



**HAL**  
open science

# SIR Economic Epidemiological Models with Disease Induced Mortality

Aditya Goenka, Lin Liu, Manh-Hung Nguyen

► **To cite this version:**

Aditya Goenka, Lin Liu, Manh-Hung Nguyen. SIR Economic Epidemiological Models with Disease Induced Mortality. *Journal of Mathematical Economics*, 2021, 93, 10.1016/j.jmateco.2021.102476 . hal-03170689

**HAL Id: hal-03170689**

**<https://hal.science/hal-03170689>**

Submitted on 16 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# SIR Economic Epidemiological Models with Disease Induced Mortality\*

Aditya Goenka<sup>†</sup>      Lin Liu<sup>‡</sup>      Manh-Hung Nguyen<sup>§</sup>

January 11, 2021

## Abstract

This paper studies an optimal growth model where there is an infectious disease with *SIR* dynamics which can lead to mortality. Health expenditures (alternatively intensity of lockdowns) can be made to reduce infectivity of the disease. We study implications of two different ways to model the disease related mortality - early and late in infection mortality - on the equilibrium health and economic outcomes. In the former, increasing mortality reduces infections by decreasing the fraction of infectives in the population, while in the latter the fraction of infectives increases. We characterize the steady states and the outcomes depend in the way mortality is modeled. With early mortality, increasing mortality leads to higher equilibrium per capita output and consumption while in the late mortality model these decrease. We establish sufficiency conditions and provide the first results in economic models with *SIR* dynamics with and without disease related mortality - a class of models which are non-convex and have endogenous discounting so that no existing results are applicable.

**Keywords:** Infectious diseases, Covid-19, SIR model, mortality, sufficiency conditions, economic growth, lockdown, prevention, health expenditure.

**JEL Classification:** *E13, E22, D15, D50, D63, I10, I15, I18, O41, C61.*

---

\*We thank the referees for their helpful comments. The usual disclaimer applies. Manh-Hung Nguyen acknowledges support from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program).

<sup>†</sup>Department of Economics, University of Birmingham, Email: a.goenka@bham.ac.uk

<sup>‡</sup>Management School, University of Liverpool, Email: lin.liu@liverpool.ac.uk

<sup>§</sup>Toulouse School of Economics, INRAE, University of Toulouse Capitole, Email: manh-hung.nguyen@tse-fr.eu

# 1 Introduction

The Covid-19 pandemic has brought the study of interaction of infectious disease with the economy, i.e. economic epidemiology models, to the frontier of economic research. The first generation of economic epidemiology models typically studied *SIS* models without disease related mortality.<sup>1</sup> However, with Covid-19 the modeling of mortality has become important as this seems to be a driver of the policy responses adopted in many countries.<sup>2</sup>

This paper analyses the *SIR* model which has been used to model Covid-19 (Ferguson, et al. (2020)) when there is disease related mortality<sup>3</sup> in a neoclassical growth model so that the model is fully general equilibrium.<sup>4</sup> The literature has largely studied immediate effects of Covid-19 and in this paper we concentrate on the medium to long run effects, i.e. effects of the disease on the steady state.<sup>5</sup> Households can save through investing in capital and production of the single consumption good uses capital and labor. Only those that are not infected (i.e. those who are susceptible and those recovered from the disease) individuals can work. There are two effects of the disease: there is morbidity, i.e. those who are ill do not work, and there is mortality, so that a fraction of those who have contracted the disease die due to it. The contact rate is endogenous in the model and is decreasing in health expenditures, which can also be interpreted as self-isolation costs (Eichenbaum, et al. (2020)).<sup>6</sup> In our model the households are homogeneous and we do not model disease related externality where households do not take into account the effect of their decisions on the evolution of the disease in the population. The recent paper, Goenka and Liu (2020) explores in detail the effect of this health externality<sup>7</sup> in a dynamic general equilibrium model with health expenditures and we abstract from it to concentrate on the role of mortality in modeling diseases of the *SIR* type. In the model we use an extended welfare function

---

<sup>1</sup>See et al. (2004), Goenka and Liu (2013, 2020), Goenka, Liu and Nguyen (2014), and Toxvaerd (2019) and the references there.

<sup>2</sup>Boucekkine and Laffargue (2010) and Chakraborty, et al. (2010) model disease related mortality in the overlapping generations framework but did not use a compartment epidemiology model as in this paper and the emerging literature.

<sup>3</sup>Whether there is long-lasting immunity in Covid-19 is contentious. Long, et al. (2020) using data from China find evidence consistent with steep decline in 2-3 months. Similar results were found in a study in the US (Ibarrando, et al. (2020)). On the other hand Wajnberg, et al. (2020) and Sekine, et al. (2020) find evidence suggesting longer immunity. Dan, et al. (2021) find evidence that suggests immunity lasts at least 8 months in a cohort study. As in our companion paper we have studied issues related to the *SIS* model, here we study implications of the alternate, *SIR* model.

<sup>4</sup>Goenka, Liu and Nguyen (2020) in a companion paper study optimal lockdown and other issues in a *SIS* model with disease related mortality.

<sup>5</sup>There are other papers using the *SIR* model with disease related mortality e.g. Acemoglu, et al. (2020), Alvarez, et al. (2020), Eichenbaum, et al. (2020), and Jones, et al. (2020). These papers concentrate on short-run models and are not fully general equilibrium as all variables (wages, interest rate, and capital) do not adjust.

<sup>6</sup>Goenka and Liu (2020) and Goenka, Liu and Nguyen (2020) modeled optimal health expenditures in a similar growth framework. These papers treated health expenditures as accumulating in health capacity which is important for understanding response to Covid-19 and the mortality due to it. This introduces another state variable and in this paper to simplify the analysis we treat health expenditures as only a flow variable.

<sup>7</sup>This has been modeled in different ways in the literature, see Geoffard and Philipson (1996), Gersovitz and Hammer (2004), Goenka and Liu (2020), and Hellwig, et al. (2020).

that depends on utility from consumption as well as a loss in welfare from disease related mortality as without it there can be counter-intuitive effects where increase in mortality is welfare improving.<sup>8</sup>

There are two main methodological issues that we feel have not received adequate attention which we study in the paper. The first is examining different modeling choices for modeling mortality in the *SIR* model. We present a general model where encompasses the two canonical ways of modeling mortality in the epidemiology *SIR* model - early mortality so that there is immediate death of the infectives, and late in infection (delayed) mortality - where death takes place later so that those who succumb to the disease are those who are not circulating in the population transmitting the disease (see Busenberg and van den Driessche (1990) and Keeling and Rohani (2008)). This is consistent with the experience in Covid-19 where most of the mortality has taken place in hospitals and care homes so that these individuals are not effectively mixing with the general population of susceptibles. The timing of mortality affects epidemiology dynamics. We examine the implications of the epidemiology modeling choices on optimal choices and on the equilibrium steady state outcomes. When there is early mortality, the direct effect of higher deaths is drop in fraction of infectives which reduces infections, while with later mortality this does not happen. Thus, there is a self-limiting effect of increase in mortality. This has implications for the cut-off for persistence,  $R_0$ , and the long run effect of the disease.

The second methodological issue is that we study the sufficiency conditions for the optimal control problem. The *SIR* epidemiology dynamics are non-convex. Endogenous mortality adds another problem as the population becomes endogenous. The economic epidemiology *SIR* model, thus, has endogenous discounting and is non-convex. As a result the standard Arrow or Mangasarian conditions do not apply. There are no results for the *SIR* model with and without mortality that can be used in economic models to our knowledge.<sup>9</sup> We directly address this issue and given the special structure of the problem, we directly show the relevant transversality conditions and establish sufficiency by adapting the method of Leitmann and Stalford (1970) that was used for convex problems.<sup>10</sup> The key to the proof in the current paper is to show the co-state variables associated with the bounded state variables converge to zero with time, and this implies a different transversality condition than in Goenka, Liu and Nguyen (2020). As a special case, we obtain the sufficiency in the *SIR* model without disease related mortality.

The model is a fully dynamic general equilibrium model and we characterize the Euler equations that govern the evolution of the economy. As our interest is beyond the very short run, we show that there are two steady states for the economy: a disease free and disease endemic steady state. The optimal health expenditure depends on a function of the parameters and the equilibrium values of the economic variables. This function is interpreted

---

<sup>8</sup>See Goenka, Liu and Nguyen (2020) for extensive discussion of the issues associated with modeling the objective function.

<sup>9</sup>The transversality and sufficiency conditions for convex endogenous population models (see Boucekkine, et al. (2018)) do not apply.

<sup>10</sup>Sufficiency for a neoclassical growth model with *SIS* disease dynamics without disease related mortality was established by Goenka, Liu and Nguyen (2014). The recent work, Goenka, Liu and Nguyen (2020) establishes sufficiency for the *SIS* model with disease related mortality. There are differences between the *SIS* and *SIR* models: that there is one less state variable for the epidemiology dynamics in the *SIS* models.

as the net marginal benefit of health expenditure (net of the marginal cost) and its position determines whether health expenditures (or non-pharmaceutical interventions such as ‘lock-downs that reduce economic activity’) are positive or not in a steady state, and if they are, their magnitude. Thus, the equilibrium reproduction rate,  $R_0^*$  will depend on both the infectivity of the disease and endogenous economic choices. As the models are too complex to solve analytically we study how the steady state values of variables of interest change as the rate of disease related mortality changes (and thus, the number of disease related deaths in equilibrium). While the actual disease dynamics will typically have both early and late in infection mortality to understand how each affects equilibrium outcomes we analyze the polar cases where there is only mortality of one kind in detail. The endogenous economic choices interact with the inherent disease dynamics. With early mortality model the death of the infectives directly reduces infections and there is an indirect concentration effect as the population size also reduces. If the direct effect dominates then the per capita capital stock, output and consumption increase as those infected only consume but do not work. When there is late mortality, death does not directly reduce infections as those who succumb to the disease are not transmitting infection in the general population of susceptibles. Capital stock, output and welfare decline as infections are increasing - the fraction of the infectives in the population is increasing. Thus, the details of the epidemiology modeling will affect economic outcomes and how to think about optimal control of epidemics.

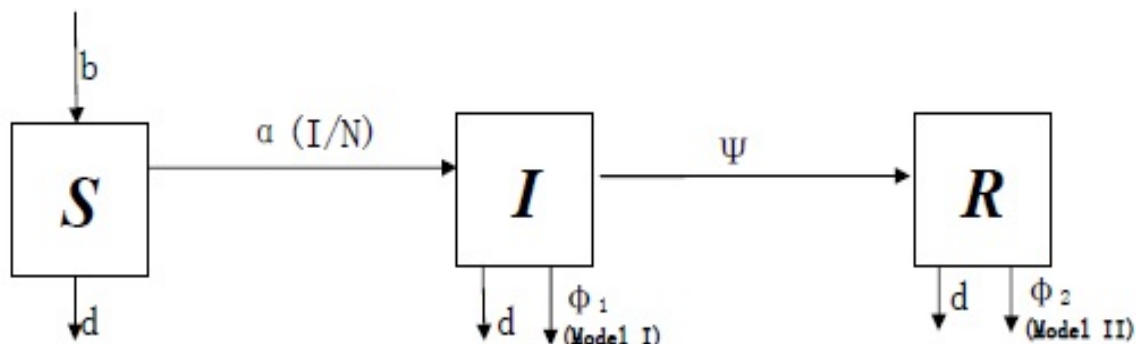
The plan of the paper is as follows: Section 2 studies the pure *SIR* model with early and late mortality. Section 3 introduces the economic epidemiology model, and characterizes the Euler equations and steady states, and Section 4 does comparative statics of equilibrium steady state outcomes when mortality is varied. Section 5 studies the transversality and sufficiency conditions, and Section 6 concludes.

## 2 The Epidemiological Model

In this section, we introduce the simplest *SIR* model with disease induced mortality. The transfer diagram is shown in Figure 1. The total population,  $N$ , is divided into three groups:  $S$ , the susceptible (healthy and susceptible to the disease),  $I$ , the infective (infected and capable of transmitting the disease), and the recovered,  $R$  (recovered with subsequent immunity to the disease). Individuals are born at the rate  $b$ , healthy and susceptible to the disease. We assume homogeneous mixing so that the likelihood of any individual contracting the disease is the same. There is horizontal incidence of the disease i.e. transmission from peers. Let  $\alpha$  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  $\alpha(I/N)S$ , depending on the fraction of the infected. This contact structure is the standard incidence or frequency dependant model, commonly used in the epidemiology literature for human diseases. It is adopted as the pattern of human interaction is relatively stable and invariant to the size of the population. The recovery of individuals is governed by the parameter  $\psi$  and the total number of individuals who recover from the disease at each time period is  $\psi I$ . Each individual faces the exogenous death rate,  $d$ , irrespective of health status.

In the epidemiology literature, there are two ways of introducing disease induced mor-

**Figure 1.** The Transfer Diagram For the *SIR* Epidemiology Model with mortality



Note: The total population is divided into three groups: the susceptibles,  $S$ , the infectives,  $I$  and recovered,  $R$ . The birth rate is  $b$  and newborns are born healthy and susceptible. All individuals irrespective of health status die at the rate  $d$ . The susceptible get infected at the rate  $\alpha I/N$  and the infected recover at the rate  $\psi$ . In model I (early mortality), the infected die at the rate  $\phi_1$  due to disease, while in model II (late mortality) the recovered die at the rate  $\phi_2$ .

tality. The first is *early mortality* where the infected die at the rate  $\phi_1$ <sup>11</sup> due to the disease. The second is *late in infection mortality* where the recovered die at the rate  $\phi_2$  due to the disease.<sup>12</sup> This difference between the two affects the disease dynamics. In the first way of modeling (Model I) it affects transmission directly as size of infectives drops due to mortality while in the second (Model II) it does not.

Mortality among the recovered but not in infectives may seem strange at first. Whether infectives are asymptomatic, have mild, or severe symptoms, it is their coming into contact with susceptibles that leads to the transmission of the disease. Early mortality reduces the fraction of infectives and thus, reduces the transmission of the disease. The interpretation of late in infection or delayed mortality is that the seriously ill have been identified and isolated so their mortality does not change the fraction of infectives mixing in the population. This seems to be the case in the case of Covid-19. In Covid-19 it is those with severe symptoms who die from the disease and it seems that a large part of this mortality has taken in either hospitals or care homes. Thus, one can think of them not circulating in the population and thus, one can treat this group in the recovered group from a modeling point of view. This is consistent with the evidence that risk of infections in the community are higher than that in hospitals, i.e. nosocomial infections (Carter, et al. (2020)). The mortality in the recovered group also covers pre-mature mortality of those who have been treated and free of the disease. While Covid-19 is too recent to have data on this, there are several other

<sup>11</sup> $\phi$  is similar to the Infection Fatality Rate which also covers those that are untested and asymptomatic rather than the Case Fatality Rate which depends on the number of individuals diagnosed with the disease.

