Abstract

We analyze how to optimally engage in social distancing in order to minimize the spread of an infectious disease. We identify conditions under which any optimal policy is single-peaked, i.e., first engages in increasingly more social distancing and subsequently decreases its intensity. We show that an optimal policy might substantially delay measures that decrease the transmission rate to create herd-immunity and that engaging in social distancing sub-optimally early can increase the number of fatalities. Finally, we find that optimal social distancing can be an effective measure to reduce the death rate of a disease.

Keywords — Social Distancing, SIR model, Time-Optimal Control of an Epidemic

1 Introduction

This paper theoretically analyzes how to optimally engage in measures to contain the spread of an infectious disease. We formalize this question in the context of a standard model from epidemiology, the Susceptible-Infected-Recovered (SIR) model (Kermack and McKendrick, 1927). This model divides the population into three groups: susceptible, infected and recovered. People transition from one group into another at given exogenously specified rates depending on the size of each sub-population. We extend this model by introducing an additional parameter controlled by the planner that affects the rate at which the disease is transmitted. This parameter captures political measures such as social distancing (SD), the lockdown of businesses, schools, universities and other institutions. While such measures reduce the spread of the disease, they often come at a substantial economic and social cost. We model this trade-off by considering a planner who faces convex cost in the number of infected (capturing the number of people whose death is caused by the disease) and the reduction in transmission rate (capturing the cost of shutting down society). Convexity of the cost allows for the possibility that the probability of dying from the disease increases in the fraction of the population which is infected, for example because fewer infected receive treatment due to capacity constraints of the healthcare system.
Our analysis identifies several features of any optimal policy: First, whenever the probability of dying from the disease does not increase too quickly in the number of infected, the optimal policy is single peaked in the sense that first the measures to reduce the transmission rate are escalated until some point in time, and after this point measures are reduced. Second, if furthermore the cost of reducing the transmission rate is linear, meaning that closing half of society for two days is equally costly as closing all of society for one day, only the most extreme policies are used at any point in time. Either, the planner imposes the maximal possible lockdown or no restrictions at all. Intuitively, the planner can achieve a greater effect by imposing a more extreme policy for a shorter time and thus does not find it optimal to use intermediate policies. These results imply that for linear cost the optimal policy has a simple structure and consists of maximally three phases: first it imposes no restrictions then it imposes as many restrictions as possible, and finally in the third phase imposes no restrictions at all. This result drastically simplifies the search for an optimal policy as the planner has to only optimize over the start and end time of the social distancing period. We furthermore show that in this case the number of infected peaks at most twice under any optimal policy.

We then illustrate some further insights using parameters that are commonly chosen to model the Covid-19 epidemic that started in 2020. We first characterize the optimal timing of the social distancing period given that the planner has access to a certain budget of days of social distancing. We find that the optimal social distancing is often substantially delayed. For example, if the planer has a budget of 100 days of social distancing in the next 360 days after 0.1% of the population are infected it is optimal to delay social distancing by 50 days. This initial period of letting the disease spread uncontrolled is useful as it creates “herd immunity” and thereby reduces the overall severity of the epidemic. We identify analytic conditions on the parameters under which benefit of herd immunity is so strong that sometimes more social distancing can increase the number of people that die from the epidemic. We provide an example in which more people die when social distancing is imposed from day 0-100 compared to day 50-100. As this example suggests, benefit of optimally timing social distancing measures is often large and we illustrate this by comparing social distancing in the first $t$ days after 0.1 percent are infected to $t$ days of optimally timed social distancing. Finally, we describe the optimal timing of social distancing.

**Related Literature** Our theoretical results extend the literature on the optimal control of an infectious disease (for an overview see chapter 5 in Wickwire, 1977). In general, there are three policy tools commonly used to control an infectious disease: (i) immunization, (ii) testing and isolation of infected individuals and (iii) lockdown measures that lead to a reduction in the contact rate for the whole population. Most of the preceding literature has focused on immunization and selective isolation measures. Abakuks (1973) considers how to optimally isolate infectious population if infectious population can be instantaneously isolated. Abakuks (1972, 1974) determine the optimal vaccination strategy in the same framework. Morton and Wickwire (1974) and Wickwire (1975) extend the previous work on vaccination and isolation by considering flow controls. Behncke (2000) considers more general functional forms and Hansen and Day (2011) allows for hard bounds on the control, while considering vaccination and isolation policies simultaneously. Ledzewicz and Schättler (2011) considers vaccination and treatment in a model with population growth. Bolzoni et al. (2017) consider the optimal use of vaccination, isolation, culling. The general insight from this literature is that for linear cost of vaccination/testing the optimal policy switches from vaccinating/testing the population at the maximal feasible intensity until some point in time to vaccinating/testing no one after that point in time.
While the previously discussed literature has analyzed vaccination and isolation policies, little is known about optimal lockdown policies. In general, the analysis of optimal lockdown policies in the standard SIR model is a challenging problem, due to its non-linear structure and the three dimensional state space. To the best of our knowledge the first article that discusses the optimal social distancing or lockdown policies in the SIR model is Chapter 4 in Behncke (2000). In a model without terminal cost, this paper observes that the optimal policy depends only on the shadow price difference between infected and susceptible. Bolzoni et al. (2017) consider the case where social distancing is costless and are interested in minimizing the time until the share of infected falls below a prespecified threshold. In contrast, we model the cost of social distancing and consider the number of people who die from the disease instead of how long the disease lasts. They establish that the optimal policy switches once from no social distancing to maximal social distancing. An optimal policy in our model with linear cost is also bang-bang, but switches at most twice: from no social distancing to maximal social distancing and back. For non-linear cost we show that the optimal policy is no longer bang-bang, but still maintains a single-peaked shape. Our paper contributes to this literature by theoretically deriving properties of an optimal lockdown policy if one takes the cost of lockdown into account about which little has been known before.

Most closely related to our work, Miclo et al. (2020) derive the optimal policy that keeps the number of infected below some exogenous threshold when the cost of reducing the transmission rate is linear. The two studies complement each other as we focus on the case where the cost of infected is not too convex, while their model intuitively corresponds to the case where the cost is infinitely convex at the threshold that captures the capacity of the healthcare system. Federico and Ferrari (2020) analyze the optimal confinement of a pandemic in a variant of the SIR model where the transmission rate is modeled as a diffusive stochastic process whose trend can be controlled via costly lockdown measures.

While we focus on the optimal timing of measures to reduce the spread of the epidemic other aspects have been studied in the recent theoretical literature: Lipnowski and Ravid (2020) study the optimal design of group tests. Acemoglu et al. (2020a) analyse the effect of increased testing on incentives to engage in risky behavior. Erol and Ordoñez (2020) study how social distancing changes the network of social and economic interactions.

Finally, our paper relates to the recent, quickly growing literature in economics that numerically studies optimal policies for the epidemic of Covid-19 in the context of SIR models (see for example Alvarez et al., 2020; Kissler et al., 2020; Toda, 2020; Acemoglu et al., 2020b; Akbarpour et al., 2020; Garibaldi et al., 2020; Kantner, 2020). While it is not a goal of this paper to make any recommendations for the current Covid-19 epidemic we hope that the formal analysis and insights into the structure of optimal policies will be useful in the rapidly evolving discussion of how to optimally react to the Covid-19 epidemic.

2 The Evolution of an Epidemic

The SIR Model To model the spread of an infectious disease we rely on a basic model from epidemiology, the Susceptible Infected Recovered (SIR) model introduced in Kermack and McKendrick (1927). We divide society into three groups: susceptible $s$, infected $i$, and the rest which is either immune to the disease as they

---

1. This insight generalizes to our analysis (see Proposition 1).
2. See Assumption 1 below.
3. See for example Atkeson (2020); Barro et al. (2020); Dewatripont et al. (2020); Piguillem et al. (2020); Stock (2020); de Walque et al. (2020); Guerrieri et al. (2020).
recovered from it or died. We denote by \( s(t) \) the fraction of the population that is healthy, but susceptible to disease at time \( t \), and by \( i(t) \) the fraction of the population that is infected. The SIR model assumes the number of people that gets infected by a single infected person is deterministic proportional to the fraction of society \( s(t) \) that is still susceptible to the disease. Intuitively, if only a small fraction of society is susceptible to the disease it is unlikely that an infected person meets a susceptible person. The mass of healthy people that become infected during \( dt \) thus equals

\[
\beta(t)i(t)s(t)dt,
\]

where the transmission rate \( \beta(t) \in [b, \bar{b}] \) captures both the infectivity of the disease, as well as measures society has taken to influence the speed at which the disease spreads (like social distancing). At rate \( \gamma > 0 \) infected become non-infected, by either recovering from the disease, or dying of it, such that during a short time span \( dt \), the fraction of infected is reduced by \( \gamma i(t)dt \). The susceptible and infected populations thus for every \( t \in [0, \infty) \) evolve according to the following dynamics

\[
\begin{align*}
s'(t) &= -\beta(t)i(t)s(t), \quad s(0) = s_0, \\
i'(t) &= \beta(t)i(t)s(t) - \gamma i(t), \quad i(0) = i_0,
\end{align*}
\]

where \( s_0, i_0 \in (0, 1) \) are given initial values satisfying \( s_0 + i_0 \leq 1 \).

**Reproduction Rate** We follow the terminology from the epidemiology literature (see e.g. Allman and Rhodes, 2004, p. 287) and define the basic reproduction rate \( R_0 \) of the disease to equal the number of people a single infected person infects when he is the only person carrying the disease and the transmission rate is maximal

\[
R_0 = \frac{\bar{b}}{\gamma}.
\]

The effective reproduction rate \( R(t) = \frac{1}{\gamma} \beta(t)s(t) \) at time \( t \) equals the number of people infected by an infected person at time \( t \) and depends on the current transmission rate. Intuitively, \( R_0 \) measures how quickly the disease would spread if no measures were taken and \( R(t) \) measures how quickly the disease spreads at time \( t \) given the policy used by the planner and the current fraction of susceptible population.

**Control of the Transmission Rate** The time-dependent transmission rate \( \beta: [0, \infty) \to B \) takes values in a compact interval \( B = [b, \bar{b}] \subset (0, \infty) \). We denote by \( \bar{b} \) the maximal transmission rate and by \( b \) the minimal transmission rate that can be achieved through policy measures, like social distancing, the closure of schools, wearing masks, etc. The set of admissible controls \( B \) consists of all measurable functions \( \beta: [0, \infty) \to B \).

