - \gamma'^* - \alpha'^*), a_{63} = -\frac{f_{22}^*}{c^*} - 2\lambda_3^*\alpha^*, a_{64} = \frac{f_2^*}{c^{*2}}, a_{66} = \frac{f_2^*}{c^*\lambda_3^*}.
\end{aligned}
$$

Let us denote $\varpi_1 = \delta + b - d, \varpi_2 = \alpha^* - \gamma^* - b$, we have the following equations which will be used in the calculation.

$$\lambda_3^* = \frac{f_2^*}{c^*(\theta - 2\alpha^* l^* + b + \gamma^* + \alpha^*)} = \frac{f_2^*}{c^*(\theta + \varpi_2)}$$

$$a_{22} = -\varpi_1, a_{33} = -\varpi_2 \tag{B.1}$$

$$a_{55} = f_1^* = \theta + (\delta + b - d) = \theta + \varpi_1 \tag{B.2}$$

$$a_{66} = \frac{f_2^*}{c^* \lambda_3^*} = \theta - b - \gamma^* + \alpha^* = \theta + \varpi_2 \tag{B.3}$$

$$a_{66}a_{54} = a_{56}a_{64} = \frac{f_1^* f_2^*}{\lambda_3^*} \frac{g'^*}{g''^*} \frac{1}{c^{*2}} \tag{B.4}$$

$$a_{41}a_{54} = c^* f_{11}^* \frac{f_1^*}{c^*} \frac{g'^*}{g''^*} = f_{11}^* \frac{g'^*}{g''^*} f_1^* = -a_{51}a_{55} = -a_{51}(\varpi_1 + \theta) \tag{B.5}$$

Since

$$f_1^* = f_2^* g'^* \frac{(1 - l^*)(\gamma'^* - \alpha'^* l^*)}{\theta + \alpha^* - b - \gamma^*} = \frac{a_{13}a_{25}a_{32}}{a_{66}}$$

we also get

$$a_{55}a_{66} = a_{13}a_{25}a_{32} = (\theta + \varpi_1)(\theta + \varpi_2). \tag{B.6}$$

$$a_{41}a_{56}a_{64} = a_{41}a_{54}a_{66} = -a_{51}(\varpi_1 + \theta)(\varpi_2 + \theta). \tag{B.7}$$

As $\lambda_3^* c^* = \frac{f_2^*}{a_{66}}$, we have

$$\begin{aligned} a_{56}a_{61} &= \frac{f_1^*}{\lambda_3^*} \frac{g'^*}{g''^*} \left(-\frac{f_{12}^*}{c^*} \right) = \left(\frac{-g'^* f_{12}^*}{g''^*} \right) \frac{(\varpi_1 + \theta)}{\lambda_3^* c^*} = \\ &= \left(\frac{-g'^* f_{12}^*}{g''^*} \right) \frac{a_{55}a_{66}}{f_2^*} = \left(\frac{-g'^* f_{12}^*}{g''^*} \right) \frac{a_{13}a_{25}a_{32}}{f_2^*} = \left(\frac{-g'^* f_{12}^*}{g''^*} \right) a_{25}a_{32}. \end{aligned}$$

Thus,

$$\begin{aligned} &a_{25}a_{56}a_{62} + a_{25}a_{32}a_{53} - a_{56}a_{61} = a_{25}a_{56}a_{62} + a_{25}a_{32}a_{53} + \frac{g'^* f_{12}^*}{g''^*} a_{25}a_{32} \\ &= a_{25}[-\frac{f_1^*}{\lambda_3^*} \frac{g'^*}{g''^*} \lambda_3^*(2\alpha'^* l^* - \gamma'^* - \alpha'^*) + (1 - l^*)(\gamma'^* - \alpha'^* l^*) \left(\frac{f_1^*(2\alpha'^* l^* - \alpha'^* - \gamma'^*)}{(1 - l^*)(\gamma'^* - \alpha'^* l^*)} - f_{12}^* \right) \frac{g'^*}{g''^*} \\ &+ \frac{g'^* f_{12}^*}{g''^*}(1 - l^*)(\gamma'^* - \alpha'^* l^*)] = 0. \end{aligned} \tag{B.8}$$

$$a_{54}a_{43}a_{25}a_{32} = \frac{f_1^*}{c^*} \frac{g'^*}{g''^*} c^* f_{12}^* a_{25}a_{32} = (-f_1^*)(\frac{-g'^* f_{12}^*}{g''^*}) a_{25}a_{32} = -(\varpi_1 + \theta)a_{56}a_{61} \tag{B.9}$$

The characteristic equation, $|\Lambda I - \mathcal{J}_3| = 0$ can be written as a polynomial of Λ:

$$P(\Lambda) = \Lambda^6 - D_1\Lambda^5 + D_2\Lambda^4 - D_3\Lambda^3 + D_4\Lambda^2 - D_5\Lambda + D_6 = 0$$

where the D_i are the sum of the i-th order minors about the principal diagonal of $\mathcal{J}_3$.