<sup>12</sup>The pure *SIR* epidemiology models with varying population size are well studied in mathematical biological literature (see Busenberg and van den Driessche (1990), Hethcote (2008), Keeling and Rohani (2008) and Mena-Lorca and Hethcote (1992)).

infectious diseases where those who have successfully been treated have shorter conditional life expectancies: for example for TB it is 3.6 years (Hoger, et al. (2016), and for HIV it is 6.8 years (9.5 years without any co-morbidities) (Marcus, et al. (2020)). There is emerging evidence that Covid-19 induces damage to lungs - Ground-glass opacities and lesions, and long term lung fibrosis (Bernheim, et al. (2020), Hosseiny, et al. (2020), and Li and Xia (2020)) and to the heart tissue (see PÃ©rez-Bermejo, et al. (2020)) and thus, one may expect impairment of life expectancy. Thus, in this case while there is only a morbidity effect on infectives there is an increased mortality effect on the recovered.

We can write the model incorporating both early and late in infection mortality the following system of differential equations :

$$\begin{aligned}\dot{S} &= bN - \frac{\alpha SI}{N} - dS \\ \dot{I} &= \frac{\alpha SI}{N} - \psi I - dI - \phi_1 I \\ \dot{R} &= \psi I - dR - \phi_2 R \\ \dot{N} &= bN - dN - (\phi_1 I + \phi_2 R).\end{aligned}$$

The outflow of susceptibles is due to new infections,  $\alpha(I/N)S$ , and exogenous death  $dS$ . All individuals are born healthy and thus there is a flow into the susceptible class,  $bN$ .<sup>13</sup> The infectives grow by  $\alpha(I/N)S$  and infectives recover at rate  $\psi$  or succumb to the disease at rate  $\phi_1$ . There is also mortality from other causes given by  $dI$ . The individuals who recover from the disease have subsequent immunity from the disease and inflow is given by  $\psi I$  and there is mortality from other causes  $dR$  as well as the disease at the rate  $\phi_2$  The population changes due to net births,  $(b - d)N$  and disease related mortality,  $\phi_1 I + \phi_2 R$ .

We impose the following assumptions on the demographic and epidemiological parameters.

**Assumption 1.** *We assume the parameters in the SIR model satisfy:*

1. *The demographic parameters  $b$  and  $d$  are non-negative, and  $b \geq d + \phi_1 + \phi_2$ ;*
2. *The recovery rate  $\psi > 0$ ;*
3. *The disease induced mortality rate  $\phi_1 \geq 0$  and  $\phi_2 \geq 0$ .*

The first assumption ensures that the total population grows over time, regardless of the level of disease prevalence, otherwise there may be extinction of the population. In a *SIR* model with mortality the dynamic properties depend on the evolution of the population size. We make the above assumption which is consistent with existence of steady states with a growing population. Busenberg and van den Driessche (1990) have a weaker condition that depends on the properties of the steady state fractions of susceptibles and infectives. Mena-Lorca and Hethcote (1992) study the case of only early mortality and have a weaker

---

<sup>13</sup>These are new entrants to the population.

condition that depends only on the parameters and not on the steady state.<sup>14</sup> Here, we assume this stronger condition which is meaningful and easy to interpret. We assume the recovery rate is positive as this is an *SIR* model. We do not impose any assumption on the contact rate  $\alpha$  as it is endogenous in the economic epidemiological models in the next section. The contact rate in the epidemiology literature is understood to depend not only on the biology of the disease but also on behavioural, social and policy (e.g. non-pharmaceutical interventions) considerations.

Since  $N = S + I + R$ , we define  $i = I/N$  and  $r = R/N$  and rewrite the model as:

$$\dot{i} = \alpha(1 - i - r)i - \psi i - bi - \phi_1 i + i(\phi_1 i + \phi_2 r) \quad (1)$$

$$\dot{r} = \psi i - br - \phi_2 r + r(\phi_1 i + \phi_2 r) \quad (2)$$

$$\frac{\dot{N}}{N} = b - d - (\phi_1 i + \phi_2 r). \quad (3)$$

Note that the proportion of the susceptible  $s = R/N = 1 - i - r$ . We denote the steady state of a variable  $x$  as  $x^\infty$  to distinguish it from the optimal value in a trajectory in the later part of the paper which is denoted as  $x^*$ .

**Proposition 1.** (*Busenberg and van den Driessche (1990), Mena-Lorca and Hethcote (1992)*)  
*Consider the epidemiological model given by equation (1) - (3). Then*

1. *The disease free steady state with  $i^\infty = 0$ ,  $r^\infty = 0$  and  $s^\infty = 1$  always exists. It is stable when  $\frac{\alpha}{b+\psi+\phi_1} \leq 1$ , and unstable when  $\frac{\alpha}{b+\psi+\phi_1} > 1$ ;*
2. *When  $\frac{\alpha}{b+\psi+\phi_1} > 1$ , there exists a unique endemic steady state with  $0 < i^\infty < 1$ ,  $0 < r^\infty < 1$  and  $0 < s^\infty < 1$ , which is stable. The endemic steady state  $(i^\infty, r^\infty)$  is the solution to the following system of equations:*

$$\left( \phi_2 + \frac{(\alpha - \phi_2)\phi_1}{\phi_1 - \alpha} \right) r^2 + \left( -b - \phi_2 + \frac{(\alpha - \phi_2)\psi}{\phi_1 - \alpha} + \phi_1 + \frac{\phi_1(\psi + b)}{\phi_1 - \alpha} \right) r + \psi + \frac{\psi(\psi + b)}{\phi_1 - \alpha} = 0$$

$$i = \frac{\alpha - \phi_2}{\phi_1 - \alpha} r + 1 + \frac{\psi + b}{\phi_1 - \alpha}$$

Note that even though fraction of recovered,  $r$ , is given by a quadratic equation there is only one admissible solution in the range  $0 < r < 1$  in the pure epidemiology model (see Busenberg and van den Driessche (1990)).

## 2.1 Comparing the effects of early and late in infection mortality

To highlight the implications of the different ways to model mortality, we focus on the polar cases - when there is only early mortality and when there is only late mortality. When

<sup>14</sup>Essentially in the first model they assume  $b \geq d + \phi_1 i^\infty$  and in the second model,  $b \geq d + \phi_2 r^\infty$  where  $i^\infty$  and  $r^\infty$  are steady state values of the two variables.



$\phi_2 = 0$  there is only early mortality, we call this case ‘Model I’. When  $\phi_1 = 0$  there is only late mortality and we call this ‘Model II’. The early in infection mortality case (Model I) and the late mortality case (Model II) differ in their modeling strategies for the disease induced mortality rate.

When the disease mortality rate  $\phi_1 = \phi_2 = 0$ , the two models coincide as it is the standard *SIR* model without disease related mortality. The total population grows at a constant rate  $b - d$ , and we have

$$i^\infty = \frac{b(\alpha - \psi - b)}{\alpha(\psi + b)}, \quad r^\infty = \frac{\psi(\alpha - \psi - b)}{\alpha(\psi + b)} \quad \text{and} \quad s^\infty = \frac{b + \psi}{\alpha}.$$

However, the two models differ dramatically when the disease mortality rates  $\phi_1$  and  $\phi_2$  are positive. The cut-off point or the reproduction number  $R_0$  for the eradication of infectious diseases is different for both the models. Proposition 1 shows that the reproduction number  $R_0 = \frac{\alpha}{b + \psi + \phi_1}$  depends on  $\phi_1$  and not on  $\phi_2$  as mortality of the recovered does not affect transmission. Thus, the two models have different  $R_0$ . In model I with  $\phi_2 = 0$ , we have  $R_0 = \frac{\alpha}{b + \psi + \phi_1}$ , and when  $R_0 < 1$  there is only one disease free steady state, which is stable. That is, the disease is eradicated. In model II with  $\phi_1 = 0$ , we have  $R_0 = \frac{\alpha}{b + \psi}$ , and when  $R_0 < 1$  the disease is eradicated. Therefore, in the early mortality model, disease induced mortality rate affects the reproduction number. The higher the disease induced mortality rate  $\phi_1$ , the more likely the disease is eradicated. This is because in model I people who die from the disease are the infected and infectious, and as more infected people die, the lower is the chance to infect others, and it is more likely the disease is eradicated. In contrast, in the late mortality model, disease induced mortality does not affect the reproduction number. This is because in model II, people who die from the disease are those who are not transmitting the disease as they are isolated or recovered from the disease. They are neither infectious nor susceptible, and thus do not have direct effect on the prevalence of infectious diseases.

Furthermore, when the reproduction number  $R_0 > 1$  in both models, that is, the disease is prevalent, the two models’ predictions differ when we vary the disease mortality rates. In particular, two models differ dramatically on the prediction of the proportion of the infected as disease induced mortality rates vary. In model I ( $\phi_2 = 0$ ), there are two effects due to the disease induced mortality rate  $\phi_1$ . There is a direct effect, which leads to a reduction in the proportion of the infected  $i$ . This is captured by the term  $-\phi_1 i$  in equation (1). The other is an indirect concentration effect which is a reduction in the population growth rate, shown in equation (3). It further leads to an increase in the proportion of the infected, captured by the term  $\phi_1 i^2$  in equation (1), and the recovered, captured by the term  $\phi_1 i r$  in equation (2). Therefore, if we assume the direct effect dominates the indirect effect, when disease induced mortality rate  $\phi_1$  increases, the proportion of the infected decreases and it becomes easier to control the spread of diseases. In contrast, in model II ( $\phi_1 = 0$ ), there are also two effects due to disease induced mortality rate  $\phi_2$ . The direct effect leads to a reduction in the proportion of the recovered  $r$ . This is captured by the term  $-\phi_2 r$  in equation (2). The other effect is a reduction in the population growth rate, shown in equation (3). It further leads to an increase in the proportion of the infected, captured by the term  $\phi_2 i r$  in equation (1), and the recovered, captured by the term  $\phi_2^2 r$  in equation (2). This is the indirect concentration effect. If we assume as before that the direct effect dominates the indirect effect, when the

disease induced mortality rate  $\phi_2$  increases, the proportion of the recovered decreases and the proportion of the infected increases and it becomes more challenging to control the spread of diseases.

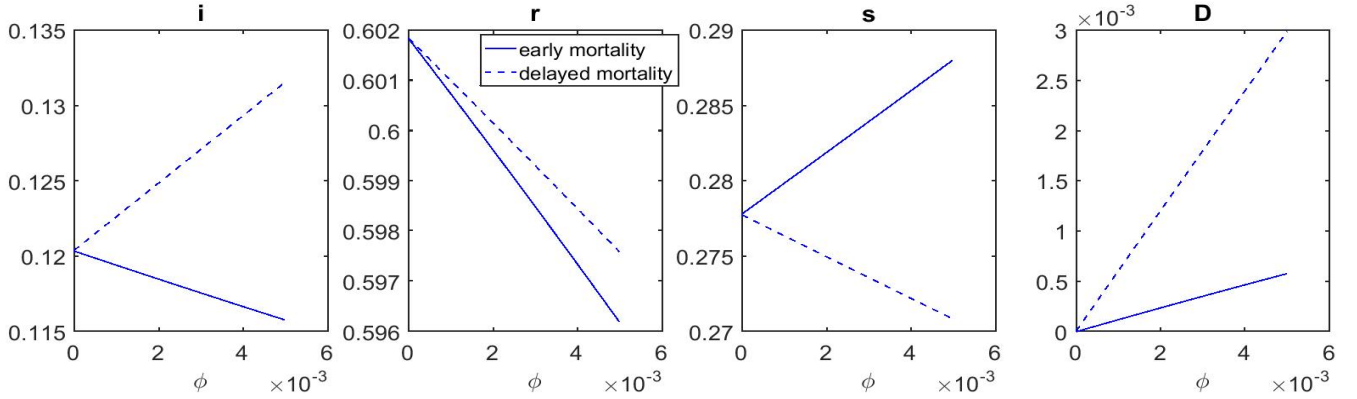
To clearly see the difference between the two models, we conduct the following simulation. Define  $D$  as the proportion of disease induced mortality rate in the total population –  $D = \phi_1 i + \phi_2 r$ . That is, in model I, we have  $D = \phi_1 i$ , while in model II, we have  $D = \phi_2 r$ . We choose the following parameters:  $b = 2\%$ , death rate  $d = 1\%$ , recovery rate  $\psi = 0.1$  (i.e. recovery takes  $\psi^{-1} = 10$  days) and contact rate  $\alpha = 0.4$ . We assume all parameter values in both models are the same, and they are chosen such that both models predict disease endemic steady states. This is shown in Figure 2 - simulation results for model I (the early mortality model) varying  $\phi_1$  and model II (the late mortality model) varying  $\phi_2$ . We show the evolution of  $i$ ,  $r$ ,  $s$  and  $D$  as disease induced mortality rates change. When  $\phi_1 = \phi_2 = 0$  and  $D = 0$ , the two models coincide. When we increase  $\phi_1$  in model I and  $\phi_2$  in model II, the total death rate  $D$  increases. For the same disease induced mortality rates  $\phi_1 = \phi_2$ , there is higher death  $D$  in model II, because in steady state the proportion of the recovered is larger than the proportion of the infected. For model I (the solid line), the general situation improves, as when  $\phi_1$  rises,  $i$  decreases. This is because the infectives transmit the disease and as their proportion drops, the transmission of the disease decreases. In contrast, for model II (the dash line), as  $\phi_2$  increases, the general situation deteriorates as  $i$  increases. This is because as there are more deaths from the group of the recovered - i.e. those that are not circulating in the population transmitting the disease, rather than the group of the infected, the proportion of the infected goes up and the proportion of the recovered goes down. Since there is a larger proportion of the infected, the susceptibles are more likely to get infected and the proportion of the susceptibles is smaller.

Note that the simulation here only focuses on the change in model predictions when we change disease mortality rate  $\phi_1$  and  $\phi_2$ . We ran simulations of changes in the other parameters  $b, d, \psi$  and  $\alpha$ , and the two models predict the similar trend for all the variables.<sup>15</sup> For this reason, the paper focuses on the comparison of the early mortality model and the late mortality model when disease induced mortality rates vary.

---

<sup>15</sup>We do not report the diagrams for conciseness of the paper.

**Figure 2.** Simulation for the early and late in infection mortality models varying disease induced mortality rates



The panels are *i*, *r*, *s* and *D*. The solid line is the simulation for model I, while the dash line is the simulation for model II. The parameter values are  $b = 0.02$ ,  $d = 0.01$ ,  $\alpha = 0.4$  and  $\psi = 0.1$ . The disease induced mortality rates  $\phi_1$  and  $\phi_2$  vary from 0 to 0.5%.

### 3 The Economic Epidemiology Model

For the pure epidemiological model, it matters how we model disease induced mortality. With early mortality, the higher the disease mortality rate, the easier it is to control the spread of infectious diseases, while for the late mortality model, the higher the disease mortality rate, the more challenging it is to control the spread of the disease. What is the consequence of the difference in modeling mortality on economic outcomes? Do we still see the sharp difference in terms of the severity of prevalence of infectious diseases when the response to control the disease is endogenous and how does this affect economic variables such as capital, output, consumption and total welfare?