We introduce two convex cost functions \( v: [0, 1] \to [0, \infty) \) and \( c: B \to [0, \infty) \), where \( v \) captures the economic and health cost of infected population and \( c \) captures the cost of reducing the transmission rate. We only make minimal assumption on \( c \) and assume that it is non-increasing, convex, and continuous. Without loss we normalize the cost associated with the highest transmission rate to zero, \( c(\bar{b}) = 0 < c(b) \) for all \( b \in [b, \bar{b}] \). We suppose that \( v \) is convex, continuously differentiable, and strictly increasing and without loss set \( v(0) = 0 \). The planner trades off the consequences of a higher number of infected with the economic
and social cost of reducing the transmission rate and thus aims at minimizing the costs
\[ \int_0^\infty v(i(t)) + c(\beta(t)) \, dt. \]  

(3)

We assume that a vaccine arrives at deterministic time \( T > 0 \) and every susceptible person is vaccinated immediately, but no cure is available at that point in time.\(^4\) As the transmission rate \( \beta \) equals zero after \( T \), an easy calculation\(^6\) shows that the objective (3) can be rewritten as
\[ J(\beta) = \int_0^T v(i(t)) + c(\beta(t)) \, dt + \bar{v}(i(T)), \]  

(4)

where \( \bar{v}(i) = \int_0^i \frac{v(z)}{z^\gamma} \, dz \) for all \( i \in [0,1] \). A policy \( \beta^* \) is optimal if it minimizes the objective \( J \)
\[ \beta^* \in \text{arg min}_{\beta \in \mathbb{R}} J(\beta). \]  

(5)

**Discussion of the Costs**  The cost \( v(i) \) measures the number of people that die per unit of time if a share \( i \) of the population is infected. Convexity of \( v \) captures that the probability of dying from the disease might be higher if a large share of the population is infected and the hospital system is overwhelmed.\(^7\) The cost function \( c \) captures the economic and social cost of measures taken to reduce the transmission rate. For example if social distancing measures are imposed which require the closure of most businesses this comes at a substantial economic cost. The convexity of \( c \) captures that some institutions, i.e. businesses, schools, etc, are less costly to close down than others and the planner could start by closing down these “least essential” institutions or impose other low cost measures such as making masks mandatory. The convexity of \( c \) creates an incentive for the planner to smooth out lockdown measures over time and it is often instructive to focus on the linear case where such an incentive is absent.

We throughout assume that both costs \( c \) and \( v \) are measured in the same units. This could be achieved by assigning to each death costs equal to the “value of life” as defined by different government agencies and insurance companies.\(^8\) To assign a dollar value to the death of a person is normal practice for governmental agencies, such as the US Environmental Protection Agency, who decide on life-saving measures that are often (very) costly to implement.

\(^4\)Note that we assume that \( T > 0 \) is a deterministic constant. We considered the case of a random arrival time of a vaccine in a working paper version of this paper. If a cure also becomes available at time \( T \), the terminal cost equals zero \( \bar{v}(i) = 0 \) as all infected get cured immediately at time \( T \). We considered this case in the working paper version of this paper.

\(^5\)Note that future costs are not discounted in our model. In order to not make our results and analysis any more involved, we refrain from introducing a discount factor of the form \( e^{-\rho t} \) in the cost functional. In the cases of a relatively short time horizon \( T \) or a small discount factor \( \rho \), the effects of discounting are limited.

\(^6\)As, after the comprehensive vaccination of the population, there are no new infections, the share of infected population evolves according to \( i'(t) = -\gamma i(t) \) and is given by \( i(t) = i(T) e^{-\gamma (t-T)} \) for \( t \geq T \). The share of the population that dies after the arrival of the vaccine at time \( T \) thus equals \[ \int_T^\infty v \left( i(T) e^{-\gamma (t-T)} \right) \, dt = \int_0^{i(T)} \frac{v(z)}{z^\gamma} \, dz. \]

\(^7\)The cost \( v \) can not only capture the people who die of the disease directly, but also those who die because other medical conditions remain untreated as an indirect consequence of the disease.

\(^8\)For example the US Environmental Protection Agency assigned a value $9.1 million to a life in 2010 and the Department of Transportation $9.2 million in 2014 see [https://en.wikipedia.org/wiki/Value_of_life#United_States](https://en.wikipedia.org/wiki/Value_of_life#United_States).
A Microfoundation for the Reduction of the Transmission Rate  A natural microfoundation for the cost of reducing the transmission rate is to consider \( k \in \mathbb{N} \) measures which can be enacted by the planner individually. Such measures for example include the closing of various institutions or making the use of masks mandatory. The planner can choose to enact each measure \( j \in \{1, \ldots, k\} \) with an intensity \( \alpha_j \in [0, 1] \). The cost of implementing a measure depends linearly on \( \alpha_j \), i.e., there exist \( c_j \in (0, \infty) \) such that the cost of enacting measure \( j \) with intensity \( \alpha_j \) amounts to \( c_j \alpha_j \). The transmission rate is now a decreasing, differentiable function \( \beta^\circ : [0, 1]^k \to B \) of the extend \( \alpha_1, \ldots, \alpha_k \) to which the different measures are enacted. The cost of reducing the transmission rate is now the solution of an optimization problem where the planner enacts measures optimally in order to achieve a given transmission rate \( b \in B \)

\[
c(b) = \min_{\alpha \in [0,1]^k} \left\{ \sum_{j=1}^k c_j \alpha_j : \beta^\circ(\alpha) \leq b \right\}.
\]

While the cost function \( c \) defined in (6) is in general decreasing and continuous it might not be convex.\(^9\) Nevertheless convexity can easily be checked in concrete examples. For example as we verify in Lemma 11 in the Appendix choosing \( \beta^\circ(\alpha) = b \sum_{j=1}^k \pi_j^\alpha_j \), for some efficacy \( \pi_1, \ldots, \pi_k \) of the different measures leads to a non-increasing, continuous convex cost \( c \). This micro-foundation highlights that due to the possible non-linearity of \( c \) our setting can encompass problems where the principal at each point in time chooses which measures to enact from a potentially complex set of policies. We henceforth do not explicitly mention the different measures which are used to reduce the transmission rate, but simply consider the reduction in transmission rate directly.

3 Optimal Policies

The next result shows existence of an optimal policy and provides necessary conditions that any solution of the optimal control problem (5) must satisfy.

**Proposition 1.** An optimal policy exists. Let \( \beta^* \in B \) be an optimal policy and denote by \( s^*, i^* : [0, T] \to [0, 1] \) the associated number of susceptible and infected satisfying (1). There exists a function \( \eta^* : [0, T] \to \mathbb{R} \) with \( \eta^*(T) = \frac{v(i^*(T))}{\sum_{t=0}^T s^*(T)} \) such that for almost all \( t \in [0, T] \) it holds

\[
(\eta^*)'(t) = \eta^*(t) \beta^*(t) i^*(t) - v'(i^*(t)) + \frac{v(i^*(t)) + c(\beta^*(t)) - \min_{b \in B} \left[ \frac{1}{i^*(t)} v(i^*(T)) s^*(T) b + c(b) \right]}{i^*(t)},
\]

(7)

\[
(\beta^*)'(t) \in \arg\min_{b \in B} \left[ \eta^*(t) i^*(t) s^*(t) b + c(b) \right].
\]

Moreover, we have \( \eta^*(t) > 0 \) for all \( t \in [0, T] \).

The proof of Proposition 1 relies on a sequence of auxiliary results we establish in the appendix using standard arguments from control theory that can, e.g., be found in Clarke (2013). The existence of an optimal policy follows as the convexity of \( c \) and \( B \) ensures compactness of the policy space which leads to

---

\(^9\)In the context of consumer choice this function is known as the expenditure function, i.e., the minimal expenditure needed to achieve a given utility level (see e.g. Mas-Colell et al., 1995, Section 3.E). As far as we know there are no known sufficient conditions for the convexity of the expenditure function in the utility.
the existence of an optimal policy. Pontryagin’s optimality principle then yields that there exist two Lagrange multipliers $\lambda^*_1$ - the marginal cost of susceptible population – and $\lambda^*_2$ – the marginal cost of infected population – such that the optimal control is only a function of these multipliers. As the transmission rate controls how fast susceptible population becomes infected, the optimal control can be determined from the difference in marginal cost $\eta^* = \lambda^*_2 - \lambda^*_1 > 0$ according to (8). This Lagrange multiplier $\eta^*(t)$ has a clear interpretation as the marginal cost of having an additional susceptible person infected. Proposition 1 goes beyond the Pontryagin maximum principle as it shows that the two Lagrange multipliers $\lambda^*_1, \lambda^*_2$ can be summarized in a single Lagrange multiplier $\eta^*$ whose dynamics can be expressed as a forward-backward equation (independent of $\lambda^*_1, \lambda^*_2$) thereby effectively reducing the dimension of the problem by 1. The downside of this simplification is that it introduces an explicit dependence of the dynamics on the terminal state of the system. Nevertheless, this reduction is what allows us to explicitly characterize features of the optimal policy later. To prove this reduction we exploit the specific structure of the dynamics of the SIR model. Furthermore, Proposition 1 shows that $\eta^*(t) > 0$ which establishes that the planner always benefits from having fewer infected. While this result seems intuitive it is not at all obvious as we later show that the planner sometimes benefits from a higher transmission rate when not using an optimal policy (see Section 3.4 and Figure 1).

Remark 1. Note that Proposition 1 makes no statement about the uniqueness of optimal policies. Proposition 1 and the following results provide properties that any optimal policy satisfies.

We henceforth restrict attention to optimal policies where (8) is satisfied for all $t \in [0, T]$. This assumption is inconsequential in the sense that changing the policy $\beta^*$ on a set of measure zero does not affect the share of susceptible and infected population ($s^*, i^*$).