Thus,

$$D_1 = a_{11} + a_{22} + a_{33} + a_{55} + a_{66} = 3\theta > 0$$

$$\begin{aligned} D_2 &= a_{11}(a_{22}+a_{33}+a_{55}+a_{66})+a_{41}+a_{51} \\ &\quad +a_{22}(a_{33}+a_{55}+a_{66})-a_{52}a_{25}+a_{33}(a_{55}+a_{66})+a_{55}a_{66} \end{aligned}$$

Replace

$$a_{22}+a_{33}+a_{55}+a_{66}=2\theta$$

$$a_{33}+a_{55}+a_{66}=2\theta+\varpi_1$$

$$a_{55}+a_{66}=2\theta+\varpi_1+\varpi_2$$

we get

$$D_2=3\theta^2-\theta(\varpi_1+\varpi_2)-\varpi_1^2-\varpi_2^2+a_{41}+a_{51}-a_{25}a_{52}.$$

$$\begin{aligned} D_3 &= a_{11}a_{22}a_{33}+a_{22}a_{41}+a_{11}a_{22}a_{55}+a_{22}a_{51}-a_{11}a_{25}a_{52}+a_{11}a_{22}a_{66} \\ &\quad +a_{33}a_{41}+a_{11}a_{33}a_{55}+a_{33}a_{51}+a_{11}a_{33}a_{66}-a_{41}a_{54}+a_{41}a_{55} \\ &\quad +a_{41}a_{66}+a_{11}a_{55}a_{66}-a_{56}a_{61}+a_{51}a_{66} \\ &\quad +a_{22}a_{33}a_{55}+a_{32}a_{25}a_{53}-a_{25}a_{52}a_{33}+a_{22}a_{33}a_{66} \\ &\quad +a_{22}a_{55}a_{66}+a_{25}a_{56}a_{62}-a_{25}a_{52}a_{66}+a_{33}a_{55}a_{66} \end{aligned}$$

We keep only $a_{41}, a_{25}a_{52}, a_{25}a_{56}a_{62}, a_{25}a_{32}a_{53}, a_{56}a_{61}$ in the expression, replace $a_{11}, a_{22}, a_{33}, a_{55}, a_{66}$ via $\varpi_1, \varpi_2, \theta$ from (B.1)-(B.3), use (B.5) and (B.8) to obtain

$$D_3=\theta[\theta^2-2\theta(\varpi_1+\varpi_2)-2(\varpi_1^2+\varpi_2^2)]+2\theta(a_{41}+a_{51}-a_{25}a_{52})$$

$$\begin{aligned} D_6 &= a_{66}[a_{55}a_{22}a_{33}a_{41}-a_{25}a_{32}a_{43}a_{51}-a_{25}a_{33}a_{41}a_{52}+a_{25}a_{32}a_{41}a_{53} \\ &\quad -a_{25}a_{54}a_{11}a_{32}a_{43}+a_{25}a_{54}a_{13}a_{32}a_{41}-a_{54}a_{22}a_{33}a_{41}]+a_{56}a_{64}a_{22}a_{33}a_{41} \\ &\quad +a_{56}a_{25}[a_{64}a_{11}a_{32}a_{43}-a_{64}a_{13}a_{32}a_{41}+a_{32}a_{43}a_{61}+a_{33}a_{41}a_{62}-a_{32}a_{41}a_{63}] \end{aligned}$$

Proof of Lemma 4

Proof. By using (B.4) we rewrite D_6 as follows :

$$\begin{aligned}
D_6 &= a_{66}(a_{55}a_{22}a_{33}a_{41} - a_{25}a_{33}a_{41}a_{52}) + a_{56}a_{25}(a_{32}a_{43}a_{61} + a_{33}a_{41}a_{62} - a_{32}a_{41}a_{63}) + a_{66}a_{25}a_{32}(a_{41}a_{53} - a_{43}a_{51}) \\
&\quad a_{11}a_{32}a_{43}a_{25}(a_{56}a_{64} - a_{66}a_{54}) + a_{13}a_{32}a_{41}a_{25}(a_{66}a_{54} - a_{56}a_{64}) + a_{22}a_{33}a_{41}(a_{56}a_{64} - a_{66}a_{54}) \\
&= a_{66}(a_{55}a_{22}a_{33}a_{41} - a_{25}a_{33}a_{41}a_{52}) + a_{56}a_{25}[a_{32}(a_{43}a_{61} - a_{41}a_{63}) + a_{33}a_{41}a_{62}] + a_{66}a_{25}a_{32}(a_{41}a_{53} - a_{43}a_{51}).
\end{aligned}$$

The first term $a_{66}(a_{55}a_{22}a_{33}a_{41} - a_{25}a_{33}a_{41}a_{52}) < 0$.