#### 3.1 The model

The model is based on the growth model with SIS disease dynamics in Goenka and Liu (2013, 2020) and Goenka, Liu and Nguyen (2014, 2020). In this paper we use an underlying neoclassical growth model with *SIR* dynamics with mortality. To avoid keeping track of the history of healthy, infected and recovered individuals, and to stay close to the canonical growth model, we adopt the framework of a large representative household.

**Households:** We assume the economy is populated by a continuum of non-atomic identical households who are the representative decision-making agents. In the absence of the disease, the size of the population in each household grows over time at the rate of  $b - (d + \phi_1 + \phi_2) \geq 0$ , where  $b$  is the birth rate and  $d$  is the death rate. Within each household, an individual is either healthy or infected or recovered from the diseases. We assume that diseases follow the *SIR* dynamics (see the discussion in Section 2).

We model the infectious disease as having two effects - reducing productivity of the infected and disease related mortality. We make the simplifying assumption that when an

infected individual is incapacitated by the disease the productivity falls to zero.<sup>16</sup> We assume the labor is supplied inelastically.<sup>17</sup>

We use an extended welfare function. If a purely utilitarian welfare function is used then mortality can lead to an increase of welfare of the survivors. (see Goenka, Liu and Nguyen (2020) for a further discussion) to avoid counter-intuitive effects due to increasing mortality.<sup>18</sup> When there is no disease prevalence, the disutility from disease mortality is of course zero. Define the total disease related mortality  $D = \phi_1 i + \phi_2 r$ . There is full insurance within the household so that consumption of all members irrespective of health status is the same. The representative household's preferences are given as:

$$\begin{aligned} & \int_0^\infty e^{-\rho t} [u(c_t) - \chi \nu(D_t)] N_t dt \\ &= \int_0^\infty e^{-\int_0^t (\rho - b + d + D(\tau)) d\tau} [u(c_t) - \chi \nu(D_t)] N_0 dt \end{aligned} \quad (4)$$

where  $\rho$  is the discount factor with  $\rho > b - d$ , the initial size of household is assumed to be one.  $\nu(D)$  is the loss in welfare from disease related mortality,  $D$ , and  $\chi \geq 0$  is the weight.

**Assumption 2.** *The felicity function  $u$ ,  $u : \mathbb{R}_+ \rightarrow \mathbb{R}$  is  $\mathcal{C}^2$  with  $u' > 0$  and  $u'' < 0$ .*

*The discount rate,  $\rho > 0$ .*

**Assumption 3.** *The loss from mortality function  $\nu(D) : \mathbb{R}_+ \rightarrow \mathbb{R}$  is a convex function with  $\nu' > 0$  and  $\nu'' \geq 0$  and  $\nu(0) = 0$ .*

**Production:** The production side of the model is a standard neoclassical growth model where households can invest in capital which is productive next period and depreciates at rate  $\delta$ .<sup>19</sup> Households own representative firms that use capital and labor as inputs.

**Assumption 4.** *The production function  $f(k, l)$ ,  $f : \mathbb{R}_+^2 \rightarrow \mathbb{R}$  is  $\mathcal{C}^2$  with*

1.  $f_k > 0, f_l > 0$ ,
2.  $f$  is concave and homogeneous of degree 1,
3. with  $f(0, \cdot) = f(\cdot, 0) = 0$ .
4.  $\lim_{k \rightarrow 0} f_k(k, \cdot) = \lim_{l \rightarrow 0} f_l(\cdot, l) = \infty$ ;  $\lim_{k \rightarrow \infty} f_k(k, \cdot) = 0$ .

---

<sup>16</sup>How much productivity is affected varies across diseases and see Goenka, Liu and Nguyen (2020) for a discussion of this. If we assume partial decrease in productivity the qualitative results are not affected.

<sup>17</sup>In Goenka and Liu (2012) we endogenize the labor-leisure choice with *SIS* disease dynamics and show that the dynamics are invariant under standard assumptions.

<sup>18</sup>This is also consistent with the welfare function used in models which are partial equilibrium, e.g. Acemoglu, et al. (2020), and Alvarez, et al. (2020). See Goenka, Liu and Nguyen (2020) for further discussion.

<sup>19</sup>Goenka and Liu (2020) have an endogenous growth model where there is human capital accumulation and households choose time to work and time for human capital accumulation. It uses *SIS* dynamics without disease related mortality.

5. The physical capital depreciates at the rate  $\delta \in (0, 1]$ .

The total labor force is

$$L = N - I \Rightarrow l = \frac{N - I}{N} = 1 - i. \quad (5)$$

The contact rate, which is key to determine the prevalence of infectious disease, is affected by health expenditure. That is, each household can control infection by investing in health expenditure  $h = H/N$ . This can be thought of preventive measures taken to prevent the transmission of the disease. This is the specification used in Goenka, Liu and Nguyen (2014) and Goenka and Liu (2020). The first paper also modeled expenditures that increase recovery rates from the disease but we abstract from this in this paper as there are no known therapies for recovery from Covid-19.<sup>20</sup> Eichenbaum, et al. (2020) interpret this loss of output,  $h$  in our context, as the cost of a lockdown which also reduces the transmission rate  $\alpha$  and thus, is also consistent with interpreting it as the cost of a lockdown which reduces transmission of the disease by forgoing output.<sup>21</sup>

**Assumption 5.** *The contact rate is a function of  $h$  - that is,  $\alpha(h)$ . We assume*

- $\alpha'(h) < 0$ ;
- $\alpha(h) \rightarrow \bar{\alpha}$  when  $h \rightarrow 0$ ;

The law of evolution of capital stock is:

$$\dot{K} = f(K, N - I) - C - H - \delta K.$$

We define for each household, physical capital per capita  $k = K/N$  and consumption per capita  $c = C/N$ . Since the two models differ in the growth rate of total population  $N$ , which is affected by death rate caused by infectious diseases, the maximization problems which is set up in terms of per household are different. Since the growth rate of total population is  $b - d - \phi_1 i - \phi_2 r$ . Then, the law of motion for physical capital can be rewritten as:

$$\dot{k} = f(k, 1 - i) - c - h - \delta k - (b - d)k + (\phi_1 i + \phi_2 r)k \quad (6)$$

We define the following variable which is the effective discount rate (see Uzawa (1968). In Das (2003) discounting depends on consumption and in Boucekkine, et al. (2018) on population similar to this paper).

$$\begin{aligned} \Delta &= \int_0^t (\rho - b + d + \phi_1 i(\tau) + \phi_2 r(\tau)) d\tau \\ \dot{\Delta} &= \rho - b + d + \phi_1 i + \phi_2 r. \end{aligned} \quad (7)$$

---

<sup>20</sup>Antivirals and anti-inflammatory drugs are now known to reduce mortality but these are inexpensive and widely available prior to the outbreak. While ventilators are used in severe cases, there is an emerging view that their use can complicate recovery and in fact cause ventilator induced lung injury (Marini and Gattinoni (2020).

<sup>21</sup>In Goenka, Liu and Nguyen (2020) we have a different way of modeling a lockdown or quarantine where a fraction of the non-infective population has to work from home with a reduced productivity.

Thus, the social planner maximizes equation (4) subject to equation (6), (1), (2) and (7) with the constraints that  $i, r$  lie in the unit interval and that health expenditure,  $h$ , is non-negative.

The maximization problem is:

$$\max \int_0^{\infty} e^{-\Delta} [u(c) - \chi\nu(\phi_1 i + \phi_2 r)] N_0 dt$$

s.t.

$$\dot{k} = f(k, 1 - i) - c - h - \delta k - (b - d)k + (\phi_1 i + \phi_2 r)k \quad (8)$$

$$\dot{i} = \alpha(h)(1 - i - r)i - \psi i - bi - \phi_1 i + i(\phi_1 i + \phi_2 r) \quad (9)$$

$$\dot{r} = \psi i - br - \phi_2 r + r(\phi_1 i + \phi_2 r) \quad (10)$$

$$\dot{\Delta} = \rho - b + d + (\phi_1 i + \phi_2 r) \quad (11)$$

$$0 \leq i \leq 1$$

$$0 \leq r \leq 1$$

$$h \geq 0.$$

The control variables are  $c, h$ , and state variables are  $k, i, r, \Delta$ . Define  $\lambda_1, \lambda_2, \lambda_3$  and  $\lambda_4$  are co-state variables for  $k, i, r$  and  $\Delta$ , respectively.  $\mu_1, \mu_2$  and  $\mu_3$  are Lagrangian multipliers for  $i \geq 0, r \geq 0$  and  $h \geq 0$ .<sup>22</sup>

The problem is one where there is endogenous discounting in a non-convex model so that the usual sufficiency conditions do not apply.

**Remark 1.** *The Hamiltonian is not jointly concave in state and control variables if the welfare function is positive, i.e. if  $u(c) - \chi\nu(\phi_1 i + \phi_2 r) > 0$ . In particular, the condition for the Hessian matrix to be semi-negative definite which require alternating signs of principal minors  $M_j (j = 1, \dots, 6)$  is not satisfied in our model.*

Let us rewrite the Hamiltonian as  $H(k, i, \Delta, r, h, c)$  then it is easy to check, the first minor  $M_1 = |H_{kk}| = \lambda_1 f_{11} < 0$ . Let  $M_2 = \begin{vmatrix} H_{kk} & H_{ki} \\ H_{ik} & H_{ii} \end{vmatrix}$  denote the second minor. We then have

$$\begin{aligned} M_3 &= \begin{vmatrix} H_{kk} & H_{ki} & H_{k\Delta} \\ H_{ik} & H_{ii} & H_{i\Delta} \\ H_{\Delta k} & H_{\Delta i} & H_{\Delta\Delta} \end{vmatrix} = \begin{vmatrix} H_{kk} & H_{ki} & 0 \\ H_{ik} & H_{ii} & H_{i\Delta} \\ 0 & H_{\Delta i} & H_{\Delta\Delta} \end{vmatrix} \\ &= H_{\Delta\Delta} M_2 + (-1)^{2+3} H_{\Delta i} \begin{vmatrix} H_{kk} & 0 \\ H_{ik} & H_{i\Delta} \end{vmatrix} \\ &= H_{\Delta\Delta} M_2 - H_{\Delta i}^2 H_{kk}. \end{aligned}$$

<sup>22</sup>We can ignore the constraints  $i \leq 1$  and  $r \leq 1$ .

Because  $H_{\Delta\Delta} = e^{-\Delta}[u(c) - \chi\nu(\phi_1i + \phi_2r)] > 0$ ,  $H_{kk} < 0$ , we have

$$M_3 = H_{\Delta\Delta}M_2 - H_{\Delta i}^2H_{kk} > 0 \text{ if } M_2 > 0.$$

So the condition for the Hessian being semi-negative definite fails.

We establish the sufficiency conditions for this model in section 5.<sup>23</sup> The equilibrium is determined by the following necessary and sufficient first order necessary and sufficient conditions and the evolution of the state variables:

$$e^{-\Delta}u'(c) = \lambda_1 \tag{12}$$

$$\lambda_1 = \lambda_2\alpha'(h)(1 - i - r)i + \mu_3 \tag{13}$$

$$\dot{\lambda}_1 = -\lambda_1[f_1(k, 1 - i) - \delta - (b - d) + \phi_1i + \phi_2r] \tag{14}$$

$$\dot{\lambda}_2 = e^{-\Delta}\chi\phi_1\nu'(\phi_1i + \phi_2r) + \lambda_1(f_2(k, 1 - i) - k\phi_1) \tag{15}$$

$$-\lambda_2(\alpha(h)(1 - i - r) - \alpha i - \psi - b - \phi_1 + 2\phi_1i + \phi_2r) - \lambda_3(\psi + \phi_1r) - \lambda_4\phi_1 - \mu_1$$

$$\dot{\lambda}_3 = e^{-\Delta}\chi\phi_2\nu'(\phi_1i + \phi_2r) - \lambda k\phi_2 + \lambda_2(\alpha(h)i - \phi_2i) \tag{16}$$

$$-\lambda_3(-b - \phi_2 + \phi_1i + 2\phi_2r) - \lambda_4\phi_2 - \mu_2$$

$$\dot{\lambda}_4 = e^{-\Delta}[u(c) - \chi\nu(\phi_1i + \phi_2r)] \tag{17}$$

$$\mu_1 \geq 0, \quad i \geq 0, \quad \mu_1i = 0 \tag{18}$$

$$\mu_2 \geq 0, \quad r \geq 0, \quad \mu_1r = 0 \tag{19}$$

$$\mu_3 \geq 0, \quad h \geq 0, \quad \mu_3h = 0, \tag{20}$$

along with equation (1), (2), (6) and (7).

The economic epidemiology models with only early mortality and only late mortality are two special cases of the model we above. When  $\phi_2 = 0$ , it is early mortality model (model I) where the infected die due to diseases, and the equilibrium conditions are given by equation (1), (2), (6), (7), (12) - (20) with  $\phi_2 = 0$ . When  $\phi_1 = 0$ , it is late mortality model (model II) where the infected die due to diseases, and the equilibrium conditions are given by equation (1), (2), (6), (7), (12) - (20) with  $\phi_1 = 0$ .

There are four ways how the disease induced mortality affects the economic equilibrium in both models. The effects are largely symmetric. That is, how disease induced mortality rate  $\phi_1$  affects the behavior of model I is similar to how disease induced mortality rate  $\phi_2$  affects the behavior of model II. First, it leads to direct death. For model I, this is captured by  $-\phi_1i$  in equation (1) and the reduction in the proportion of the infected further leads to change in shadow value for for  $i$ , which is captured by  $\lambda_2\phi_1$  in equation (15). For model II, the direct death effect is captured by  $-\phi_2r$  in equation (2) with the reduction in the proportion of the recovered further leading to a change in the shadow value for  $r$ , which is captured by  $\lambda_3\phi_2$  in equation (16). Second, the disease induced mortality leads to a reduction in the population growth rate. Since the variables in the model are per capita terms, a reduction in population

<sup>23</sup>Goenka, Liu and Nguyen (2014) prove sufficiency in a *SIS* model without disease related mortality and Goenka, Liu and Nguyen (2020) show sufficiency in a *SIS* model with disease related mortality. These are the only sufficiency conditions for optimal control in economic epidemiology models we are aware of.



growth means an increase in the value of the variables. This is what we call concentration effect. If the population growth rate decreases by the amount  $D = \phi_1 i + \phi_2 r$  due to infectious diseases, the proportion of the infected increases by  $i(\phi_1 i + \phi_2 r)$  shown in equation (1), the proportion of the recovered increases by  $r(\phi_1 i + \phi_2 r)$  shown in equation (2), and the physical capital per capita  $k$  increases by  $k(\phi_1 i + \phi_2 r)$  shown in equation (6). They further lead to change in their shadow values. The change in the shadow value for  $k$  is captured by the term  $-\lambda_1(\phi_1 i + \phi_2 r)$  in equation (14). The change in the shadow value for  $i$  is captured by the term  $-\lambda_1 k \phi_1 - \lambda_2(2\phi_1 i + \phi_2 r) - \lambda_3 \phi_1 r$  in equation (15) and the change in the shadow value for  $r$  is captured by the term  $-\lambda_1 k \phi_2 - \lambda_2 \phi_2 i - \lambda_3(\phi_1 i + 2\phi_2 r)$  in equation (16). Third, the disease induced mortality means additional loss in utility. Since the objective function incorporates the disutility from the disease related death  $\chi\nu(\phi_1 i + \phi_2 r)$ , the shadow values of  $i$  and  $r$  include additional value loss when more people die as a result of increases in  $i$  and  $r$ . This is captured by the term  $e^{-\Delta}\chi\phi_1\nu'(\phi_1 i + \phi_2 r)$  in equation (15) and  $e^{-\Delta}\chi\phi_2\nu'(\phi_1 i + \phi_2 r)$  in equation (16). Fourth, with disease induced mortality, the probability of dying is higher and people become less patient. That is, disease induced mortality leads to higher discount rates shown in 7, and to further changes its shadow value  $\lambda_4$  in equation (17).