3.1 Gain from Reducing the Transmission Rate

An economically important quantity is $g^*(t) = \eta^*(t)i^*(t)s^*(t)$. Proposition 1 implies that $g^* : [0, T] \to (0, \infty)$ is strictly greater than zero and determines the set of optimal controls through

$$\beta^*(t) \in \arg \min_{b \in B} \left[ g^*(t)b + c(b) \right].$$

(9)

The above equation implies that $g^*$ admits a direct economic interpretation as the gain of reducing the transmission rate in an optimal policy. From $g^*(t)$ the optimal policy can be simply determined as follows: Engage in all measures that reduce the transmission rate $b$ until the marginal cost of measures $-c'(b)$ that further reduce the transmission rate exceed $g^*(t)$. To simplify notation we define $M(g) = \min_{b \in B} [gb + c(b)]$ for all $g \in \mathbb{R}$. Our next result characterizes how the gain from reducing the transmission rate evolves over time.

Proposition 2 (Properties of the Gain from Reducing the Transmission Rate).

\footnote{The nonlinear nature of the SIR dynamics make the minimization of $\beta \mapsto J(\beta)$ in general non-convex. Therefore, standard uniqueness results for optimal controls do not apply here. Establishing uniqueness results in non-convex situations is a challenging task and can in many cases only be carried out for small time horizons $T$ (see, e.g., Kandhway and Kuri (2014) and Fister et al. (1998)).}

\footnote{It is a priori not clear that an optimal $\beta^*$ can be modified such that (8) is satisfied for all $t \in [0, T]$ and such that $\beta^*$ remains measurable. We prove this in Lemma 17 in the appendix.}

3
(i) The gain from reducing the transmission rate \( g^* : [0,T] \to (0,\infty) \) evolves according to

\[
(g^*)'(t) = -\gamma g^*(t) + s^*(t) (M(g^*(t)) - M(g^*(T))) - [v'(i^*(t))i^*(t) - v(i^*(t))] \, ,
\]

with terminal condition \( g^*(T) = \frac{1}{\tau} v(i^*(T))s^*(T) \).

(ii) For all \( t \in [0,T) \) the gain \( g^* \) satisfies the following bounds

\[
0 < g^*(T) < g^*(t) < \frac{1}{\gamma} \left[ v(1) + v'(1) \left( e^{(T-t)\gamma} - 1 \right) \right] .
\]

(iii) Any optimal control \( \beta^* \) is non-increasing in \( g^* \), i.e. \( g^*(t) > g^*(t') \implies \beta^*(t) \leq \beta^*(t') \).

Equation (10) characterizes how the incentive to reduce the transmission rate evolves over time. The first term \( -\gamma g^*(t) \) shows that the incentive is exponentially decaying at the rate at which the infected die or become cured. Intuitively, when there are fewer infected the benefits from reducing the transmission rate is smaller.

The term \( M(g^*(t)) - M(g^*(T)) \) quantifies how much the objective function of the planner increases as a function of \( g^*(t) \). This term is strictly positive, and increasing in \( g^*(t) \). Finally, the third term \( v'(i^*(t))i^*(t) - v(i^*(t)) \) is non-negative, increasing in \( i^* \) and a measure of the increase in cost due to the convexity of \( v \). The term implies that the gain from reducing the transmission rate falls more quickly when more people are currently infected.

Part (ii) of the proposition states that the incentive to reduce the transmission rate is minimal at the final time \( T \). Part (iii) establishes that the optimal control is non-increasing in the transmission rate. Together part (ii) and (iii) imply that the optimal control is maximal at the terminal time \( T \).

**Corollary 3.** Any optimal control \( \beta^* \in B \) satisfies \( \beta^*(t) \leq \beta^*(T) \) for all \( t \in [0,T] \).

Intuitively, as the benefit of having fewer infected is smallest at time \( T \) as people which get infected at time \( T \) do not cause further infections due to the availability of a vaccine at time \( T \).

**Remark 2.** While Proposition 2 shows that for the optimal policy the gain of reducing the transmission rate is positive, this result does not extend to sub-optimal policies. In Proposition 9 we identify parameters and policies such that a point-wise reduction in the transmission rate leads to a higher cost of infected and reducing the transmission rate is thus not beneficial.

### 3.2 Herd-Immunity

We denote by \( \tau \) the first time society achieves herd-immunity, that is the first time every infected person infects fewer than one person on average without intervention of the planner.\(^{13,14}\)

\[
\tau = \min \left\{ t \in [0,T] : \frac{5}{\gamma} s(t) \leq 1 \right\} .
\]

\(^{12}\)These facts are proved in Lemma 16. As an example suppose that the percentage of infected who dies grows linearly in the fraction of the population that is infected \( v(i) = ki^2 \). In this case \( v'(i^*(t))i^*(t) - v(i^*(t)) = ki^2 \).

\(^{13}\)Throughout the article we use the conventions that \( \inf \emptyset = \min \emptyset = \infty \) and \( \sup \emptyset = \max \emptyset = -\infty \).

\(^{14}\)In the context of the SIR model the definition that a society has achieved herd-immunity if \( s(t) \leq \frac{5}{\gamma} \) is standard in the literature on mathematical epidemiology (see, e.g., Section 7.3 in Allman and Rhodes (2004)).
It is immediate from the definition of \( \tau \) that the share of infected population is strictly decreasing after \( \tau \), independent of the policy chosen after time \( \tau \). Our next result shows that the optimal policy always implies a non-decreasing transmission rate after \( \tau \), i.e. that lockdown measures get less stringent after herd-immunity has been achieved in society.

**Proposition 4.** Fix an optimal policy \( \beta^* \). The gain from reducing the transmission rate \( g^* \) is strictly decreasing on \([\tau,T]\) and the optimal transmission rate \( \beta^* \) is non-decreasing on \([\tau,T]\).

Intuitively, after herd-immunity has been achieved the incentives to reduce the transmission rate are decreasing as each infected has a smaller chance of infecting someone in the future as over time one gets closer to the arrival of a vaccine, and the share of susceptible population decreases.

For any disease with basic reproduction rate \( R_0 \) below 1, herd immunity is achieved at time zero and an immediate corollary of Proposition 4 is that any optimal policy reduces the lockdown measures over time. Especially, in the case of linear costs, i.e., there exists \( \delta > 0 \) such that \( c(\beta) = \delta (\bar{b} - \beta) \), this implies that the planner optimally enforces the maximal lockdown until some time \( t^* \) and afterwards uses no measures to reduce the transmission rate.

**Corollary 5.** Assume that \( R_0 \leq 1 \) and let \( \beta^* \) be an optimal policy. Then, \( \beta^* \) is non-decreasing on \([0,T]\) and if \( c \) is linear then there exists \( t^* \in [0,T] \) such that\(^{15}\)

\[
\beta^*(t) = \begin{cases} 
\bar{b} & \text{for } t \in [0,t^*) \\
\underline{b} & \text{for } t \in (t^*,T].
\end{cases}
\]

### Less Social Distancing Leads to Faster Herd-Immunity

As we argue next, the fact that for \( R_0 \leq 1 \) the optimal policy engages in a lockdown as quickly as possible does not generalize to diseases with \( R_0 > 1 \). Intuitively, for such diseases there is an additional motive at play for the planner before herd immunity is reached. By increasing the transmission rate the planner generates more infections, which eventually leads to more people who are immune to the disease, and faster herd-immunity.

Consider two policies \( \beta \) and \( \hat{\beta} \) and their associated paths of susceptible and infected \((s^\beta, i^\beta)\) and \((s^{\hat{\beta}}, i^{\hat{\beta}})\). We say that \( \beta \) induces less social distancing than \( \hat{\beta} \) if for all \( t,t' \in [0,T] \) with \( s^{\hat{\beta}}(t) = s^{\beta}(t') \) we have\(^{16}\)

\[ \beta(t) \geq \hat{\beta}(t') \].

Our definition of less social distancing requires that for any given share of susceptible population the transmission rate is higher if the planner enforces less social distancing.

**Proposition 6.** If \( \beta \) induces less social distancing than \( \hat{\beta} \) then there will be fewer susceptible \( s^{\beta}(t) \leq s^{\hat{\beta}}(t) \) at every point in time \( t \in [0,T] \) and herd-immunity will be reached faster under the policy \( \beta \).

Proposition 6 shows that less social distancing leads to faster herd-immunity. A natural conjecture following this observation is that too much social distancing can increase the number of people who die from a disease. We establish that this is indeed the case in Section 3.4.

\(^{15}\)Here and in the sequel we use the convention that \([x,x] = (x,x) = (x,x) = \emptyset \) for all \( x \in \mathbb{R} \). In particular, any of the intervals in (12) might be empty.

\(^{16}\)Note that \( s^{\hat{\beta}} \) is strictly decreasing. Consequently, for fixed \( t' \in [0,T] \) there exists at most one \( t \in [0,T] \) such that \( s^{\hat{\beta}}(t) = s^{\beta}(t') \) which is given by \( t = (s^{\hat{\beta}})^{-1}(s^{\beta}(t')) \). The requirement \( \beta(t) \geq \hat{\beta}(t') \) thus becomes \( \beta((s^{\hat{\beta}})^{-1}(s^{\beta}(t'))) \geq \hat{\beta}(t') \) in this case.
3.3 Limited Capacity Effects

We next introduce an assumption that restricts to environments where the cost is not too convex in the share of infected population.\textsuperscript{17}

Assumption 1 (Limited Capacity Effects). The cost $v$ is twice continuously differentiable and satisfies
\[
\frac{v''(i)}{\gamma} < \frac{\gamma}{4} \frac{b^2 - b}{b^2} |c'(b^-)| \quad \text{for all } i \in [0, 1].
\]

Assumption 1 admits an intuitive interpretation: As $\frac{1}{\gamma} v'(i)$ is the probability with which an additional infected dies when a share $i$ of the population is infected. The condition thus requires that the probability of dying from the disease can not be too sensitive to the share of infected population, compared to the marginal cost of reducing the transmission rate. It is always satisfied if the probability of dying from the disease is independent of which fraction of the population is infected, which corresponds to the case where $v$ is linear. The main reason for non-linear $v$ is that the death rate increases for larger number of infected due to limited hospital capacity. Assumption 1 thus rules out too large capacity effects that arise from the overload of the healthcare system. It is thus a reasonable approximation if the number of infected exceeds the hospital capacity.

Our next result shows that in this case the optimal transmission rate is quasi-convex.