Note that $a_{56}a_{25} < 0$, $a_{33} = (b + \gamma^* - \alpha^*) = \alpha^*(l^* - 1)$ and by concavity of f, $f_{12}^{*2} < f_{11}^* f_{22}^*$. We show the second term

$$\begin{aligned}
& a_{56}a_{25}[a_{32}(a_{43}a_{61} - a_{41}a_{63}) + a_{33}a_{41}a_{62}] \\
=\ & a_{56}a_{25}[a_{32}(-f_{12}^{*2} + f_{11}^* f_{22}^* + 2c^* f_{11}^* \lambda_3^* \alpha^*) - a_{33}c^* f_{11}^* \lambda_3^* (2\alpha'^* l^* - \gamma'^* - \alpha'^*)] \\
<\ & a_{56}a_{25}[c^* f_{11}^* \lambda_3^* (2\alpha^* a_{32} - a_{33}(2\alpha'^* l^* - \gamma'^* - \alpha'^*))] \\
=\ & a_{56}a_{25}[c^* f_{11}^* \lambda_3^* (2\alpha^*(1 - l^*)(\gamma'^* - \alpha'^* l^*) - \alpha^*(l^* - 1)(2\alpha'^* l^* - \gamma'^* - \alpha'^*))] \\
=\ & a_{56}a_{25}c^* f_{11}^* \lambda_3^* \alpha^*(1 - l^*)(\gamma'^* - \alpha'^*) \\
=\ & f_1^* \frac{g'^{*2}}{g''^*} c^* f_{11}^* \alpha^*(1 - l^*)(\gamma'^* - \alpha'^*)
\end{aligned}$$

and the third term

$$\begin{aligned}
& a_{66}a_{25}a_{32}(a_{41}a_{53} - a_{43}a_{51}) \\
=\ & a_{66}a_{25}a_{32}[c^* f_{11}^* \frac{g'^*}{g''^*}(\frac{f_1^*(2\alpha'^* l^* - \alpha'^* - \gamma'^*)}{(1 - l^*)(\gamma'^* - \alpha'^* l^*)} - f_{12}^*) + c^* f_{12}^* f_{11}^* \frac{g'^*}{g''^*}] \\
=\ & a_{66}a_{25}c^* f_{11}^* \frac{g'^*}{g''^*} f_1^*(2\alpha'^* l^* - \alpha'^* - \gamma'^*) \\
=\ & \frac{f_2^*}{\lambda_3^*} g'^* f_{11}^* \frac{g'^*}{g''^*} f_1^*(2\alpha'^* l^* - \alpha'^* - \gamma'^*).
\end{aligned}$$

So we have

$$\begin{aligned}
& a_{56}a_{25}[a_{32}(a_{43}a_{61} - a_{41}a_{63}) + a_{33}a_{41}a_{62}] + a_{66}a_{25}a_{32}(a_{41}a_{53} - a_{43}a_{51}) \\
<\ & \frac{g'^{*2}}{g''^*} f_1^* c^* f_{11}^* [\alpha^*(1 - l^*)(\gamma'^* - \alpha'^*) + \frac{f_2^*}{c^* \lambda_3^*}(2\alpha'^* l^* - \alpha'^* - \gamma'^*)] \\
=\ & \frac{g^{2\prime *}}{g''^*} f_1^* c^* f_{11}^* [\alpha^*(1 - l^*)(\gamma'^* - \alpha'^*) + (\theta - b - \gamma^* + \alpha^*)(2\alpha'^* l^* - \alpha'^* - \gamma'^*)]. \\
=\ & \frac{g^{2\prime *}}{g''^*} f_1^* c^* f_{11}^* [(\alpha^* - b - \gamma^*)(\gamma'^* - \alpha'^*) + (\theta - b - \gamma^* + \alpha^*)(\frac{2\alpha'^*(b + \gamma^*)}{\alpha^*} - \alpha'^* - \gamma'^*)] \\
<\ & 0 \text{ by A.11(1).}
\end{aligned}$$

Hence we shown $\det \mathcal{J}_3 = D_6 < 0$. ■

Proof of Lemma 5

Proof.

$$\begin{aligned}D_1D_2 - D_3 &= 3\theta[3\theta^2 - \theta(\varpi_1 + \varpi_2) - \varpi_1^2 - \varpi_2^2] - \theta[\theta^2 - 2\theta(\varpi_1 + \varpi_2) - 2(\varpi_1^2 + \varpi_2^2)] \\ &\quad +3\theta[a_{41} + a_{51} - a_{52}a_{25}] - [2\theta a_{41} + 2\theta a_{51} - 2\theta a_{25}a_{52}] \\ &= \theta[8\theta^2 - \theta(\varpi_1 + \varpi_2) - \varpi_1^2 - \varpi_2^2] + \theta(a_{41} + a_{51} - a_{25}a_{52}),\end{aligned}$$

which is negative since $a_{41} + a_{51} - a_{25}a_{52} < 0$ and $8\theta^2 - \theta(\varpi_1 + \varpi_2) - \varpi_1^2 - \varpi_2^2 < 0$ from A.11(2). Furthermore

$$\theta^2 - 2\theta(\varpi_1 + \varpi_2) - 2(\varpi_1^2 + \varpi_2^2) < 3\theta^2 - \theta(\varpi_1 + \varpi_2) - \varpi_1^2 - \varpi_2^2 < 8\theta^2 - \theta(\varpi_1 + \varpi_2) - \varpi_1^2 - \varpi_2^2 < 0.$$

So we get $D_2 < 0$ and $D_3 < 0$. ∎