### 3.2 Steady states

In this subsection, we look at the steady states - how epidemiological and economic variables behave in the long-run. Please see the appendix for all the proofs.

When the contact rate without any disease control is small enough such that the reproduction number  $R_0 = \frac{\bar{\alpha}}{b+\psi+\phi_1}$  is smaller than one, the infectious diseases are eradicated in the long-run. The economy is the same as the standard new-classical economy without disease prevalence.

**Proposition 2.** *There always exists a unique disease free steady state with  $i^\infty = 0$ ,  $r^\infty = 0$  and  $h^\infty = 0$ . The economic variables  $k^\infty$  and  $c^\infty$  are determined by*

$$\begin{aligned} f_1(k, 1) &= \rho + \delta \\ f(k, 1) &= c + \delta k + (b - d)k. \end{aligned}$$

*The disease free steady state is stable when  $\frac{\bar{\alpha}}{b+\psi+\phi_1} \leq 1$ , and unstable when  $\frac{\bar{\alpha}}{b+\psi+\phi_1} > 1$ .*

*Proof.* Please see the Appendix. □

The physical capital, total output and consumption in the disease free steady state are the benchmark for us to evaluate those economic variables in a disease endemic steady state later. Note that when  $\frac{\bar{\alpha}}{b+\psi+\phi_1} > 1$ , there also exists a disease free steady state, though unstable.

Next, we look at the disease endemic steady state, which exists only when  $\frac{\bar{\alpha}}{b+\psi+\phi_1} > 1$ . The key variable to determine in the disease endemic steady state is the level of health



expenditure  $h$ . We start by defining a function  $G$  - the net marginal benefit of health expenditure. The marginal benefit of health expenditure is in reducing the contact rate  $\alpha(h)$  and hence the proportion of the infected by  $\alpha'(h)(1 - i - r)i$  due to the standard incidence assumption. Since the shadow value of the infected  $i$  is  $\lambda_2$ , the value of additional one unit investment in health expenditure is  $\lambda_2\alpha'(h)(1 - i - r)i$ . The marginal cost of health expenditure is the economic resource which otherwise could have been consumed and captured by the term  $\lambda_1$ . Therefore, the net marginal benefit of health expenditure is  $\lambda_2\alpha'(h)(1 - i - r)i - \lambda_1$ . Moreover, since  $\lambda_1 = e^{-\Delta}u'(c)$ , we can redefine the net marginal benefit of health expenditure as  $e^{-\Delta} \left[ \tilde{\lambda}_2\alpha'(h)(1 - i - r)i - u'(c) \right]$  with  $\tilde{\lambda}_2 = \lambda_2/e^{-\Delta}$ .

**Definition 1.** Define the function  $G(h)$  :

$$G(h) = \tilde{\lambda}_2\alpha'(h)(1 - i - r)i - u'(c), \quad (21)$$

where

$$\left(\phi_2 + \frac{(\alpha - \phi_2)\phi_1}{\phi_1 - \alpha}\right)r^2 + (-b - \phi_2 + \frac{(\alpha - \phi_2)\psi}{\phi_1 - \alpha} + \phi_1 + \frac{\phi_1(\psi + b)}{\phi_1 - \alpha})r + \psi + \frac{\psi(\psi + b)}{\phi_1 - \alpha} \quad (22)$$

$$i = \frac{\alpha - \phi_2}{\phi_1 - \alpha}r + 1 + \frac{\psi + b}{\phi_1 - \alpha} \quad (23)$$

$$f_1(k, 1 - i) = \rho + \delta \quad (24)$$

$$f(k, 1 - i) = c + h + \delta k + (b - d)k + (\phi_1 i + \phi_2 r)k \quad (25)$$

$$\tilde{\lambda}_2 = \lambda_2/e^{-\Delta} \quad (26)$$

$$= \frac{M_1(-\rho - d - \phi_2 + \phi_2 r) - M_2(\psi + \phi_1 r)}{(\alpha(1 - i - r) - \rho - d - \psi - \phi_1 + \phi_1 i - \alpha i)(-\rho - d - \phi_2 + \phi_2 r) - (\psi + \phi_1 r)(\phi_2 i - \alpha i)}$$

$$M_1 = \chi\phi_1\nu'(\phi_1 i + \phi_2 r) + u'(c)(f_2(k, 1 - i) - k\phi_1) + \phi_1 \frac{u(c) - \chi\nu(\phi_1 i + \phi_2 r)}{\rho - b + d + \phi_1 i + \phi_2 r} \quad (27)$$

$$M_2 = \chi\phi_2\nu'(\phi_1 i + \phi_2 r) - u'(c)k\phi_2 + \phi_2 \frac{u(c) - \chi\nu(\phi_1 i + \phi_2 r)}{\rho - b + d + \phi_1 i + \phi_2 r}. \quad (28)$$

By equations (22) - (28), all variables in the model are functions of  $h$ . Thus, the function  $G$  is essentially a function of health expenditure  $h$ . In other words, once we determine the level of health expenditure in the model, we can solve for all the other variables.

**Proposition 3.** If  $\frac{\bar{\alpha}}{b + \psi + \phi_1} > 1$ , there are two scenarios

- If  $G(h)|_{h=0} \leq 0$ , there exists a disease endemic steady state with no health expenditure  $h^\infty = 0$ .
- $G(h)|_{h=0} > 0$ , there exists a disease endemic steady state with positive health expenditure, which is determined by  $G(h) = 0$ .

Given the optimal  $h^\infty$ , the steady state variables  $i^\infty, r^\infty, k^\infty$  and  $c^\infty$  are determined by equations (22) - (28).

*Proof.* Please see the Appendix. □

From equation (13), we have

$$\begin{aligned}\mu_3 &= -[\lambda_2\alpha'(h)(1-i-r)i - \lambda_1] \\ &= -e^{-\Delta}[\tilde{\lambda}_2\alpha'(h)(1-i-r)i - u'(c)] \\ &= -e^{-\Delta}G(h)\end{aligned}$$

where  $\lambda_1$  is marginal cost of health expenditure and  $\lambda_2\alpha'(h)(1-i-r)i$  is marginal benefit of health expenditure. When  $G(h)|_{h=0} \leq 0$ , that is,  $\mu_3 \geq 0$ , the marginal cost of health expenditure is larger than the marginal benefit, which implies we have a corner solution with the optimal health expenditure  $h^\infty = 0$ . When  $G(h)|_{h=0} > 0$ , that is,  $\mu_3 < 0$ , the marginal cost of health expenditure is smaller than the marginal benefit, which implies we have an interior solution ( $h^\infty > 0$ ) determined by  $G(h) = 0$ .

Given the optimal health expenditure  $h^\infty$ , we can solve for the contact rate  $\alpha^\infty$ . Once we know the contact rate, from the epidemiological part of the model - equation (22) and (23), we can derive the proportion of the infected  $i^\infty$  and the recovered  $r^\infty$ . Then, we can get the labor force  $l^\infty = 1 - i^\infty$ . Furthermore, from the economic part of the model - equation (24) and (25), we can get the optimal physical capital  $k^\infty$ , total output  $y^\infty$  and consumption  $c^\infty$ .

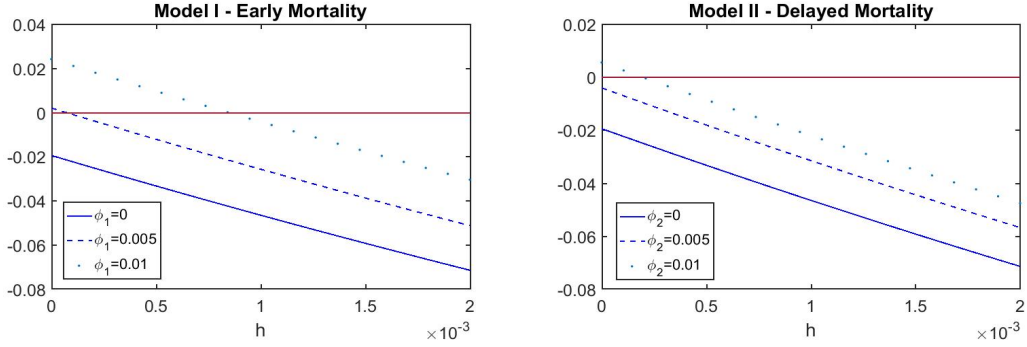
The analysis for steady states here applies for both the early mortality model and the late mortality model. In the propositions above, if  $\phi_2 = 0$ , we get steady states for the early mortality model, and if  $\phi_1 = 0$ , we get steady states for the late mortality model.

## 4 Comparing the Early and Late in Infection Mortality Models

The marriage of the economic and epidemiological models provides us a framework to understand the close link between the the economy and disease prevalence. In this paper, we focus on how the modeling strategy in epidemiology affects our understanding about the interaction between the prevalence of infectious diseases and the economy. As the model is too complex for closed form solutions, in this section, we parameterize the model and run comparative statics. We focus on examining the effects of varying disease induced mortality rates, rather than other parameters. The reason is that the effects of varying other parameters are similar in both models. Moreover, the analysis here focuses on the equilibrium steady states before and after the change as we want to capture the medium to longer term effects when investment and returns to labor and capital have adjusted.

The following functional forms and parameters are chosen in line with the literature. The production function is assumed to be Cobb Douglas:  $y = f(k, 1 - i) = Ak^\beta(1 - i)^{1-\beta}$  with  $A = 1$  and  $\beta = 0.36$ . Physical capital depreciates at the rate  $\delta = 0.05$  and discount rate  $\rho = 0.055$ . The utility function is of the CES form  $U(c) = \frac{c^{1-\sigma}}{1-\sigma}$  and we choose  $\sigma = 1$ , that

**Figure 3.** A depiction of G function in Model I and II

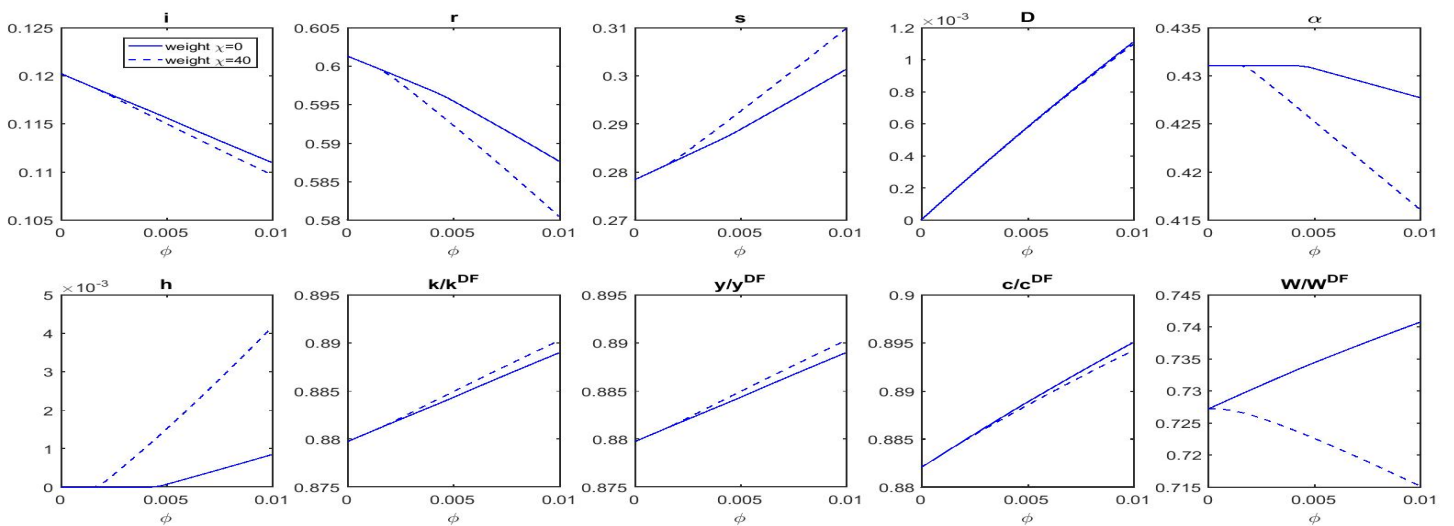


The figure depicts G function, which is the net marginal benefit of health expenditure. The left panel is G function for the early mortality model (model I), and the right panel is G function for the late mortality model (model II). In each panel, we plot G function for three different disease mortality rates - 0, 0.5% and 1%. The weight is  $\chi = 0$ .

is, the utility function is log utility. We assume the disutility from disease induced death takes the form:  $\nu(D) = D^{\nu_0}$  with  $\nu_0 = 1.2$  in line with the assumptions in the paper. In the simulation, we vary the weight we attach to this disutility  $\chi$ . We set the birth rate  $b = 2\%$  and death rate  $d = 1\%$ . The recovery rate is  $\psi = 0.1$ . We do not have enough empirical evidence in suggesting for functional form for contact rate function. Therefore, the contact rate function is chosen in line with the assumption on  $\alpha(h)$  and large enough to generate a disease endemic steady state in the simulation. We assume contact rate function is a power function:  $\alpha(h) = \epsilon_0(h + \epsilon_2)^{\epsilon_1}$  with  $\epsilon_0 = 0.2$ ,  $\epsilon_1 = -0.2$  and  $\epsilon_2 = 0.0215$ .