**Proposition 7.** Suppose that Assumption 1 is satisfied and let $\beta^* \in B$ be an optimal control. Then there exists $t^* \in [0, T]$ such that

(i) $\beta^*$ is non-increasing on $[0, t^*]$ and non-decreasing on $[t^*, T]$, i.e., $\beta^*$ is quasi-convex, and

(ii) the fraction of infected population $i^*$ is strictly log-concave on $[0, t^*]$ and thus admits at most one local maximum on $[0, t^*]$.

Proposition 7 establishes that any optimal policy is single peaked, in the sense that the measures to decrease the transmission rate are first escalated until some point in time and then reduced over time. Any policy where a reduction in measures is followed by an increase is suboptimal. Furthermore, in the initial period where SD measures are escalated and the transmission rate falls the fraction of infected is single peaked. The proof of Proposition 7 is quite involved. The main idea of the proof is to differentiate the dynamics of the gain from reducing the transmission rate derived in (10) and then bounding the second derivative of $g^*$ using Assumption 1 to establish that $g^*$ is quasi-concave.

We next show that if Assumption 1 is satisfied and $c$ is linear then the optimal policy involves only the two most extreme controls $\beta^*(t) \in \{\underline{b}, \overline{b}\}$. The assumption that the cost $c$ of measures that reduce the transmission rate is linear has a simple interpretation in the context of social distancing: Shutting down half of the economy for two days is equally costly as shutting down the whole economy for a single day.\textsuperscript{18} While there is no normative reason for this assumption we think of it as a natural baseline for the analysis, and potentially a good approximation of the trade-offs once “cheap measures” to reduce the transmission rate (such as wearing masks) are exhausted.\textsuperscript{19}

\textsuperscript{17}We denote by $c'(b^-) = \lim_{b \to b^-} \frac{c(b)-c(b^-)}{b-b^-} \leq 0$ the left-derivative of $c$ at $b$ (which exist due to convexity of $c$).

\textsuperscript{18}This implicitly assumes that the transmission rate $\beta$ depends linearly on the shut down of the economy.

\textsuperscript{19}We note here that the theoretical literature on optimal policies to manage an infectious disease has almost exclusively focused on the case of linear cost for which we provide a solution in Proposition 8.
We call a policy $\beta_{[t_1, t_2]}$ an interval policy if it first engages in no social distancing until time $t_1$, then in maximal social distancing from time $t_1$ to $t_2$, and in no social distancing after time $t_2$, i.e. for a.e. $t \in [0, T]$

$$
\beta_{[t_1, t_2]}(t) = \begin{cases} 
\delta & \text{for } t \in [0, t_1) \\
\frac{b}{\delta} & \text{for } t \in (t_1, t_2) \\
\frac{\delta - b}{\delta} & \text{for } t \in (t_2, T] 
\end{cases}
$$

(13)

**Proposition 8.** Suppose that Assumption 1 is satisfied and $c$ is linear\(^{20}\) and let $\beta^* \in B$ be an optimal control.

(i) Then $\beta^*$ is an interval policy, i.e. there exist $0 \leq t_1^* \leq t_2^* \leq T$ such that $\beta^* \equiv \beta_{[t_1^*, t_2^*]}$.

(ii) The fraction of infected $i^*$ under the policy $\beta^*$ is strictly increasing on $[0, t_1^*]$, and has at most one local maximum in $[0, t_2^*]$, and at most one local maximum in $[t_2^*, T]$.

Proposition 8 drastically simplifies the search for an optimal policy as it implies that any optimal policy is characterized by the two points in time $(t_1^*, t_2^*)$. Note that the proposition does not rule out that any of the intervals is empty. In particular, reducing the transmission rate by the maximal amount at every point in time as well as taking no measures at all to reduce the transmission rate can be optimal.

The first economic insight of Proposition 8 is that when the cost of reducing the transmission rate is linear and the hospital capacity does not have a too strong effect on the death rate, then it is never optimal to use intermediate measures for a longer time (i.e. closing only parts of the economy) as doing so is dominated by implementing maximal measures for a shorter time. The second part of the proposition establishes that under the optimal policy the number of infected peaks at most twice. The second peak always happens after the planner has stopped using lockdown measures. Thus, whenever one observes the number of infected peak more than twice one can conclude that either the planner acted suboptimally or one of the assumptions of the model must be violated.

### 3.4 More Social Distancing Can Increase the Number of Death

We next prove that, perhaps surprisingly, more social distancing can lead to more people getting infected.

**Assumption 2.** The basic reproduction number $R_0 = \frac{\delta}{\gamma}$ without social distancing and the reproduction number $\dot{R}_0 = \frac{b}{\gamma}$ with maximal social distancing satisfy

$$s(0)R_0 > cR_0.$$  

Assumption 2 is satisfied if initially the share of infected population is sufficiently small (i.e. $s(0) \approx 1$), maximal social distancing is effective enough to control the spread of the virus ($R_0 \leq 1$), and without social distancing the virus is spreading with a reproduction rate $R_0$ higher than $c \approx 2.72$. For reference the estimated $R_0$ for the initial Covid-19 variant ranged between 1.5 and 6.68 (see Liu and Gayle, 2020). Intuitively, Assumption 2 applies to diseases which are very infectious without social distancing and whose spread can be controlled well through social distancing.

\(^{20}\)By linearity we mean that there exists $\delta > 0$ such that $c(b) = \delta(b) - b$ for all $b \in B$. 

11
As we have shown in Proposition 8 whenever the cost of infected is not too convex every optimal policy is an interval policy. Our next result establishes that, in the class of interval policies, more social distancing can lead to more people getting infected.

**Proposition 9.** Assume that Assumption 2 is satisfied and let

\[ t_2 > \frac{e^{R_0} - s(0)}{(1 - s(0))\gamma[R_0 s(0) - e^{R_0}]} \]

Then there exists a time \( \tilde{t}_1 < t_2 \) such that every interval policy \( \beta_{[\tilde{t}_1, t_2]} \) with \( t_1 \in (0, \tilde{t}_1] \) leads to a strictly lower total number of infected \( \gamma \int_0^T i(t) \, dt \) than the interval policy \( \beta_{[0, t_2]} \) for \( T \) large enough.

Consider the case where a constant fraction of infected die from the disease and therefore, the total number of death is proportional to the total number of infected. Intuitively, Proposition 9 establishes that for a very infectious disease \(^{21}\) if a vaccine is only expected to arrive in the far out future the number of people who die from the disease can be reduced by delaying engaging in social distancing (and thus engaging in less social distancing overall)! An incomplete intuition for this result is as follows: by delaying social distancing, more people get infected in the initial phase which leads to fewer people being infected in the future. This intuition is of course highly incomplete as it does not provide a reason why the aforementioned benefit should exceed the number of people who die due to the additional infected in the initial phase.

The proof of this result follows from a detailed analysis of the SIR dynamics. As a first auxiliary result we establish in Proposition 19 in the appendix that if a vaccine never arrives \( (T = \infty) \) the number of people who die over the course of the epidemic in an interval policy \( \beta_{[t_1, t_2]} \) is decreasing in the number of people who die between time \( t_1 \) and \( t_2 \). To minimize total death a social planner should thus aim at maximizing the total number of infected in the period of social distancing. Starting social distancing later, i.e. increasing \( t_1 \) thus has two effects: (i) it decreases the total number of infected in \([t_1, t_2]\) as the time-interval gets shorter (ii) it increases the total number of infected in \([t_1, t_2]\) as the number of infected at time \( t_1 \) is larger. Through a careful analysis of the change in the corresponding ODEs we establish that under Assumption 2 the latter effect dominates the former and thus delaying social distancing is optimal. This establishes the result if a vaccine never arrives and it follows from continuity that this result extends to the case where the vaccine arrives sufficiently late in the future.

**An Example Where Less Social Distancing is Better** Proposition 9 establishes analytically tractable sufficient conditions for more social distancing to increase the number of people who die from the disease. While these conditions are demanding this effect can arise even for plausible parameters as we illustrate in the next numerical example depicted in Figure 1. The example considers a disease with a basic reproduction rate of 2.5, an average infection length of 12 days, where social distancing reduces the transmission rate by 65%, and 1% of infected die. These parameters are chosen in line with estimates for the Covid-19 pandemic in 2020.\(^{22}\) A vaccine is assumed to arrive after 360 days. Starting with 0.1% of the

\(^{21}\)As we have establishes in Corollary 5 starting social distancing immediately is optimal (and the result does thus not hold) for diseases of low infectiousness where \( R_0 \leq 1 \) and thus Assumption 2 which requires \( R_0 > \frac{1}{s(0)} e^{R_0} > 1 \) is violated.

\(^{22}\)Wu et al. (2020) estimate \( R_0 \) for Covid-19 to be between 2.47 and 2.86 or between 2.32 and 2.71 depending on the modelling assumptions. For an average infection length of 10 days Toda (2020) finds a median reproduction rate of 2.9 and documents significant variation across countries. Kantner (2020) calibrate an SIR model to Germany and find a basic reproduction rate of \( R_0 = 2.7 \).
population infected at time zero it compares the number of infected if social distancing is enacted from day $0-100$ to social distancing from day $50-100$. On the left one can see how the number of infected evolves over time under either policy. If there is no social distancing initially the number of infected peaks early and then decreases until day 100, after which it mildly increases until herd immunity is reached at day 126, after which the number of infected decrease. If social distancing is enacted from day 0-100 the number of infected peaks much later and higher at day 168 at which herd-immunity is reached. The right-hand-side plot illustrates the accumulated share of people who died. While with social distancing more people die early from the disease, fewer people die later on. Overall, the share of the population that dies from the disease is 0.177% less when social distancing is enacted from day 50-100 in this example. Notably, this example does not assume convex $v$, which corresponds to limited hospital capacity, and would amplify the difference in dead due to the different height of the peak number of infected.

### 3.5 The Optimal Timing of Social Distancing

We next illustrate the optimal timing of social distancing. For this illustration we again consider a disease with a basic reproduction rate of 2.5, an average infection length of 12 days, where social distancing reduces the transmission rate by 65%, and 1% of infected die.\(^{23}\) A vaccine is assumed to arrive after 360 days.

Throughout this section we only consider linear cost $c$ of reducing the transmission rate and linear cost $v$ in the number of infected. Denote by $B^*$ the set of policies which engage in maximal social distancing on an optimally chosen time interval of some length $l \in [0,T]$ and otherwise in no social distancing, i.e.,

$$B^* = \bigcup_{l \in [0,T]} \arg \min_{t_1 \in [0,T-l]} J(\beta_{[t_1,t_1+l]}).$$

A direct Corollary Proposition 8 (i) is the observation that every optimal policy must be in $B^*$.\(^{24}\)

---

\(^{23}\)These parameter values are chosen in line with the Covid-19 pandemic in 2020 c.f. Footnote 22.