We start by examining the G function, shown in Figure 3. It depicts net marginal benefit of health expenditure as a function of health expenditure  $h$ . The left panel is for model I, and the right panel is for model II. In each panel, we plot G function with three different levels of disease induced mortality rate. In all three scenarios, we can see that the net marginal benefit of health expenditure decreases as health expenditure increases. This is rather intuitive and a direct consequence of the assumptions in the paper. The functional form of the contact rate function  $\alpha(h)$  implies that  $\alpha'(h) < 0$  and  $\alpha''(h) > 0$ . Thus, when health expenditure  $h$  increases, the contact rate  $\alpha(h)$  decreases, but it decreases at a diminishing rate. This suggests the marginal benefit of health expenditure decreases as  $h$  increases. At the same time, as more resource spent in controlling diseases, consumption is lower and marginal utility of consumption is higher. This suggests the marginal cost of health expenditure rises as  $h$  increases. Therefore, the net marginal benefit of health expenditure  $G(h)$  decreases as health expenditure  $h$  increases. Furthermore, when we increase the disease induced mortality rate, the net marginal benefit of health expenditure increases in both models. In the case of  $\phi_1 = 0$  in model I and  $\phi_2 = 0$  in model II, the two models coincide and the G function in both panels are exactly the same. At the point  $h = 0$ , the net marginal benefit of health expenditure is negative. It implies that the marginal benefit of health expenditure is smaller than marginal cost, and the optimal health expenditure is zero. Thus, we have an endemic steady state with corner solution. In the case of  $\phi_1 = 1\%$  in model I, at the point of  $h = 0$ ,

**Figure 4.** The simulation for the early mortality model (model I) varying disease induced mortality rate  $\phi_1$



The figure shows changes in steady state as we vary disease induced mortality rate  $\phi_1$  in the early mortality model. The top panels are the proportion of the infected  $i$ , the proportion of the recovered  $r$ , the proportion of the susceptible  $s$ , the proportion of disease related death in total population  $D$  and the contact rate  $\alpha$ . The bottom panels are health expenditure  $h$ , physical capital  $k/k^{DF}$ , total output  $y/y^{DF}$ , consumption  $c/c^{DF}$ , total welfare  $W/W^{DF}$ . The solid line is the simulation when  $\chi = 0$ , while the dash line is the simulation when the weight  $\chi = 40$ .

the net marginal benefit of health expenditure is positive, and it implies model I has an endemic steady state with interior solution. This is true in the right panel of model II with  $\phi_2 = 1\%$ . Figure 3 provides the net marginal benefit of health expenditure when we choose the weight  $\chi = 0$ . For both models, when we increase the weight  $\chi$ , the net marginal benefit health expenditure increases, due to the additional gain from the drop in the disutility of disease related death. It suggests with larger weight in disutility of disease related death, it is more likely to invest in health expenditure to control the spread of infectious disease.

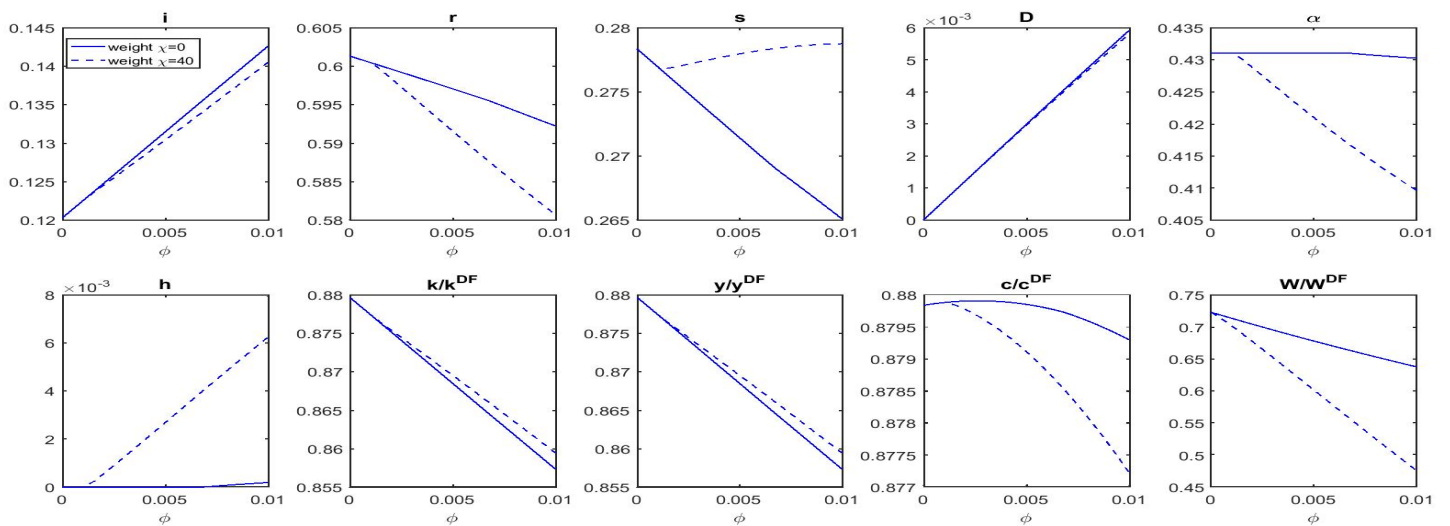
To further investigate how the steady state variables change as we change the disease induced mortality rate, we provide simulation results in Figure 4 for model I and Figure 5 for model II. In each of these figures, the solid line is the scenario when the household only derives utility from consumption with weight  $\chi = 0$ , and the dashed line is the scenario where there is additional disutility from disease related death with weight  $\chi = 40$ . When there are no disease related deaths ( $\phi_1 = \phi_2 = 0$ ), both models are the same, which predict zero health expenditure as the net marginal benefit of health expenditure is negative, shown in Figure 3. Therefore, in both figures, all variables have the same starting point. For instance, in model I when  $\phi_1 = 0$  the proportion of the infected is 12% of the population, and in model II when  $\phi_2 = 0$  the proportion of the infected is 12% as well.

Now lets focus on how the epidemiological and economic variables vary as we change the disease induced mortality rates in both models. First, when the disease induced mortality rate rises, households are more likely to invest in health expenditure in controlling diseases or increase heath expenditure. This is reflected in the G function - Figure 3. When mortality rate rises, the net marginal benefit of controlling diseases increases. This is true for both models.

Second, for the epidemiological variables, their variations largely mirror those in the pure epidemiological models. This mirroring effect is not immediately obvious, as in the economic epidemiological model the health expenditure and hence the contact rate change as the disease mortality rate rises, while in the pure epidemiological model (Figure 2) the contact rate is a constant. In model I, an increase in mortality of infectives decreases the proportion of infectives in the population. This is the self-limiting effect of mortality on pandemics. In contrast, in model II, when disease mortality rate  $\phi_2$  increases, the proportion of the infected increases though the health expenditure increases and the contact rate decreases.

Third, the evolution of economic variables are drastically different between the two models as the disease induced mortality rates varies. To explore the change in economic variables - physical capital  $k$ , output  $y$  and consumption  $c$ , we plot them as a proportion of the disease free steady state, which implies how much worse off of the disease endemic steady state is as compared to the disease free steady state. Note that, for both models the disease free steady states are exactly the same. The prevalence of infectious diseases affect the economy mainly because of three reasons: (i) the infected can not participate in labor force, (ii) that investing in health in controlling diseases can squeeze the economic resources and (iii) the disease related mortality changes population size, which has two consequences: first, the per capita amount of resources change and second, the change in population affects the discount rate. As we can see, for model I, as disease death rate  $\phi_1$  increases, physical capital stock, total output, consumption all increase. This seems to be counter intuitive. In fact, this is like a gift of dying effect. The individuals who die due to infectious diseases are those who are infected, infectious and can not work, but consume. So from a purely economic point of

**Figure 5.** The simulation for the late mortality model (model II) varying disease induced mortality rate  $\phi_2$



The figure shows changes in steady state as we vary disease induced mortality rate  $\phi_2$  in the early mortality model. The top panels are the proportion of the infected  $i$ , the proportion of the recovered  $r$ , the proportion of the susceptible  $s$ , the proportion of disease related death in total population  $D$  and the contact rate  $\alpha$ . The bottom panels are health expenditure  $h$ , physical capital  $k/k^{DF}$ , total output  $y/y^{DF}$ , consumption  $c/c^{DF}$ , total welfare  $W/W^{DF}$ . The solid line is the simulation when  $\chi = 0$ , while the dash line is the simulation when the weight  $\chi = 40$ .

view (when the objective is only about maximizing consumption per capita), the society is better off when disease mortality rate is high. This is because with a decreasing proportion of the infected, the labor force  $1 - i$  increases, which implies higher marginal product of capital, and higher physical capital. Since both labor force and physical capital are higher, the total output and consumption also increase as mortality rate rises. Moreover, the total welfare in fact rises as the disease mortality rate increases. However, this gift of dying effect is absent in model II. In model II, when disease death rate  $\phi_2$  increases, physical capital stock, total output, consumption and total welfare all decrease. This is because in model II, increasing mortality does not limit the spread of infections and the pandemic gets worse rather than better. This effect not only directly affects per capita availability of resources but also increases discounting. Thus, this leads to lower investment in capital, and lower total output and consumption. The total welfare also decrease as mortality rate rises.

We now look at the case where there is a welfare loss from mortality,  $\chi > 0$ , shown as the dashed lines in Figure 4 and 5. With positive weight on disutility, the evolution of epidemiological and economics variables largely remain the same. One difference is that when we take into account the additional utility loss from the disease related death, in model I the total welfare decreases as  $\phi_1$  increase. This is because with positive weight on disease related death, the rise in disutility from death outweighs the rise in utility from consumption, which leads the total welfare to fall. However, how physical capital, output and consumption change does not depend on the weight attached to the disutility.

To sum up, differences in epidemiological modeling matters not only for disease evolution but also for the equilibrium economic outcomes. As the difference in modeling mortality affects evolution of the infective population, the optimal health expenditure is affected. The extent of the effect depends on how much weight is given to the welfare loss due to mortality. This affects other economic outcomes and the welfare in equilibrium. Thus, there is a non-trivial interaction between economic and epidemiological choices in understanding the effects of pandemics on welfare and economic outcomes.

## 5 Sufficient conditions

In this section we study the sufficiency of the first order conditions with disease related mortality. It is well known in the literature that with *SIS* or *SIR* dynamics the constraints are not convex and it is unclear if either the Arrow or the Mangasarian sufficiency conditions will be satisfied (Gersovitz and Hammer (2003)). Goenka, Liu and Nguyen (2014) provided a sufficiency result in a neo-classical framework, such as in the current paper, with *SIS* dynamics but no disease mortality<sup>24</sup> and Goenka, Liu and Nguyen (2020) show sufficiency in the *SIS* with disease related mortality. We give the first sufficiency result for the *SIR* model. The result is with disease related mortality and the case without disease related mortality is a special case. We give the result for the general model so that it covers both the case of early and late mortality. The problem is non-trivial because including disease related mortality effectively makes the effective discount rate,  $\Delta$ , endogenous. The Hamiltonian is

---

<sup>24</sup>This paper also included the additional state variable health capital which can reduce contact rate and increase recovery rate.

non-concave so in this situation the Arrow and Mangasarian conditions do not apply (see below) as well as the conditions endogenous population models that are convex (for example Boucekkine, et al. (2018)). While the proof follows the strategy of the *SIS* disease related model, there are now two state variables for disease evolution,  $i, r$  as opposed to only one for the *SIS* model. The control variable  $h$  is also different from the control variable for quarantine in the same paper.

We directly show the inequality of local optimality of the Hamiltonian along *any* interior path that satisfies the first order necessary and transversality conditions. This is done by adapting the method of Leitmann and Stalford (1971). As a corollary, the disease endemic steady state will be locally optimal. Optimality of the disease free steady state is not in question as it is the neoclassical steady state.

Denote the interior state variables  $\mathbf{x}_t^* = (k_t^*, i_t^*, r_t^*, \Delta_t^*)$  where  $\mathbf{x}_0^* = (k_0^*, i_0^*, r_0^*, \Delta_0^*)$ , the control variables  $\mathbf{z}_t^* = (c_t^*, h_t^*)$  and co-state variables  $\lambda_t = (\lambda_{1,t}, \lambda_{2,t}, \lambda_{3,t}, \lambda_{4,t})$ .

The Hamiltonian becomes

$$\begin{aligned} H(\mathbf{x}_t, \mathbf{z}_t, \lambda_t) &= e^{-\Delta} \{u(c) - \chi\nu(\phi_1 i + \phi_2 r)\} + \lambda_1 \{f(k, (1-i) - c - h - \delta k - (b-d - \phi_1 i - \phi_2 r)k) \\ &\quad + \lambda_2 \{\alpha(h)(1-i-r)i - \psi i - bi - \phi_1 i + i(\phi_1 i + \phi_2 r)\} \\ &\quad + \lambda_3 \{\psi i - br - \phi_2 r + r(\phi_1 i + \phi_2 r)\} + \lambda_4 \{\rho - b + d + (\phi_1 i + \phi_2 r)\} \\ &= e^{-\Delta} \{u(c) - \chi\nu(\phi_1 i + \phi_2 r)\} + \langle \lambda_t, \dot{\mathbf{x}}_t \rangle \end{aligned}$$

where  $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_1^n x_j y_j$  is the dot product of two vectors  $\mathbf{x} = (x_1, \dots, x_n), \mathbf{y} = (y_1, \dots, y_n)$ .

The first-order necessary conditions are satisfied at  $(\mathbf{x}_t^*, \mathbf{z}_t^*)$  :

$$e^{-\Delta} u'(c) = \lambda_1 \tag{29}$$

$$\lambda_1 = \lambda_2 \alpha'(h)(1-i-r)i \tag{30}$$

$$\dot{\lambda}_1 = -\lambda_1 [f_1(k, 1-i) - \delta - (b-d - \phi_1 i - \phi_2 r)] \tag{31}$$

$$\begin{aligned} \dot{\lambda}_2 &= e^{-\Delta} \chi \phi_1 \nu'(\phi_1 i + \phi_2 r) + \lambda_1 (f_2(k, 1-i) - k\phi_1) \\ &\quad - \lambda_2 [\alpha(h)(1-i-r) - \alpha i - \psi - b - \phi_1 + 2\phi_1 i + \phi_2 r] - \lambda_3 (\psi + \phi_1 r) - \lambda_4 \phi_1 \end{aligned} \tag{32}$$

$$\begin{aligned} \dot{\lambda}_3 &= e^{-\Delta} \chi \phi_2 \nu'(\phi_1 i + \phi_2 r) - \lambda_1 k\phi_2 + \lambda_2 (\alpha i - \phi_2 i) \\ &\quad - \lambda_3 (-b - \phi_2 + \phi_1 i + 2\phi_2 r) - \lambda_4 \phi_2 \end{aligned} \tag{33}$$

$$\dot{\lambda}_4 = e^{-\Delta} [u(c) - \chi\nu(\phi_1 i + \phi_2 r)] \tag{34}$$

The standard transversality conditions are

$$\lim_{t \rightarrow \infty} \lambda_{j,t} x_{j,t}^* = 0, j = 1, \dots, 3. \tag{35}$$

It is standard that  $0 \leq k_t \leq \max\{k_0, \hat{k}\}$  where  $\hat{k}$  is the maximum sustainable capital



stock<sup>25</sup> Then is  $c_t$  is bounded by a constant<sup>26</sup>,  $c_t \leq A$ , and hence

$$u(c) - \chi\nu(\phi_1 i + \phi_2 r) \leq u(A) + \chi\nu(\phi_1 + \phi_2) < +\infty. \quad (36)$$

Similarly,  $h_t$  is also bounded by  $A$ .

The proof proceeds via three Lemmas.

**Lemma 1.** *We have*

$$\lim_{t \rightarrow \infty} \lambda_{4,t}(\Delta_t - \Delta_t^*) = 0.$$

*Proof.* Please see the Appendix. □

**Lemma 2.** *When  $\frac{\alpha}{b+\psi+\phi_1} > 1$  we have*

$$\begin{aligned} i) \quad & \lim_{t \rightarrow \infty} \lambda_1(k^* - k) \leq 0. \\ ii) \quad & \lim_{t \rightarrow \infty} \lambda_2(i^* - i) = 0. \\ iii) \quad & \lim_{t \rightarrow \infty} \lambda_3(r^* - r) = 0. \end{aligned}$$

*Proof.* Please see the Appendix. □

In the *SIS* model there is only one state variable,  $i$ , describing the disease dynamics and we can show the appropriate transversality condition for it directly (see Goenka, Liu and Nguyen (2020, Lemma 2)). In the *SIR* model there are two variables  $i, r$  thus we need the additional parametric restriction that ensures an endemic steady state in order to derive the limiting properties of the co-state variables.

We will adapt the method developed by Leitmann and Stalford (1971)<sup>27</sup> for a sufficiency condition to our (non-convex) infinite-horizon optimal control problem for the endogenous

---

<sup>25</sup>Definition of maximal capital stock is  $\hat{k} \in (0, \infty)$  such that  $f(k, l) > k$  for all  $k \in (0, \hat{k})$  and  $f(k, l) < k$  for all  $k > \hat{k}$ . It implies  $k \leq \max\{k_0, \hat{k}\} := \bar{k}$ .

<sup>26</sup>If investment is irreversible, then  $c_t \leq f(k_t, l_t) \leq f(\hat{k}, 1) := A$ . Otherwise, as in Goenka, Liu and Nguyen (2014), we can assume that there exists  $\kappa \geq 0, \kappa \neq \infty$  such that  $-\kappa \leq \dot{k}/k$ . 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)). Let us define the net investment  $\iota = \dot{k} + (\delta + b - d)k = f(k, l) - c - m$ , it 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 it holds with  $\kappa = \delta + b - d$ . Therefore, assuming non-negative investment is stronger in the sense that  $\kappa$  can take any value except for infinity. And we have  $c_t \leq f(\bar{k}, 1) + \kappa \bar{k} := A$ .

<sup>27</sup>We re-state the Leitmann and Stalford result. (Leitmann-Stalford, 1971): Consider the problem:

$$\max \int_0^{\infty} g_0(x(t), z(t))$$

subject to

$$\dot{x}(t) = g(x(t), z(t)), x(0) = x_0, z(t) \in Z$$

discounting problem. The Leitmann-Stalford result allows for potential non-convexities but not endogenous discounting. This condition is weaker than standard Arrow-Mangasarian sufficient conditions (see Theorem V, Peterson and Zalkind (1978), page 595).

Define the augmented Hamiltonian  $\bar{H}(\mathbf{x}_t, \mathbf{z}_t, \lambda_t) = H(\mathbf{x}_t, \mathbf{z}_t, \lambda_t) + \langle \dot{\lambda}_t, \mathbf{x}_t \rangle$  and  $M(\mathbf{x}_t, \lambda_t) = \max_{\mathbf{z}_t} \bar{H}(\mathbf{x}_t, \mathbf{z}_t, \lambda_t)$  as the augmented maximized Hamiltonian.

We need the following Lemma.

**Lemma 3.** *We have  $\bar{H}(\mathbf{x}_t^*, \mathbf{z}_t^*, \lambda_t) \geq \bar{H}(\mathbf{x}_t^*, \mathbf{z}_t, \lambda_t)$  for all  $\mathbf{z}_t$ . In other word, given  $\mathbf{x}_t^*$  then  $\mathbf{z}_t^* = \arg \max \bar{H}(\mathbf{x}_t^*, \mathbf{z}_t, \lambda_t)$  and thus  $M(\mathbf{x}_t^*, \lambda_t) = \bar{H}(\mathbf{x}_t^*, \mathbf{z}_t^*, \lambda_t)$ .*

*Proof.* Please see the Appendix. □

In line with Leitmann and Stalford (1971), we will use the following assumption.

**Assumption 6.** *Assume that*

$$\bar{H}(\mathbf{x}_t^*, \mathbf{z}_t^*, \lambda_t) \geq \bar{H}(\mathbf{x}_t, \mathbf{z}_t, \lambda_t) \tag{37}$$

**Remark 2.** *Assumption 6 is weaker than assumption on the concavity of maximized Hamiltonian  $M(\mathbf{x}_t, \lambda_t)$  in  $\mathbf{x}_t$  as in Arrow's sufficiency condition. Indeed, assuming  $M(\mathbf{x}_t, \lambda_t)$  is concave in  $\mathbf{x}_t$ . Since  $M(\mathbf{x}_t, \lambda_t) \geq \bar{H}(\mathbf{x}_t, \mathbf{z}_t, \lambda_t)$  and by Lemma 3  $M(\mathbf{x}_t^*, \lambda_t) = \bar{H}(\mathbf{x}_t^*, \mathbf{z}_t^*, \lambda_t)$  and*

$$\begin{aligned} \bar{H}_{x_j,t}(\mathbf{x}_t^*, \mathbf{z}_t^*, \lambda_t) &= H_{x_j,t}(\mathbf{x}_t^*, \mathbf{z}_t^*, \lambda_t) + \dot{\lambda}_{j,t} \\ &= -\dot{\lambda}_{j,t} + \dot{\lambda}_{j,t} = 0 \end{aligned}$$

---

Define the Hamiltonian

$$H(x, z, \lambda) = g_0(x(t), z(t)) + \langle \lambda, g(x(t), z(t)) \rangle$$

where  $\langle \lambda, g(x(t), z(t)) \rangle$  is the inner product in  $\mathbb{R}^n$ . Let  $z^* \in Z$ , let  $x^* = x(z^*)$  be the corresponding trajectory, and let  $\lambda : [0, \infty) \rightarrow \mathbb{R}^n$  be absolutely continuous. Let following conditions be fulfilled for every  $z \in Z$  and  $x = x(z)$

$$i) \int_0^\infty e^{-\rho t} [H(x^*(t), z^*(t), \lambda(t)) - H(x(t), z(t), \lambda(t)) + \langle \dot{\lambda}, x^*(t) - x(t) \rangle] dt \geq 0,$$

and

$$ii) \lim_{t \rightarrow \infty} \langle \lambda, x^*(t) - x(t) \rangle \leq 0.$$

Then  $(x^*, z^*)$  is an optimal solution.

we get

$$\begin{aligned}
\bar{H}(\mathbf{x}_t^*, \mathbf{z}_t^*, \lambda_t) - \bar{H}(\mathbf{x}_t, \mathbf{z}_t, \lambda_t) &\geq M(\mathbf{x}_t^*, \lambda_t) - M(\mathbf{x}_t, \lambda_t) \\
&\geq \langle M_x(\mathbf{x}_t^*, \lambda_t), \mathbf{x}_t^* - \mathbf{x}_t \rangle \\
&= \langle \bar{H}_x(\mathbf{x}_t^*, \mathbf{z}_t^*, \lambda_t), \mathbf{x}_t^* - \mathbf{x}_t \rangle \\
&= 0.
\end{aligned}$$

Also, if the Hamiltonian is jointly concave in state and control variables as in the Mangasarian sufficient condition, we easily get (37) by the properties of a concave function and the FOCs (29)-(34):

$$\bar{H}(\mathbf{x}_t^*, \mathbf{z}_t^*, \lambda_t) - \bar{H}(\mathbf{x}_t, \mathbf{z}_t, \lambda_t) \geq \langle \bar{H}_x(\mathbf{x}_t^*, \mathbf{z}_t^*, \lambda_t), \mathbf{x}_t^* - \mathbf{x}_t \rangle + \langle \bar{H}_z(\mathbf{x}_t^*, \mathbf{z}_t^*, \lambda_t), \mathbf{z}_t^* - \mathbf{z}_t \rangle = 0.$$

However, in our model, the Hamiltonian is not jointly concave if the welfare function is positive, i.e. if  $u(c) - \chi\nu(\phi_1 i + \phi_2 r) > 0$ . (see Remark 1 above).

**Remark 3.** SIR model without disease related mortality is a special case and with exogenous discounting Assumption 6 is not needed.

We are now ready to prove the sufficient condition.

**Proposition 4.** Consider the maximization problem (4) and suppose that an interior continuous  $(\mathbf{x}_t^*, \mathbf{z}_t^*)$  and associated costate variables  $\lambda_t$  exist and satisfy (29)-(34). Then under Assumptions 1-6,  $(\mathbf{x}_t^*, \mathbf{z}_t^*)$  is a locally optimal solution.

*Proof.* The results of Lemmas 1-2 yield

$$\lim_{t \rightarrow \infty} \langle \lambda_t, \mathbf{x}_t^* - \mathbf{x}_t \rangle \leq 0. \quad (38)$$

From Assumption A6 we get

$$\int_0^\infty \{H(\mathbf{x}_t^*, \mathbf{z}_t^*, \lambda_t) - H(\mathbf{x}_t, \mathbf{z}_t, \lambda_t)\} + \langle \dot{\lambda}_t, \mathbf{x}_t^* - \mathbf{x}_t \rangle dt \geq 0.$$

It implies

$$\int_0^\infty e^{-\Delta^*} [u(c^*) - \chi\nu((\phi_1 i^* + \phi_2 r^*))] dt - \int_0^\infty e^{-\Delta} [u(c) - \chi\nu((\phi_1 i + \phi_2 r))] dt \geq - \lim_{t \rightarrow \infty} \langle \lambda_t, \mathbf{x}_t^* - \mathbf{x}_t \rangle.$$

Therefore, it follows from (38) that

$$\int_0^\infty e^{-\Delta^*} [u(c^*) - \chi\nu((\phi_1 i^* + \phi_2 r^*))] dt - \int_0^\infty e^{-\Delta} [u(c) - \chi\nu((\phi_1 i + \phi_2 r))] dt \geq 0$$

and we get the sufficient condition.  $\square$

As the endemic steady state satisfies the necessary conditions, we have shown that it is indeed optimal.

**Remark 4.** *Using properties of the function  $G$  the proposition can be amended to cover the case where there are no health expenditures in the steady state. Similarly, the optimality of the disease free steady state can be established. These are omitted for brevity.*

## 5.1 Discussion

We are using the FOCs and the standard transversality conditions

$$\lim_{t \rightarrow \infty} \lambda_{j,t} x_{j,t}^* = 0, j = 1, \dots, 3. \quad (39)$$

to provide a direct proof for sufficient condition. Note that this condition holds only at the optimal path  $x_{j,t}^*$ . For any admissible path  $x_{j,t}$  it may not be satisfied. Moreover,  $\lambda_t$  is only identified by the FOCs at optimal solutions  $(\mathbf{x}_t^*, \mathbf{z}_t^*)$ .

The literature also used a transversality condition where along the optimal paths

$$\lim_{t \rightarrow \infty} H(\mathbf{x}_t^*, \mathbf{z}_t^*, \lambda_t) = 0. \quad (40)$$

The transversality condition (40) is taken from Michel (1982) for a constant discount rate. In general, these two conditions are not equivalent. Six and Wirl (2015) in a pollution model with endogenous discounting model using the convergence of the state variable to a steady state show that if (35) holds then (40) also holds. We now also show a similar result but for our non-convex model based on the usual transversality condition and additional assumptions on function  $\alpha(h)$ .

**Lemma 4.** *If  $\alpha'(0) \neq -\infty$  and  $\alpha'' > 0$  then the usual transversality condition (35) implies the (40) transversality condition.*

*Proof.* Please see the Appendix. □

Note we only assume that  $\alpha'(0) < 0$  for existence of steady state (Assumption 5). The assumption that  $\alpha'(0) \neq -\infty$  is ruling out an Inada type condition on the  $\alpha$  function. This is consistent with our results on the existence of a steady state with zero health expenditure. We did not need to impose a condition on the second derivative of  $\alpha$  for the earlier results.

Note that the transversality condition (40) can also imply the usual transversality condition (35) but we need more assumptions. For example, Aseev and Kryazhimskiy (2004) show that (40) implies (35) if additional assumptions on the constraint of state variables imposed. (see Corollary 4, page 1111).

For the sufficiency, we assume only (35) holds. However, since our model is non-convex with endogenous discounting, this condition is not enough for a sufficiency as the framework

of the earlier results do not hold. Using the special structure of the autonomous problem, we provide a direct proof of sufficiency by proving the transversality condition for state variables for any admissible  $x_t$ ,

$$\lim_{t \rightarrow \infty} \lambda_{j,t}(x_{j,t}^* - x_{j,t}) \leq 0. \quad (41)$$

These kind of transversality conditions were assumed directly in Cartigny and Michel (2003), Acemoglu (2009) (Theorem 7.11, page 246) for a sufficiency condition but for convex problems and standard discounting. This condition is difficult to check because the admissible path  $x_{j,t}$  does not necessarily satisfy the FOCs while the co-state  $\lambda_{j,t}$  is only determined at the optimal path  $x_{j,t}^*$ . We do not get any information for  $x_{j,t}$  from two standard transversality conditions (35) and (40). However, if  $x_{j,t}$  is bounded, then the condition  $\lim_{t \rightarrow \infty} \lambda_{j,t} = 0$  implies (41). If  $\lambda_{j,t} \geq 0$  and  $x_{j,t} \geq 0$  then (35) implies (41). Thus, Acemoglu (2009) (Theorem 7.14) makes this assumption as  $\lim_{t \rightarrow \infty} \lambda_{j,t} x_{j,t} \geq 0$ . In our model, the co-state variable associated with the infective is negative so this inequality is only satisfied as a zero identity which will be proven in our model.

Finally, it is crucial when we check the maximality of the Hamiltonian we can decompose it into two parts: the first just relies on the separability of control and state variables and the concavity in control variables of the objective function, 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 state variables and the Hamiltonian is also non-concave. As indicated, we show this term converges to a negative value, and we are able to obtain sufficiency of the first order conditions.

## 6 Conclusion

Due to the ongoing Covid-19 pandemic there is an increasing interest in economic epidemiology which study the interaction of infectious diseases using compartmental disease modeling and economic outcomes. Many papers use the Kermack-McKendrick model with mass action to model the epidemiology. However, this model due to the linear effect on infections of population size may not be the most suitable to look at medium to long run effects as population size change due to birth and deaths - both due to the disease and other causes. The standard incidence model does not have this effect and is the model of choice by epidemiologists. We investigated two models: where there are early deaths so that deaths of infectives which shortens the duration of the epidemics and late deaths when the individuals who succumb to the disease are not circulating in the population - either because they are hospitalized or there is pre-mature mortality of the recovered population. It seems that in Covid-19 an increasing part of the mortality is of the later type. These are early days of the pandemic and long run effects of the disease on pre-mature mortality are not fully understood. We studied the optimal response which can be either interpreted as preventive health expenditures or self-isolation. In a fully general equilibrium neoclassical growth model, the optimal response and equilibrium outcomes are sensitive to the modeling choice as they have very different implications for disease evolution. This generates different equilibrium effects even

when mortality is the same in both models. Thus, economists should pay close attention to what choices are made for modeling infectious diseases.