\(^{24}\)This result extends beyond linear $v$ as long as the marginal cost of social distancing is not too small relative to
Corollary 10. Suppose that \(c\) and \(v\) are linear, and \(\beta^*\) is an optimal policy, then \(\beta^* \in B^*\).

Each policy in the set \(B^*\) is completely described by two numbers the start date of social distancing and the length of the social distancing period. A direct implication is that we can determine an optimal policy (given linear cost) by a simple 2-step procedure: First, fix a length of the social distancing period \(l\) and optimize over the start date (this is a simple problem as the start date is a real number). As the length of the social distancing period is fixed to \(l\), linearity of \(c\) and \(v\) entail that the minimization of \(J\) is equivalent to the minimization of the integral \(\int_0^T i(t)dt\), which in turn is equivalent to the minimization of the total number of people that die from the disease. Second, optimize over the length of the social distancing period (which is again a simple optimization problem as the length is a real number). Furthermore, by computing the set \(B^*\) we can describe the set of policies that are optimal given for \(any\) possible linear cost of reducing the transmission rate and an additional infected person.

We depict the set \(B^*\) in right graph of Figure 2. As a first example, consider the case where the planner optimally engages in \(l = 100\) days of social distancing. In this case the optimal policy is to start social distancing on day 50 and end it on day 150. As one can see in the left graph of Figure 2 this leads to a flatter curve of infected over time than social distancing in the first 100 days or no social distancing. Interestingly, the effect of suboptimal social distancing is marginal compared to no social distancing in the sense that while it initially reduces the number of infected substantially, it essentially only delays the peak of infected, but does not substantially flatten it. Optimal social distancing leads to a substantial reduction in the implied death rate within a year: 0.607% under optimal social distancing, 0.892% with social distancing in the first hundred days, and 0.893% without social distancing.

Returning to the right graph of Figure 2, one can see it is often optimal to delay social distancing beyond the date where 0.1% of the population is infected. For example even if it is optimal for the planner to enforce 150 days of social distancing within the next year it is only optimal to start social distancing after 50 days. This observation might be surprising as it implies that if it is not optimal to maintain permanent social distancing, the convexity of the cost of infected life may play a significant role in the optimal policy.

\[
|c'| \geq \frac{4}{\pi^2} \frac{\beta - \beta}{\beta + \beta} \max_{i \in [0,1]} v''(i).
\]
The Value of Optimal Social Distancing

We next illustrate the importance of the optimal timing of social distancing. The left graph of Figure 3 plots the death rate as a function of the number of days of social distancing.\footnote{We used the same parameters as in Figure 1 and 2.} The solid line depicts the death rate when the social distancing measures are optimally timed and the dashed line depicts the death rate if the planner enforces social distancing immediately. As one can see in the figure social distancing can be an effective measure to prevent the death of population. For example, 50 days of optimally timed social distancing (from day 50 to day 100) reduce the death rate by roughly 0.2%. The figure however shows that without optimal timing social distancing is much less effective and to achieve an equal reduction in the example one needs 270 days of social distancing.

The right graph of Figure 3 plots the reduction in cost from engaging in an additional day of optimally timed social distancing. As one can see in the picture the benefit is non-concave. It is initially high and the first days of social distancing safe around 0.005% of the population. It then decreases and peaks again around 300 days of social distancing. Intuitively, this non-concavity results from the fact that engaging in social distancing over almost the whole time until a vaccine arrives can prevent the disease from ever substantially spreading. In contrast engaging in social distancing over a shorter time horizon can not avoid a substantial spread of the disease before the arrival of a vaccine.

4 Conclusion

We derived the optimal policy for social distancing during an epidemic. Our analysis revealed several features of the optimal policy. If the death rate is not too sensitive to the number of infected, the optimal policy consists of two phases: a first phase where the planner enforces more social distancing over time and a second phase where the planner enforces less social distancing over time. Furthermore, if the cost of reducing the
transmission rate is linear, the optimal policy is always extreme. At any point in time either social distancing is carried out to the maximal extent possible or not at all. The intuitive reason for this result is that more extreme measures over a shorter time horizon are more effective than less extreme measures over a longer horizon. We illustrated through an example that the effectiveness of social distancing depends crucially on its optimal timing. Within the context of this example optimal social distancing is often substantially delayed in order to generate herd immunity. Maybe surprisingly, engaging in more, but too early social distancing can increase the number of infected and the fatalities caused by the disease.

As the SIR model models a homogeneous population, it abstracts away from the fact that prevalence of the disease might be different in different sub-populations and that measures to reduce the transmission rate might affect the spread of the disease differently across sub-populations. An interesting question for future research is what measures are optimal given such heterogeneity. Birge et al. (2020) consider this question when sub-populations correspond to different geographical neighborhoods.
A Appendix

Lemma 11. Let \( \beta^\circ(\alpha) = \overline{\beta} \prod_{j=1}^{k} \pi_j^{\alpha_j}, \) with \( \pi \in [0,1]^k. \) The cost function defined (6) is non-increasing, continuous and convex.

Proof of Lemma 11. The claim that \( c \) is non-increasing follows from (6). Next let \( b, b', b'' \in [\underline{b}, \overline{b}], \gamma \in [0,1] \) such that \( \gamma \log(b') + (1 - \gamma) \log(b'') = \log(b). \) Denote by \( \alpha', \alpha'' \in [0,1]^k \) the optimal intensities in (6) given the transmission rate \( b' \) or \( b''), \) i.e., \( c(b') = \sum_{j=1}^{k} c_j \alpha_j' \) and \( c(b'') = \sum_{j=1}^{k} c_j \alpha_j''. \) We have that the log of the transmission rate associated with the intensities \( \gamma \alpha' + (1 - \gamma) \alpha'' \in [0,1]^k \) equals

\[
\log(b) + \sum_{j=1}^{k} \log(\pi_j)[\gamma \alpha'_j + (1 - \gamma) \alpha''_j] = \log(b) + \gamma \sum_{j=1}^{k} \log(\pi_j) \alpha'_j + (1 - \gamma) \sum_{j=1}^{k} \log(\pi_j) \alpha''_j
\]

\[
= \gamma \log(b') + (1 - \gamma) \log(b'') = \log(b).
\]

As the convex combination of the policies \( \alpha' \) and \( \alpha'' \) is feasible this implies that

\[
c(b) \leq \gamma c(b') + (1 - \gamma)c(b'').
\]

Hence, \( x \mapsto c(e^x) \) is convex. This together with the fact that \( c \) is non-increasing implies that \( c \) is convex. This implies continuity of \( c \) on the open interval \( (\underline{b}, \overline{b}) \). Continuity on the boundary follows from the continuity of the objective and constraint in (6).

We next prove a sequence of auxiliary results which together imply Proposition 1. Recall that the function \( M: \mathbb{R} \to \mathbb{R} \) is defined by \( M(y) = \min_{b \in B} [yb + c(b)], y \in \mathbb{R}, \) and note that \( M \) is non-decreasing.

Lemma 12. An optimal policy \( \beta^* \) exists that solves (5). Let \( s^*, i^*: [0,T] \to [0,1] \) be the state processes associated with an optimal control satisfying (1). Then there exist absolutely continuous functions \( \lambda_1^*, \lambda_2^*: [0,T] \to \mathbb{R} \) which satisfy for almost all \( t \in [0,T] \) the dynamics

\[
(\lambda_1^*)(t) = (\lambda_1^*(t) - \lambda_2^*(t)) \beta^*(t) i^*(t), \quad \lambda_1^*(T) = 0,
\]

\[
(\lambda_2^*)(t) = (\lambda_1^*(t) - \lambda_2^*(t)) \beta^*(t) s^*(t) + \gamma \lambda_2^*(t) - v'((i^*)'(t)), \quad \lambda_2^*(T) = \frac{v((i^*)'(T))}{\gamma i^*(T)},
\]

and the optimality condition

\[
\beta^*(t) \in \arg \min_{b \in B} \left[ (\lambda_2^*(t) - \lambda_1^*(t)) i^*(t) s^*(t)b + c(b) \right].
\]

Moreover, for all \( t \in [0,T] \) we have

\[
M(\lambda_2^*(t) - \lambda_1^*(t) i^*(t) s^*(t)) - \gamma \lambda_2^*(t) i^*(t) + v((i^*)'(t)) = M \left( \frac{v((i^*)'(T)) s^*(T)}{\gamma} \right).
\]

Proof of Lemma 12. Suppose an optimal policy \( \beta^* \) exists. The existence of functions \( \lambda_1^*, \lambda_2^*: [0,T] \to \mathbb{R} \) that satisfy (15), (16) and (17) follows from the Pontryagin principle (see, e.g., Clarke, 2013, Theorem 22.2 and Corollary 22.3). We show the existence of an optimal policy by verifying the conditions of Theorem 23.11 in Clarke (2013).
(a) $g(t, (s, i)) = \left(\frac{-is}{+is}\right)$ which implies that $|g(t, (s, i))| \leq 2|is| < 2$.

(b) $B = [h, \bar{b}]$ is closed and convex by definition.

(c) The sets $E = \{(s_0, i_0)\} \times \mathbb{R}_+$ and $Q = [0, T] \times [0, 1]^2$ are closed and $\ell(s_0, i_0, s_T, i_T) = \bar{v}(i_T)$ is lower semicontinuous.

(d) The running cost $\beta \mapsto v(i) + c(\beta)$ is convex as $c$ is convex. Furthermore, $v(i) + c(\beta) \geq 0$.

(e) The projection set is given by $\{(s_0, i_0)\}$ and thus bounded.

(f) As $\beta \in B$ it follows that $|\beta| \leq \bar{b}$. This verifies (f) (ii).

Moreover, the constant control $\beta(t) = \bar{b}$ has finite costs. We have hence verified that there exists an optimal policy.