In economic epidemiology models, the interest is in optimal choices. The models with *SIR* dynamics are non-convex and if there is disease related mortality, discounting is also endogenous. Thus, the usual conditions for sufficiency of first order conditions do not apply. We present the first results on sufficiency for this model. The *SIR* model without disease related mortality is a special case. The results differ from the *SIS* model as there are now two state variables,  $s, i$  to describe disease dynamics rather than just one.

## References

- [1] Acemoglu, D. (2009) *Introduction to Modern Economic Growth*, Princeton University Press.
- [2] Acemoglu, D., Chernozhukov, V., Werning, I., and Whinston, M.D. (2020) Optimal targeted lockdowns in a multi-group SIR model, *NBER Working Paper* No. 27102.
- [3] Alvarez, F.E., Argente, D., and Lippi, F. (2020) A simple planning problem for Covid-19 lockdown, *NBER Working Paper* No. 26981.
- [4] Anderson, R.M. and May, R.M. (1991) *Infectious diseases of humans: Dynamics and control*, Oxford: Oxford University Press.
- [5] Aseev, S and Kryazhimskiy, A. (2004) The Pontryagin maximum principle and transversality conditions for a class of optimal control problems with infinite time horizons. *SIAM Journal on Control and Optimization*, 43(3), 1094–1119.
- [6] Bernheim, A. et al. (2020) Chest CT findings in Coronavirus disease-19 (Covid-19): Relationship to duration of infection, *Radiology* 295, 695-691.
- [7] Boucekkine, R. and Laffargue, J.-P. (2010) The distributional consequences of epidemics, *Journal of Economic Dynamics and Control*, 34(2) 231-245.
- [8] Boucekkine, R., Martínez, B., and Ruiz-Tamarit, J.R. (2018) Optimal population growth as an endogenous discounting problem: The Ramsey case, In: Feichtinger G., Kovacevic R., Tragler G. (eds) *Control Systems and Mathematical Methods in Economics. Lecture Notes in Economics and Mathematical Systems*, vol 687. Springer.
- [9] Brauer, F. (2008) Compartmental models in epidemiology, in Brauer, F., van den Driessche, P. and Wu, J. (Eds) *Mathematical epidemiology* Lecture notes in Mathematics 1945, Berlin: Springer-Verlag.
- [10] Busenberg, S. and van den Driessche, P. (1990) Analysis of a disease transmission model in a population with varying size, *Journal of Mathematical Biology* 28, 257-290.
- [11] Carter, B., et al. (2020) Nosocomial COVID-19 infection: examining the risk of mortality. The COPE-Nosocomial Study (COVID in Older people), *Journal of Hospital Infection* <https://doi.org/10.1016/j.jhin.2020.07.013>.

- [12] Cartigny, P. and Michel, P. (2003) On a sufficient transversality condition for infinite horizon optimal control problems, *Automatica* 39, 1007-1010.
- [13] Chakraborty, S., Papageorgiou, C., and Perez-Sebastian, F., (2010) “Diseases, infection dynamics and development” *Journal of Monetary Economics*, 859-872.
- [14] 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.
- [15] Dan, J., et al. (2021) Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection, *Science*, DOI: 10.1126/science.abf4063.
- [16] Das, M. (2003) Optimal growth with decreasing marginal impatience, *Journal of Economic Dynamics and Control*, 27, 1881-1898.
- [17] Ferguson, N., et al. (2020) Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand, Imperial College Covid-19 Response Team, DOI: <https://doi.org/10/25561/77482>.
- [18] Eichenbaum, M.S., Rebelo, S., and Trabandt, M. (2020) The macroeconomics of epidemics, *NBER Working Paper* No. 26882.
- [19] Funk, J.E. and Gilbert, E.G. (1970) Some sufficient conditions for optimality in control problems with state space constraints, *SIAM Journal on Control*, 8(4), 489-504.
- [20] Geoffard, P-Y. and Philipson, T. (1996) Rational Epidemics and Their Public Control, *International Economic Review*, 37(3): 603-24.
- [21] Gersovitz, M. and Hammer, J.S., (2004) The economical control of infectious diseases, *Economic Journal*, 1-27.
- [22] Goenka, A., Liu, L. (2013) Infectious diseases and endogenous fluctuations, *Economic Theory*, 125-149.
- [23] Goenka, A., Liu, L. (2020) Human capital, infectious diseases and economic growth, *Economic Theory*, 70, 1-47.
- [24] Goenka, A., Liu, L., Nguyen, M.H. (2014) Infectious diseases and economic growth, *Journal of Mathematical Economics* 50, 34-53.
- [25] Goenka, A., Liu, L., Nguyen, M.H. (2020) Modeling optimal quarantines under infectious disease related mortality, TSE Working Paper, No. 20-1136, August 2020.
- [26] Hellwig, C. and the Toulouse Macroeconomics Group (2020) The hammer and the dance: Equilibrium and optimal policy during a pandemic crisis, TSE Working Paper, no. 20-1099, 2020.
- [27] Hethcote, H.W. (2000) The mathematics of infectious disease, *SIAM Review* 42(4), 599-653.

- [28] Hethcote, H.W. (2008) “The basic epidemiology models, Epidemiology models with variable population size, and Age-structured epidemiology models, in Ma, S. and Xia, Y. (Eds.) *Mathematical Understanding of Infectious Disease Dynamics*, Singapore: World Scientific.
- [29] Høger, S., Lykens, K., Beavers, S. F., Katz, D., and Miller, T. L. (2014) Longevity loss among cured tuberculosis patients and the potential value of prevention, *The International Journal of Tuberculosis and Lung Disease* 18(11) 1347-1352.
- [30] Hosseiny, M., et al. (2020) Radiology perspective of coronavirus disease 2019 (Covid-19): Lessons from Severe Acute Respiratory Syndrome and Middle East Respiratory Syndrome, *American Journal of Roentgenology*, 214, 1078-1082.
- [31] Ibarra, F.J., et al. (2020) Rapid decay of Anti-SARS-Cov-2 antibodies in persons with mild Covid-19, *New England Journal of Medicine*, letter, DOI: 10.1056/NEJMc2025179.
- [32] Keeling, M.J. and Rohani, P. (2008) *Modeling infectious diseases in humans and animals*, Princeton: Princeton University Press.
- [33] Leitmann, G. and Stalford, H. (1971) A sufficiency condition for optimal control, *Journal of Optimization Theory and Applications*, 8(3), 169-174.
- [34] Li, Y. and Xia, L. (2020) Coronavirus disease 2019 (Covid-19): Role of chest CT in diagnosis and management, *American Journal of Roentgenology* 214: 1280-1286.
- [35] Long, Q., Tang, X., Shi, Q. et al. (2020) Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections, *Nature Medicine* 26, 1200-1204. <https://doi.org/10.1038/s41>.
- [36] Marcus J.L., Leyden, W.A., Alexeeff S.E., et al (2020) Comparison of overall and comorbidity-free life expectancy between insured adults with and without HIV Infection, 2000-2016, *Journal of American Medical Association Network Open* 3(6):e207954. doi:10.1001/jamanetworkopen.2020.7954
- [37] Marini, J.J. and Gattinoni, L. (2020) Management of COVID-19 respiratory distress, *Journal of American Medical Association*, 2323(22), 2329-2330. doi:10.1001/jama.2020.6825
- [38] Mena-Lorca, J. and Hethcote, H.W. (1992) Dynamic models of infectious diseases as regulators of population sizes, *Journal of Mathematical Biology* 30: 693-716.
- [39] Michel, P. (1982) On the transversality condition in infinite horizon optimal problems, *Econometrica* 50(4), 975-85.
- [40] Pèrez-Bermejo, J.A., et al. (2020) SARS-CoV-2 infection of human iPSC-derived cardiac cells predicts novel cytopathic features in hearts of COVID-19 patients, bioRxiv preprint, <https://www.biorxiv.org/content/10.1101/2020.08.25.265561v2>.



- [41] Peterson, D.W. and Zalkind, J.H. (1978) A review of direct sufficient conditions in optimal control, *International Journal of Control* 28(4), 589-610.
- [42] Sekine, et al. (2020) Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19, *bioRxiv* 2020.06.29.174888; doi: <https://doi.org/10.1101/2020.06.29.174888>
- [43] Six, M. and Wirl, F. (2015), Optimal pollution management when discount rates are endogenous, *Resource and Energy Economics*, 53–70.
- [44] Toxvaerd, F. (2019) Rational disinhibition and externalities in prevention, *International Economic Review* 60(4), 1737-1755.
- [45] Uzawa, H. (1968) Time preference, the consumption function and optimum asset holdings, in S. Wolfe (Ed.) *Value, Capital, and Growth: Essays in honour of Sir John Hicks*, Chicago: Aldine.
- [46] Wajnberg, A., et al. (2020) Humoral immune response and prolonged PCR positivity in a cohort of 1343 SARS-CoV 2 patients in the New York City region, *medRxiv* 2020.04.30.20085613; doi: <https://doi.org/10.1101/2020.04.30.20085613>

## 7 Appendix

### 7.1 Proof of Proposition 2: Disease free steady state

From  $\dot{i} = 0$  and  $\dot{r} = 0$ , we have one disease free steady state with  $i^\infty = 0$  and  $r^\infty = 0$ . From equation (13), we have

$$\mu_3 = \lambda_1 - \lambda_2 \alpha'(h)(1 - i - r)i = \lambda_1 > 0$$

Therefore,  $\mu_3$  is strictly positive and implies  $h^\infty = 0$ . From equation (14), we have

$$\frac{\dot{\lambda}_1}{\lambda_1} = -[f_1(k, 1) - \delta - b + d].$$

Moreover, from equation (12), we have

$$\frac{\dot{\lambda}_1}{\lambda_1} = -(\rho - b + d) + \frac{u''(c)}{u'(c)} \dot{c}.$$

Since the economy is bounded, all economic variables including  $k$ ,  $c$  and  $l$  are constant in the steady state. That is,  $\dot{c} = 0$  in the steady state. Thus, combining the above two equations, we have

$$f_1(k, 1) = \rho + \delta,$$

from which we can solve for  $k^\infty$ .

Therefore, there always exists a unique disease free steady state with  $i^\infty = 0$ ,  $r^\infty = 0$ ,  $s^\infty = 1$  and  $h^\infty = 0$ . The economic variables  $k^\infty$  and  $c^\infty$  are determined by

$$\begin{aligned} f_1(k, 1) &= \rho + \delta \\ f(k, 1) &= c + \delta k + (b - d)k. \end{aligned}$$

The disease free steady state is stable when  $\frac{\bar{\alpha}}{b+\psi+\phi_1} \leq 1$ , and unstable when  $\frac{\bar{\alpha}}{b+\psi+\phi_1} > 1$ .

## 7.2 Proof of Proposition 3: Disease endemic steady state

If  $\frac{\bar{\alpha}}{b+\psi+\phi_1} \leq 1$ , the infectious diseases are eradicated and only the disease free steady state exists. Therefore, the disease endemic steady states only exist when  $\frac{\bar{\alpha}}{b+\psi+\phi_1} > 1$ .

From equation (13), we have

$$\mu_3 = \lambda_1 - \lambda_2 \alpha'(h)(1 - i - r)i,$$

where  $\lambda_1$  is the marginal cost(MC) of health expenditure and  $\lambda_2 \alpha'(h)(1 - i - r)i$  is marginal benefit (MB) of health expenditure. There are two scenarios. When  $\mu_3 > 0$ , that is MC is greater than MB, we have a corner solution with  $h^\infty = 0$ , and the other is an interior solution ( $h^\infty > 0$ ) determined by  $\mu_3 = 0$ , that is,

$$\lambda_1 = \lambda_2 \alpha'(h)(1 - i - r)i. \quad (42)$$

In a steady state, all economic variables converge to the steady state, and all co-state variables grow at the same rate with  $\frac{\dot{\lambda}_1}{\lambda_1} = \frac{\dot{\lambda}_2}{\lambda_2} = \frac{\dot{\lambda}_3}{\lambda_3} = \frac{\dot{\lambda}_4}{\lambda_4}$ .

From equation (12), we have

$$\frac{\dot{\lambda}_1}{\lambda_1} = -(\rho - b + d + \phi_1 i + \phi_2 r).$$

Combined with equation (14), we have

$$f_1(k, 1 - i) = \rho + \delta. \quad (43)$$

From equation (17), we have

$$\begin{aligned} \frac{\dot{\lambda}_4}{\lambda_4} &= \frac{e^{-\Delta}[u(c) - \chi \nu(\phi_1 i + \phi_2 r)]}{\lambda_4} = -(\rho - b + d + \phi_1 i + \phi_2 r) \\ \lambda_4 &= \frac{e^{-\Delta}[u(c) - \chi \nu(\phi_1 i + \phi_2 r)]}{-(\rho - b + d + \phi_1 i + \phi_2 r)} \end{aligned}$$

From equation (15), we have

$$\frac{\dot{\lambda}_2}{\lambda_2} = \frac{e^{-\Delta}\chi\phi_1\nu'(\phi_1i + \phi_2r)}{\lambda_2} + \frac{\lambda_1(f_2(k, 1 - i) - k\phi_1)}{\lambda_2} - [\alpha(h)(1 - i - r) - \alpha i - \psi - b - \phi_1 + 2\phi_1i + \phi_2r] - \frac{\lambda_3(\psi + \phi_1r)}{\lambda_2} - \frac{\lambda_4\phi_1}{\lambda_2}$$

Moreover, since

$$\frac{\dot{\lambda}_2}{\lambda_2} = \frac{\dot{\lambda}_1}{\lambda_1} = -(\rho - b + d + \phi_1i + \phi_2r),$$

we can rearrange the above equation as

$$[\alpha(h)(1 - i - r) - \rho - d - \psi - \phi_1 + \phi_1i - \alpha i] \lambda_2 + (\psi + \phi_1r)\lambda_3 = e^{-\Delta}M1, \quad (44)$$

where

$$M1 = \chi\phi_1\nu'(\phi_1i + \phi_2r) + u'(c)(f_2(k, 1 - i) - k\phi_1) + \phi_1 \frac{u(c) - \chi\nu(\phi_1i + \phi_2r)}{(\rho - b + d + \phi_1i + \phi_2r)}.$$

From equation (16), we have

$$\frac{\dot{\lambda}_3}{\lambda_3} = \frac{e^{-\Delta}\chi\phi_2\nu'(\phi_1i + \phi_2r)}{\lambda_3} - \frac{\lambda_1k\phi_2}{\lambda_3} + \frac{\lambda_2(\alpha i - \phi_2i)}{\lambda_3} - (-b - \phi_2 + \phi_1i + 2\phi_2r) + \frac{\lambda_4\phi_2}{\lambda_3}$$

Moreover, since

$$\frac{\dot{\lambda}_3}{\lambda_3} = \frac{\dot{\lambda}_1}{\lambda_1} = -(\rho - b + d + \phi_1i + \phi_2r),$$

we can rearrange the above equation as

$$(\phi_2i - \alpha i)\lambda_2 + (-\rho - d - \phi_2 + \phi_2r)\lambda_3 = e^{-\Delta}M2, \quad (45)$$

where

$$M2 = \chi\phi_2\nu'(\phi_1i + \phi_2r) - u'(c)k\phi_2 + \phi_2 \frac{u(c) - \chi\nu(\phi_1i + \phi_2r)}{(\rho - b + d + \phi_1i + \phi_2r)}.$$

Define  $\tilde{\lambda}_2 = \lambda_2/e^{-\Delta}$ . Then, combing equations (44) and (45), we can solve for  $\lambda_2$ :

$$\begin{aligned} \tilde{\lambda}_2 &= \lambda_2/e^{-\Delta} \\ &= \frac{M1(-\rho - d - \phi_2 + \phi_2r) - M2(\psi + \phi_1r)}{[\alpha(h)(1 - i - r) - \rho - d - \psi - \phi_1 + \phi_1i - \alpha i](-\rho - d - \phi_2 + \phi_2r) - (\psi + \phi_1r)(\phi_2i - \alpha i)}. \end{aligned} \quad (46)$$

Substituting  $\lambda_1$  and  $\lambda_2$  into equation (13), and we have

$$\mu_3/e^{-\Delta} = \tilde{\lambda}_2\alpha'(h)(1 - i - r)i - u'(c).$$

Thus, the steady state  $(i^\infty, r^\infty, k^\infty, c^\infty, h^\infty)$  is determined by the following five equations

$$\left[ \phi_2 + \phi_1 \frac{\alpha - \phi_2}{\phi_1 - \alpha} \right] r^2 + \left[ -(b + \phi_2) + \psi \frac{\alpha - \phi_2}{\phi_1 - \alpha} + \phi_1 + \phi_1 \frac{\psi + b}{\phi_1 - \alpha} \right] r + \psi + \psi \frac{\psi + b}{\phi_1 - \alpha} = 0 \quad (47)$$

$$i = \frac{\alpha - \phi_2}{\phi_1 - \alpha} r + 1 + \frac{\psi + b}{\phi_1 - \alpha}, \quad (48)$$

$$f_1(k, 1 - i) = \rho + \delta, \quad (49)$$

$$f(k, 1 - i) = c + h + \delta k + (b - d)k + (\phi_1 i + \phi_2 r)k, \quad (50)$$

$$\tilde{\lambda}_2 \alpha'(h)(1 - i - r)i - u'(c) \geq 0, \quad h \geq 0, \quad (\tilde{\lambda}_2 \alpha'(h)(1 - i - r)i - u'(c))h = 0. \quad (51)$$

Let us focus on equation (51) and define a function  $G$ :

$$G(h) = \tilde{\lambda}_2 \alpha'(h)(1 - i - r)i - u'(c), \quad (52)$$

where  $\tilde{\lambda}_2$  is defined in equation (46), and  $i, r, k$  and  $c$  are all functions of  $h$  defined by equations (47)-(50). Therefore, the function  $G$  is a function of health expenditure  $h$ , i.e. the net marginal benefit of health expenditure. There are two scenarios.

One is the disease endemic steady state with a corner solution  $h^\infty = 0$ . When  $G(h)|_{h=0} \leq 0$ , that is, the MC is larger than the MB, we have a corner solution with  $h^\infty = 0$  and  $\alpha(h^\infty) = \bar{\alpha}$ . With  $h^\infty = 0$ , we can solve for the steady state  $r^\infty, i^\infty, k^\infty$  and  $c^\infty$  with equations (47) - (50).

The other is the disease endemic steady state with an interior solution  $h^\infty > 0$ . When  $G(h)|_{h=0} > 0$ , that is, the MB is larger than the MC, we have an interior solution and  $h^\infty$  is determined by  $G(h) = 0$ . Given the optimal  $h^\infty$ , the steady state variables  $r^\infty, i^\infty, k^\infty$  and  $c^\infty$  are determined by equation (47) - (50).

### 7.3 Proof of Lemma 1

*Proof.* Consider any feasible path  $(\mathbf{x}_t, \mathbf{z}_t)$  with the same initial condition  $\mathbf{x}_0^*$ .

It follows from (34) that

$$\lambda_{4,t} = \lambda_{4,0} + \int_0^t e^{-\Delta_\tau^*} [u(c_\tau^*) - \chi \nu (\phi_1 i_\tau^* + \phi_2 r_\tau^*)] d\tau.$$

The transversality condition (41) implies

$$\lim_{t \rightarrow \infty} [\lambda_{4,0} + \int_0^t e^{-\Delta_\tau^*} [u(c_\tau^*) - \chi \nu (\phi_1 i_\tau^* + \phi_2 r_\tau^*)] d\tau] \Delta_t^* = 0.$$

Since  $\lim_{t \rightarrow \infty} \Delta_t^* = +\infty$ , the identity above is satisfied only if

$$\lambda_{4,0} = - \int_0^t e^{-\Delta_\tau^*} [u(c_\tau^*) - \chi\nu(\phi_1 i_\tau^* + \phi_2 r_\tau^*)] d\tau$$

which in turn implies

$$\begin{aligned} \lambda_{4,t} &= - \int_0^\infty e^{-\Delta_\tau^*} [u(c_\tau^*) - \chi\nu(\phi_1 i_\tau^* + \phi_2 r_\tau^*)] d\tau + \int_t^0 -e^{-\Delta_\tau^*} [u(c_\tau^*) - \chi\nu(\phi_1 i_\tau^* + \phi_2 r_\tau^*)] d\tau \\ &= - \int_t^\infty e^{-\Delta_\tau^*} [u(c_\tau^*) - \chi\nu(\phi_1 i_\tau^* + \phi_2 r_\tau^*)] d\tau. \end{aligned}$$

For any  $\Delta$ , since  $d\Delta = (\rho - b + d + \phi_1 i + \phi_2 r)dt$  we have

$$\int_t^\infty e^{-\Delta_\tau} d\tau = \int_t^\infty \frac{e^{-\Delta_\tau} d\Delta_\tau}{\rho - b + d + \phi_1 i_\tau + \phi_2 r_\tau}. \quad (53)$$

Let denote  $q_\tau = \Delta_\tau$ , if  $\tau = t$  then  $q_t = \Delta_t$ . If  $\tau = \infty$  then  $q_\infty = \Delta_\infty = \infty$ . Moreover, since  $0 \leq i, r \leq 1$  from (53) we get

$$\begin{aligned} \frac{1}{\rho - b + d + \phi_1 + \phi_2} \int_{\Delta_t}^\infty e^{-q} dq &\leq \int_t^\infty e^{-\Delta_\tau} d\tau \leq \frac{1}{\rho - b + d} \int_{\Delta_t}^\infty e^{-q} dq \\ \Leftrightarrow \frac{e^{-\Delta_t}}{\rho - b + d + \phi_1 + \phi_2} &\leq \int_t^\infty e^{-\Delta_\tau} d\tau \leq \frac{e^{-\Delta_t}}{\rho - b + d}. \end{aligned} \quad (54)$$

It follows from (36), (54) and using the l'Hôpital's rule we have

$$\begin{aligned} 0 &\leq \lim_{t \rightarrow \infty} \Delta_t \int_t^\infty e^{-\Delta_\tau^*} [u(c_\tau^*) - \chi\nu(\phi_1 i_\tau^* + \phi_2 r_\tau^*)] d\tau \leq (u(A) + \chi\nu(\phi_1 + \phi_2)) \lim_{t \rightarrow \infty} \Delta_t \int_t^\infty e^{-\Delta_\tau^*} d\tau \\ &\leq (u(A) + \chi\nu(\phi_1 + \phi_2)) \lim_{t \rightarrow \infty} \frac{\Delta_t e^{-\Delta_t^*}}{\rho - b + d} \\ &= \frac{u(A) + \chi\nu(\phi_1 + \phi_2)}{\rho - b + d} \lim_{t \rightarrow \infty} \frac{\Delta_t}{e^{\Delta_t^*}} = \frac{u(A) + \chi\nu(\phi_1 + \phi_2)}{\rho - b + d} \lim_{t \rightarrow \infty} \frac{\dot{\Delta}_t}{\Delta_t^* e^{\Delta_t^*}} \\ &= \frac{u(A) + \chi\nu(\phi_1 + \phi_2)}{\rho - b + d} \lim_{t \rightarrow \infty} \frac{\rho - b + d + \phi_1 i + \phi_2 r}{\rho - b + d + \phi_1 i^* + \phi_2 r^*} \frac{1}{e^{\Delta_t^*}} = 0 \end{aligned}$$

because

$$\frac{\rho - b + d}{\rho - b + d + \phi_1 + \phi_2} \leq \frac{\rho - b + d + \phi_1 i + \phi_2 r}{\rho - b + d + \phi_1 i^* + \phi_2 r^*} \leq \frac{\rho - b + d + \phi_1 + \phi_2}{\rho - b + d} \text{ and } e^{\Delta_t^*} \rightarrow \infty \text{ as } t \rightarrow \infty.$$

Therefore, for any feasible  $\Delta_t$ ,

$$\lim_{t \rightarrow \infty} \lambda_{4,t} \Delta_t = - \lim_{t \rightarrow \infty} \Delta_t \int_t^\infty e^{-\Delta \tau} [u(c_\tau) - \chi \nu (\phi_1 i_\tau + \phi_2 r_\tau)] d\tau = 0. \quad (55)$$

Together with (41) we have

$$\lim_{t \rightarrow \infty} \lambda_{4,t} (\Delta_t - \Delta_t^*) = 0.$$

Note that , since  $\lim_{t \rightarrow \infty} \Delta_t = \infty$  so from (55) we get  $\lim_{t \rightarrow \infty} \lambda_{4,t} = 0$ . □

## 7.4 Proof of Lemma 2

*Proof.* i) From (29) we get  $\lambda_1 \geq 0$ . Therefore  $\lambda_1 k \geq 0$  and (35) implies

$$\lim_{t \rightarrow \infty} \lambda_1 (k^* - k) \leq 0.$$

ii) From Proposition 3, either  $i^*$  converges to a positive steady state or the sequence lies in the unit interval and does not converge to 0. In both cases, it follows from (35) that

$$\lim_{t \rightarrow \infty} \lambda_2 = 0.$$

As  $i$  is bounded, we have

$$\lim_{t \rightarrow \infty} \lambda_2 (i^* - i) = 0.$$

Similarly, we have

$$\lim_{t \rightarrow \infty} \lambda_3 = 0 \text{ and hence } \lim_{t \rightarrow \infty} \lambda_3 (r^* - r) = 0.$$

□

## 7.5 Proof of Lemma 3

*Proof.* We have

$$\begin{aligned} & \bar{H}(\mathbf{x}_t^*, \mathbf{z}_t^*, \lambda_t) - \bar{H}(\mathbf{x}_t, \mathbf{z}_t, \lambda_t) \\ = & e^{-\Delta^*} [u(c_t^*) - u(c_t)] - \lambda_1 (c_t^* - c_t) - \lambda_1 (h_t^* - h_t) + \lambda_2 (1 - i^* - r^*) i^* (\alpha(h^*) - \alpha(h)) \end{aligned}$$

As  $u(c)$  is concave we have

$$e^{-\Delta^*} [u(c_t^*) - u(c_t)] \geq e^{-\Delta^*} u'(c_t^*) (c_t^* - c_t) = \lambda_1 (c_t^* - c_t). \quad (56)$$

Since  $\alpha(h)$  is convex, we have

$$\alpha(h^*) - \alpha(h) \leq \alpha'(h^*)(h^* - h). \quad (57)$$

From (29) and (30),  $0 < e^{-\Delta}u'(c^*) = \lambda_1 = \lambda_2\alpha'(h^*)(1 - i^* - r^*)i^*$  thus  $(1 - i^* - r^*)i^*\lambda_2 < 0$  because  $\alpha'(h^*) < 0$ . Therefore, from (57) we have

$$(1 - i^* - r^*)i^*\lambda_2(\alpha(h^*) - \alpha(h)) \geq (1 - i^* - r^*)i^*\lambda_2\alpha'(h^*)(h^* - h) = \lambda_1(h^* - h). \quad (58)$$

It follows from (56) and (58) that  $\bar{H}(\mathbf{x}_t^*, \mathbf{z}_t^*, \lambda_t) - \bar{H}(\mathbf{x}_t^*, \mathbf{z}_t^*, \lambda_t) \geq 0$ .  $\square$

## 7.6 Proof of Lemma 4

*Proof.* As  $0 \leq h^* \leq A$  and  $\alpha'' > 0$ , we have  $\alpha'(0) \leq \alpha'(h^*) \leq \alpha'(A) \leq 0$ . Therefore,  $0 \leq |\alpha'(h^*)(1 - i^* - r^*)i^*\lambda_2| \leq |\alpha'(0)i^*\lambda_2|$ . Therefore the usual transversality condition implies  $\lim_{t \rightarrow \infty} |\alpha'(h^*)(1 - i^* - r^*)i^*\lambda_2| = 0$ . Thus from (30) we have  $\lim_{t \rightarrow \infty} \lambda_1 = \lim_{t \rightarrow \infty} \alpha'(h^*)\lambda_2(1 - i^* - r^*)i^* = 0$ .

We have

$$\begin{aligned} \lim_{t \rightarrow \infty} H(\mathbf{x}_t^*, \mathbf{z}_t^*, \lambda_t) = & \lim_{t \rightarrow \infty} e^{-\Delta^*} [u(c^*) - \chi\nu(\phi_1 i^* + \phi_2 r^*)] + \lim_{t \rightarrow \infty} \lambda_1 \{f(k^*, 1 - i^*) - c^* - h^* - \delta k^* - (b - d - \phi_1 i^* - \phi_2 r^*)k^*\} \\ & + \lim_{t \rightarrow \infty} \lambda_2 i^* E + \lambda_3 F + \lim_{t \rightarrow \infty} \lambda_4 \{\rho - b + d + (\phi_1 i^* + \phi_2 r^*)\} \end{aligned}$$

where

$$\begin{aligned} E &= \alpha(1 - i^* - r^*) - \psi - b - \phi_1 + \phi_1 i^* + \phi_2 r^*, \\ F &= \psi i^* - b r^* - \phi_2 r^* + r^*(\phi_1 i^* + \phi_2 r^*). \end{aligned}$$

It is easy to see that

$$\begin{aligned} 0 &\leq |E| \leq \alpha + \psi + b + 2\phi_1 + \phi_2 < \infty, \\ 0 &\leq |F| \leq \psi + b + 2\phi_2 + \phi_1 < \infty. \end{aligned}$$

Using the results of Lemma 1 and Lemma 2 with the fact that  $k^*, c^*, h^*, i^*, u(c^*), \nu(\phi_1 i^* + \phi_2 r^*)$  and  $f$  are bounded, it implies that the transversality condition (40) is satisfied.  $\square$