In the following we suppose that $\beta^* \in \mathcal{B}$ is an optimal control and denote by $s^*, i^* : [0, T] \to [0, 1]$ the associated state processes satisfying (1). Moreover, we denote by $\lambda_1^*, \lambda_2^* : [0, T] \to \mathbb{R}$ the Lagrange variables from Lemma 12. Compactness of $B$ and continuity of $c$ ensure that for all $t \in [0, T]$ the function $b \mapsto [\lambda_2^*(t) - \lambda_1^*(t)]i^*(t)s^*(t)b + c(b)$ attains its minimum on $B$.

We introduce the new Lagrange variable

$$\eta^*(t) = \lambda_2^*(t) - \lambda_1^*(t).$$

The variable $\eta^*(t)$ has a clear interpretation: it measures the marginal change in the cost with respect to infecting susceptible population. Intuitively speaking, $\eta^*(t)$ measures the additional cost if one additional person is infected at time $t$ given the optimal policy is used. Note that by (16) at each time $t$ the optimal control $\beta^*(t)$ depends on $\lambda_1^*(t)$ and $\lambda_2^*(t)$ only through their difference $\eta^*(t) = \lambda_2^*(t) - \lambda_1^*(t)$.

**Lemma 13.** Let $\beta^* \in \mathcal{B}$ be an optimal control and suppose that the optimality condition (16) holds for all $t \in [0, T]$. Suppose that $t_0 \in [0, T]$ satisfies $\eta^*(t_0) \leq 0$. Then it holds that $\lim_{t \to t_0} \beta^*(t) = \beta^*(t_0) = \bar{b}$.

**Proof of Lemma 13.** First note that the assumption $\eta^*(t_0) \leq 0$ ensures that the function $b \mapsto \eta^*(t)i^*(t)s^*(t)b + c(b)$ attains its global minimum on $B$ at $\bar{b}$. Hence (16) implies that $\beta^*(t_0) = \bar{b}$. Next let $(t_n)$ be a sequence such that $t_n \to t_0$ as $n \to \infty$. Suppose by contradiction that there exists a subsequence such that $\lim_{n \to \infty} \beta^*(t_n) =: b_0 < \bar{b}$. Next note that (16) ensures for all $n \in \mathbb{N}$ that (recall that $c(\bar{b}) = 0$)

$$\eta^*(t_n)i^*(t_n)s^*(t_n)\beta^*(t_n) + c(\beta^*(t_n)) \leq \eta^*(t_n)i^*(t_n)s^*(t_n)\bar{b}.$$  

This implies that

$$\eta^*(t_n)i^*(t_n)s^*(t_n) \geq \frac{c(\beta^*(t_n))}{\bar{b} - \beta^*(t_n)}.$$  

Taking the limit $n \to \infty$ yields the contradiction

$$0 \geq \lim_{n \to \infty} \eta^*(t_n)i^*(t_n)s^*(t_n) \geq \lim_{n \to \infty} \frac{c(\beta^*(t_n))}{\bar{b} - \beta^*(t_n)} = \frac{c(b_0)}{\bar{b} - b_0} > 0.$$  

Therefore, we have $\lim_{t \to t_0} \beta^*(t) = \bar{b} = \beta^*(t_0)$.

The next result shows that the cost of additional infected $\eta^*$ is characterized by a differential equation that does not depend on $\lambda_1^*$ and $\lambda_2^*$. Moreover, we show that both $\lambda_1^*$ and $\lambda_2^*$ can be recovered from $\eta^*$.
Lemma 14. The variable $\eta^*$ solves

\[
(\eta^*)' (t) = \eta^*(t)\beta^*(t)i^*(t) - v'(i^*(t)) + \frac{v(i^*(t)) + c(\beta^*(t)) - M \left( \frac{v(i^*(T))s^*(T)}{\gamma} \right)}{\gamma} \tag{22}
\]

with terminal condition $\eta^*(T) = \frac{v(i^*(T))}{\gamma i^*(T)}$. Conversely, suppose that $i, s, \beta, \eta : [0, T] \to \mathbb{R}$ satisfy

\[
s'(t) = -\beta(t)i(t)s(t), \quad s(0) = s_0, \\
i'(t) = \beta(t)i(t)s(t) - \gamma i(t), \quad i(0) = i_0, \\
\eta'(t) = \eta(t)\beta(t)i(t) - v'(i(t)) + \frac{v(i(t)) + c(\beta(t)) - M \left( \frac{v(i(T))s(T)}{\gamma} \right)}{\gamma}, \quad \eta(T) = \frac{v(i(T))}{\gamma i(T)}, \\
\beta(t) \in \arg\min_{b \in B} \left[ \eta(t)i(t)s(t) \right] + c(b),
\]

then

\[
\lambda_1(t) = \frac{1}{\gamma} \left[ \eta(t)\beta(t)s(t) + \frac{v(i(t)) + c(\beta(t)) - M \left( \frac{v(i(T))s(T)}{\gamma} \right)}{\gamma} \right] \tag{24}
\]

solves (15).

Proof of Lemma 14. First note that it follows from (17) and (16) that

\[
-\eta^*(t)\beta^*(t)i^*(t)s^*(t) + \gamma \lambda_2^*(t)i^*(t) = -M(\eta^*(t)i^*(t)s^*(t)) + \gamma \lambda_2^*(t)i^*(t) + c(\beta^*(t))
\]

\[
= v(i^*(t)) + c(\beta^*(t)) - M \left( \frac{v(i^*(T))s^*(T)}{\gamma} \right). \tag{25}
\]

Then (15) implies that

\[
(\eta^*)'(t) = (\lambda_2^*)'(t) - (\lambda_1^*)'(t) = -\eta^*(t)\beta^*(t)s^*(t) + \gamma \lambda_2^*(t) - v'(i^*(t)) + \eta^*(t)\beta^*(t)i^*(t)
\]

\[
= \eta^*(t)\beta^*(t)i^*(t) - v'(i^*(t)) + \frac{v(i(t)) + c(\beta(t)) - M \left( \frac{v(i^*(T))s^*(T)}{\gamma} \right)}{\gamma}. \tag{26}
\]

Next suppose that $s, i$ and $\eta$ solve (23) and that $\lambda_1$ and $\lambda_2$ are given by (24). Observe that it holds that $\lambda_1(T) = 0$ and $\lambda_2(T) = \frac{v(i(T))}{\gamma i(T)}$. Next note that the envelope theorem ensures that

\[
\frac{\partial}{\partial \gamma} \left[ \eta(t)i(t)s(t) + c(\beta(t)) \right] = \frac{\partial}{\partial \gamma} \min_{b \in B} \left[ \eta(t)i(t)s(t) + c(b) \right] = \beta(t) \frac{\partial}{\partial \gamma} \left[ \eta(t)i(t)s(t) \right]. \tag{27}
\]
Then it holds that
\[
\lambda_2'(t) = \frac{\partial}{\partial t} \left[ \frac{\eta(t)\beta(t)i(t)\gamma_1(t) + v(i(t)) + c(\beta(t)) - M \left( \frac{v(i(t))s(T)}{\gamma} \right)}{\gamma_1(t)} \right]
\]
\[
= \frac{\beta(t) \frac{\partial}{\partial t} [\eta(t)i(t)s(t)] + v'(i(t)) \gamma_1'(t)}{\gamma_1(t)}
\]
\[
- \frac{\left( \eta(t)\beta(t)i(t)s(t) + v(i(t)) + c(\beta(t)) - M \left( \frac{v(i(t)s(T))}{\gamma} \right) \right) \gamma_1'(t)}{\gamma_1(t)^2}
\]
\[
= \frac{1}{\gamma} \left( \beta(t) \eta'(t)i(t) + \beta(t) \eta(t)s'(t) \right)
\]
\[
+ \frac{v'(i(t)) - v(i(t)) + c(\beta(t)) - M \left( \frac{v(i(t)s(T))}{\gamma} \right)}{i(t)} \gamma_1'(t)
\]
\[
= \frac{1}{\gamma} \left( \beta(t) \eta'(t)i(t) + \beta(t) \eta(t)s'(t) \right)
\]
\[
+ \frac{\gamma_1'(t)}{\gamma_1(t)} \left( \eta(t)\beta(t)i(t) - \eta'(t) \right)
\]
\[
= \beta(t) \eta(t) \frac{\gamma_1'(t)}{\gamma_1(t)} \left( s'(t) + i'(t) \right) + \frac{\eta'(t)}{\gamma} \left( \beta(t)s(t) - \frac{i'(t)}{i(t)} \right)
\]
\[
= -\beta(t) \eta(t)i(t) + \eta'(t).
\]

Therefore we obtain that
\[
\lambda_2'(t) = \frac{v(i(t)) + c(\beta(t)) - M \left( \frac{v(i(t))s(T)}{\gamma} \right)}{i(t)}
\]
\[
= \gamma \lambda_2(t) - \eta(t)\beta(t)s(t) - v'(i(t)) = \left( \lambda_1(t) - \lambda_2(t) \right) \beta(t)s(t) + \gamma \lambda_2(t) - v'(i(t)).
\]

Similarly, \( \lambda_1 \) satisfies
\[
\lambda_1'(t) = \lambda_2'(t) - \eta'(t) = [\eta'(t) - \eta(t)\beta(t)i(t)] - \eta'(t) = (\lambda_1(t) - \lambda_2(t)) \beta(t)i(t).
\]

**Lemma 15** (More Infected are Costly). The function \( \eta^* \) satisfies \( \eta^*(t) > 0 \) for all \( t \in [0, T] \).

**Proof of Lemma 15.** Suppose that there exists \( t \in [0, T] \) such that \( \eta^*(t) \leq 0 \). Then Lemma 13 shows that \( \beta^*(t) = \bar{\beta} \). Moreover, Lemma 13 ensures that \( \beta^* \) is continuous at \( t \) and hence \( \eta^* \) is differentiable at \( t \). Then (22) shows (recall that \( c(\bar{\beta}) = 0 \))
\[
(\eta^*)'(t) = \eta^*(t)\beta^*(t)i^*(t) + \frac{v(i^*(t)) - M \left( \frac{v(i^*(T))s^*(T)}{\gamma} \right)}{i^*(t)} - v'(i^*(t))i^*(t).
\]
Since \( v \) is convex and since \( M \left( \frac{v(i^*(T))s^*(T)}{\gamma} \right) > 0 \) we thus obtain that
\[
(\eta^*)'(t) < \eta^*(t)\beta^*(t)i^*(t) \leq 0
\]
We conclude from the terminal condition \( \eta^*(T) = \frac{v(i^*(T))}{s^*(T)} > 0 \) that \( \eta^*(t) > 0 \) for all \( t \in [0, T] \).
Proof of Proposition 2.

Proof of Lemma 16.

Lemma 16. The function \( Q \) is non-negative and non-decreasing.

Proof of Lemma 16. The monotonicity of \( Q \) is a straightforward consequence of the monotonicity and the convexity of \( v \). Indeed, convexity of \( v \) implies for all \( i_1, i_2 \in [0, 1] \) that

\[
v(i_1) - v(i_2) \geq v'(i_2)(i_1 - i_2).
\]

This and monotonocity of \( v \) show for all \( 0 \leq i_1 \leq i_2 \leq 1 \)

\[
Q(i_2) - Q(i_1) = v'(i_2)i_2 - v(i_2) - v'(i_1)i_1 + v(i_1) = [v(i_1) - v(i_2) - v'(i_2)(i_1 - i_2)] + (v'(i_2) - v'(i_1))i_1 \geq 0.
\]

Finally this proves for all \( i \in [0, 1] \)

\[
Q(i) \geq Q(0) = -v(0) = 0.
\]

Proof of Proposition 2. Recall that \( g^*(t) = \eta^*(t)i^*(t)s^*(t), \ t \in [0, T] \). We compute \( (g^*)'(t) \)

\[
(g^*)'(t) = (\eta^*)'(t)i^*(t)s^*(t) + (\eta^*(t))(i^*)'(t)s^*(t) + \eta^*(t)i^*(t)(s^*)'(t)
\]

\[
= [(\lambda_1^*(t) - \lambda_2^*(t))\beta^*(t)s^*(t) + \lambda_2^*(t)\gamma - v'(i^*(t))) - (\lambda_1^*(t) - \lambda_2^*(t))\beta^*(t)i^*(t)]i^*(t)s^*(t)
\]

\[
+ (\lambda_2^*(t) - \lambda_1^*(t))s^*(t)[\beta^*(t)i^*(t)s^*(t) - \gamma i^*(t)]
\]

\[
- (\lambda_2^*(t) - \lambda_1^*(t))i^*(t)\beta^*(t)i^*(t)s^*(t)
\]

\[
= (\lambda_1^*(t) - \lambda_2^*(t))\beta^*(t)\gamma i^*(t)s^*(t) - \gamma i^*(t)\beta^*(t)i^*(t)s^*(t)
\]

\[
+ [\gamma \lambda_2^*(t) - v'(i^*(t)))]i^*(t)s^*(t) - \gamma (\lambda_2^*(t) - \lambda_1^*(t))s^*(t) = \gamma (\lambda_1^*(t) - v'(i^*(t)))i^*(t)s^*(t).
\]

This and the fact that \( g^*(t) = (\lambda_2^*(t) - \lambda_1^*(t))i^*(t)s^*(t) \) imply

\[
(g^*)'(t) = \gamma \lambda_1^*(t)i^*(t)s^*(t) - v'(i^*(t))i^*(t)s^*(t) = -\gamma g^*(t) + \gamma \lambda_2^*(t)i^*(t)s^*(t) = v(i^*(t))i^*(t)s^*(t).
\]

Next (17) implies that

\[
(g^*)'(t) = -\gamma g^*(t) + s^*(t) \left[ M(g^*(t)) - M \left( \frac{v(i^*(t))s^*(t)}{\gamma} \right) + v(i^*(t)) - v'(i^*(t))i^*(t) \right].
\]

Using that \( g^*(T) = \eta^*(T)i^*(T)\gamma^{s^*(T)} = \frac{\gamma i^*(T)s^*(T)}{\gamma} \) we obtain

\[
(g^*)'(t) = -\gamma g^*(t) + s^*(t) \left[ M(g^*(t)) - M(g^*(T)) - v(i^*(t))i^*(t) \right].
\]
We next argue that we have that $g^*(t) > g^*(T) > 0$ for all $t \in [0, T)$. Let $t \in [0, T)$ and suppose that $g^*(t) \leq g^*(T)$. Since $M$ is non-decreasing we have $M(g^*(t)) \leq M(g^*(T))$. Convexity of $v$ ensures that $v'(i^*(t))i^*(t) - v(i^*(t)) \geq 0$ (see Lemma 16). Moreover, by Lemma 15 we have $g^*(t) > 0$ and consequently (10) implies that $(g^*)'(t) < 0$. This contradicts continuity of $g^*$ and therefore proves $g^*(t) > g^*(T) > 0$ for all $t \in [0, T)$.

Next, we verify the upper bound in (11). To this end let $Q: [0, 1] \to \mathbb{R}$ satisfy $Q(i) = v'(i)i - v(i)$. Note that Lemma 16 implies that $Q$ is non-decreasing. Note that (10), monotonicity of $M$ and the fact that $g^*(t) > g^*(T)$ ensure that

\[ (g^*)'(t) = -\gamma g^*(t) + s^*(t)(M(g^*(t)) - M(g^*(T))) - [v'(i^*(t))i^*(t) - v(i^*(t))] \]

\[ > -\gamma g^*(t) - s^*(t)Q(i^*(t)) \geq -\gamma g^*(t) - Q(1). \]

Gronwall’s lemma shows that

\[ g^*(t) < e^{(T-t)\gamma}g^*(T) + \frac{Q(1)}{\gamma} \left( e^{(T-t)\gamma} - 1 \right). \]

Using that $g^*(T) = \frac{\nu(i^*(T))s^*(T)}{\gamma} \leq \frac{v(1)}{\gamma}$ and $Q(1) = v'(1)1 - v(1)$ we get

\[ g(t) < \frac{1}{\gamma} \left( e^{(T-t)\gamma}v(i^*(T))s^*(T) + (v'(1) - v(1)) \left( e^{(T-t)\gamma} - 1 \right) \right) \leq \frac{1}{\gamma} \left[ v(1) + v'(1) \left( e^{(T-t)\gamma} - 1 \right) \right]. \]

Finally, take two points in time $t, t' \in [0, T]$. Then (16) shows that

\[ g^*(t)\beta^*(t) + c(\beta^*(t)) \leq g^*(t)\beta^*(t') + c(\beta^*(t')) \]

and

\[ g^*(t')\beta^*(t') + c(\beta^*(t')) \leq g^*(t')\beta^*(t) + c(\beta^*(t)). \]

Adding these two inequalities yields that

\[ (g^*(t) - g^*(t'))(\beta^*(t) - \beta^*(t')) \leq 0 \]

Thus, $g^*(t) > g^*(t')$ implies $\beta^*(t) \leq \beta^*(t')$. 

By (16) we have that $\beta^*(t) \in \arg \min_{b \in B} g^*(t)b - c(b)$ for almost all $t \in [0, T]$. By potentially changing $\beta^*$ on a set of measure zero we suppose in the sequel that $\beta^*(t)$ attains the minimum for all $t \in [0, T]$ (i.e., (16) holds for all $t \in [0, T]$). Note that this change does not affect the trajectories of $s^*, i^*, \lambda_1^*$ and $\lambda_2^*$. In Lemma 17 below we show that this modification preserves the measurability of $\beta^*$.

**Lemma 17.** There exists an optimal policy such that $\beta^*(t) \in \arg \min_{b \in B}[g^*(t)b + c(b)]$ for all $t \in [0, T]$.

**Proof.** Let $\beta^*$ be an arbitrary optimal policy and $g^*$ the associated gain from reducing the transmission rate. Let $S: \mathbb{R} \to [b, \overline{b}]$ be a function that satisfies for all $y \in \mathbb{R}$ that $M(y) = yS(y) + c(S(y))$. Such a function $S$ exists by the axiom of choice and since $b \mapsto yb + c(b)$ is continuous and hence attains a minimum on $[b, \overline{b}]$. Then we can show as in Proposition 2 that $S$ is non-increasing. Hence any such $S$ is measurable. Now let $\beta^*$ satisfy $\beta^*(t) \in \arg \min_{b \in B}[g^*(t)b + c(b)]$ for all $t \in T$ where $T$ is a measurable set of full Lebesgue measure. Then define $\hat{\beta}$ for all $t \in [0, T]$ by

\[ \hat{\beta}(t) = S(g^*(t))1_{\overline{T}}(t) + \beta^*(t)1_{\overline{T}}(t). \]
Then \( \hat{\beta}(t) \in \arg \min_{b \in B} [g^*(t)b + c(b)] \) for all \( t \in [0, T] \), \( \hat{\beta} \) is measurable and \( g^* \) is continuous, and \( \hat{\beta}(t) = \beta^*(t) \) for almost all \( t \in [0, T] \). 

**Proof of Proposition 4.** Let \( Q : [0, 1] \to \mathbb{R} \) satisfy \( Q(i) = v'(i)i - v(i) \). Note that for all \( t > \tau \) we have that \( Q(s^*_t) - \gamma \leq 0 \). This implies that

\[
(g^*)'(t) = -\gamma g^*(t) + s^*(t) [M(g^*(t)) - M(g^*(T)) - Q(i^*(t))] \\
\leq -\gamma g^*(t) + s^*(t) [M(g^*(t)) - M(g^*(T))] \\
< -\gamma g^*(t) + s^*(t) \beta^*(T)(g^*(t) - g^*(T)) \\
< g^*(t)[s^*(t)\delta - \gamma] \leq 0.
\]

This establishes that \( g^* \) is strictly decreasing on \( [\tau, T] \) and as \( \beta^* \) is non-increasing in \( g^* \) the result follows. 

**Proof of Corollary 5.** As \( 1 \geq R_0 = \frac{\delta}{\gamma} \) we have that \( \tau = \min \{ t \in [0, T] : \frac{\delta}{\gamma}s(t) \leq 1 \} = 0 \). Thus, Proposition 4 implies that \( g^* \) is strictly decreasing on \( [0, T] \) and \( \beta^* \) is non-decreasing on \( [0, T] \). Let \( t^* = \inf \{ t \in [0, T] : g(t) = \delta \} \wedge T \). We note that \( g^*(t) < \delta \) for \( t < t^* \) and \( g^*(t) > \delta \) for \( t > t^* \). Since \( c(b) = \delta(b - \bar{b}) \) it follows immediately from \( \beta^*(t) \in \arg \max_{b \in B} [g^*(t)b + c(b)] = \arg \max_{b \in B} [g^*(t) - \delta]b + \delta\bar{b} \) that \( \beta^*(t) = \bar{b} \) for \( t < t^* \) and \( \beta^*(t) = 0 \) for \( t > t^* \).

**Proof of Proposition 6.** Denote by \( \tau^\beta(q) = \inf \{ t : s^\beta(t) \leq q \} \) and let \( \tilde{i}^\beta(q) = i(\tau^\beta(q)) \) for all \( q \geq s^\beta(T) \) and define \( \tilde{i}^\beta \) analogously. The number of infected \( \tilde{i}^\beta \) as a function of the number of susceptible solves

\[
\tilde{i}^\beta(s) = i_0 + s_0 - s - \gamma \int_s^{s_0} \frac{1}{\beta(\tau^\beta(z))} dz
\]

for \( s \in [s^\beta(T), s_0] \). As \( \beta \) induces less social distancing than \( \hat{\beta} \) we have that \( \beta(\tau^\beta(z)) \geq \hat{\beta}(\tau^\beta(z)) \) for all \( z \geq \max \{ s^\beta(T), \bar{s}^\beta(T) \} \). It thus follows that \( \tilde{i}^\beta(z) \geq \hat{i}^\beta(z) \) for all \( z \geq \max \{ s^\beta(T), \bar{s}^\beta(T) \} \). This implies that any given share of susceptible population is reached faster with less social distancing, i.e. for all \( s \in [\max \{ s^\beta(T), \bar{s}^\beta(T) \}, s_0] \)

\[
\tau^\beta(s) = \int_s^{s_0} \frac{1}{\beta(\tau^\beta(z))\tilde{i}^\beta(z)} dz \leq \int_s^{s_0} \frac{1}{\hat{\beta}(\tau^\beta(z))\tilde{i}^\beta(z)} dz = \tau^\beta(s).
\]

This implies that \( s^\beta(t) \leq \bar{s}^\beta(t) \) for all \( t \in [0, T] \). This implies that herd immunity is achieved faster under a higher transmission rate \( \tau^\beta(\gamma/\tilde{\gamma}) < \tau^\beta(\gamma/\bar{\gamma}) \).

**Proof of Proposition 7.** Let \( Q : [0, 1] \to \mathbb{R} \) satisfy \( Q(i) = v'(i)i - v(i) \). As \( g \) solves

\[
(g^*)'(t) = -\gamma g^*(t) + s^*(t) [M(g^*(t)) - M(g^*(T)) - Q(i^*(t))]
\]

26 By convention inf \( \emptyset = \infty \).
and $M$ and $Q$ are continuous it follows that $(g^*)'$ is continuous. Moreover, we obtain from (34) that

\[
(g^*)'(t) = -\gamma g^*(t) + s^*(t) [g^*(t)\beta^*(t) + c(\beta^*(t)) - M(g^*(T)) - Q(i^*(t))]
\]

\[
= \frac{(i^*)'(t)}{i^*(t)} g^*(t) + s^*(t) [c(\beta^*(t)) - M(g^*(T)) - Q(i^*(t))].
\]

(35)

By the envelope theorem we get that $M'(g^*(t)) = \beta^*(t)$. We thus get by (34) that for a.e. $t \in [0,T]$

\[
(g^*)''(t) = -\gamma (g^*)'(t) + (s^*)'(t) [M(g^*(t)) - M(g^*(T)) - Q(i^*(t))]
\]

\[
+ s^*(t) [\beta^*(t)(g^*)'(t) - Q'(i^*(t))(i^*)'(t)].
\]

Using (34) once again we obtain for a.e. $t \in [0,T]$

\[
(g^*)''(t) = -\gamma (g^*)'(t) + \frac{(s^*)'(t)}{s^*(t)} [(g^*)'(t) + \gamma g^*(t)]
\]

\[
+ s^*(t) [\beta^*(t)(g^*)'(t) - Q'(i^*(t))(i^*)'(t)]
\]

\[
= [-\gamma + (s^*)'(t) s^*(t) + s^*(t)\beta^*(t)] (g^*)'(t) + \frac{\gamma g^*(t)(s^*)'(t)}{s^*(t)} s^*(t) Q'(i^*(t))(i^*)'(t)
\]

\[
= [-\gamma + \beta^*(t)(s^*(t) - i^*(t))] (g^*)'(t) - \gamma g^*(t)\beta^*(t)i^*(t) - s^*(t)Q'(i^*(t))(i^*)'(t).
\]

Next it follows from (35) that for a.e. $t \in [0,T]$

\[
(g^*)''(t) = [-\gamma + \beta^*(t)(s^*(t) - i^*(t)) - \frac{Q'(i^*(t))i^*(t)s^*(t)}{g^*(t)}] (g^*)'(t) - \gamma g^*(t)\beta^*(t)i^*(t)
\]

\[
+ \frac{(s^*)'(t)^2 i^*(t) Q'(i^*(t)) [c(\beta^*(t)) - M(g^*(T)) - Q(i^*(t))]}{g^*(t)}
\]

\[
= [-\gamma + \beta^*(t)(s^*(t) - i^*(t)) - \frac{Q'(i^*(t))i^*(t)s^*(t)}{g^*(t)}] (g^*)'(t)
\]

\[
+ g^*(t)\beta^*(t)i^*(t) \left[ \frac{(s^*(t))^2 i^*(t) Q'(i^*(t)) [c(\beta^*(t)) - M(g^*(T)) - Q(i^*(t))]}{\beta^*(t) (g^*(t))^2} \right] - \gamma.
\]

It follows from the fact that $Q'(i) = v''(i)i$ that for a.e. $t \in [0,T]$

\[
(g^*)''(t) = [-\gamma + \beta^*(t)(s^*(t) - i^*(t)) - \frac{v''(i^*)(i^*)^2 s^*(t)}{g^*(t)}] (g^*)'(t)
\]

\[
+ g^*(t)\beta^*(t)i^*(t) \left[ \frac{(s^*(t))^2 i^*(t) v''(i^*) [c(\beta^*(t)) - M(g^*(T)) - Q(i^*(t))]}{\beta^*(t) (g^*(t))^2} \right] - \gamma.
\]

(36)

Next note that the facts that $s^*(t) + i^*(t) \leq 1$ and for all $s \in [0,1] : s^2(1-s) \leq \frac{4}{27}$ imply that

\[
(s^*(t))^2 i^*(t) \leq (s^*(t))^2 (1-s^*(t)) \leq \frac{4}{27}.
\]

(37)
Moreover, note that convexity of \( v \) and concavity of \( M \) ensure that
\[
c(\beta^*(t)) - M(g^*(T)) - Q(i^*(t)) \leq c(\beta^*(t)) - M(g^*(T))
\]
\[
= M(g^*(t)) - M(g^*(T)) - g^*(t)\beta^*(t)
\]
\[
\leq \beta^*(T)(g^*(t) - g^*(T)) - \beta^*(t)g^*(t) \leq (\beta^*(T) - \beta^*(t))g^*(t).
\]
(38)

We next consider the two cases (i) \( g^*(t) < -c'(\bar{b}^-) \) and (ii) \( g^*(t) \geq -c'(\bar{b}^-) \).

*Case (i):* Note that convexity of \( c \) ensures that for all \( b \in [\underline{b}, \bar{b}] \) we have
\[
\frac{c(\bar{b}) - c(b)}{\bar{b} - b} \leq c'(\bar{b}^-).
\]
In the case \( g^*(t) < -c'(\bar{b}^-) \) we thus obtain for all \( b \in [\underline{b}, \bar{b}] \) that
\[
g^*(t)\bar{b} + c(\bar{b}) < g^*(t)b + c(b).
\]
This implies that \( \beta^*(t) = \bar{b} \). By Proposition 2 we have \( g^*(t) > g^*(T) \) and hence we obtain similarly that \( \beta^*(T) = \bar{b} \). It follows from (38) that in the case \( g^*(t) < -c'(\bar{b}^-) \) we have
\[
\frac{c(\beta^*(t)) - M(g^*(T)) - Q(i^*(t))}{\beta^*(t)(g^*(t))^2} \leq 0.
\]
(39)

*Case (ii):* In the case \( g^*(t) \geq -c'(\bar{b}^-) \) we have with (38) that\(^{27}\)
\[
\frac{c(\beta^*(t)) - M(g^*(T)) - Q(i^*(t))}{\beta^*(t)(g^*(t))^2} \leq \frac{\beta^*(T) - \beta^*(t)}{\beta^*(t)g^*(t)} \leq -\frac{1}{c'(\bar{b}^-)} \left( \frac{\bar{b}}{\bar{b}} - 1 \right).
\]
(40)

Combining (37), (39), (40) and (36) implies that in both cases (i) and (ii) we have that for a.e. \( t \in [0, T] \)
\[
(g^*')''(t) \leq \left[ -\gamma + \beta^*(t)(s^*(t) - i^*(t)) - \frac{v''(i^*(t))(i^*(t))^2s^*(t)}{g^*(t)} \right] (g^*)'(t)
\]
\[
- g^*(t)\beta^*(t)i^*(t) \left[ \frac{4v''(i^*(t))}{27c'(\bar{b}^-)} \left( \frac{\bar{b}}{2} - 1 \right) + \gamma \right].
\]
(41)

To show that \( g^* \) is strictly quasi concave consider a point \( t^* \in (0, T) \) such that \( (g^*)'(t^*) = 0 \). It follows from boundedness of \( \beta^*, i^* \) and \( s^* \), the fact that \( g^* > 0 \) and continuity of \( (g^*)' \) that
\[
\lim_{t \to t^*} \left[ -\gamma + \beta^*(t)(s^*(t) - i^*(t)) - \frac{v''(i^*(t))(i^*(t))^2s^*(t)}{g^*(t)} \right] (g^*)'(t) = 0.
\]
Moreover, \( \gamma \beta^* g^* i^* \) is bounded away from 0 on \([0, T]\). Therefore Assumption 1 that \( \frac{v''(t)}{t} < \frac{27k}{4(\bar{b} - \bar{b})}(-c'(\bar{b}^-)) \) and (41) ensure that there exist \( \epsilon, \delta > 0 \) such that \( (g^*)''(t) \leq -\delta \) for a.e. \( t \in (t^* - \epsilon, t^* + \epsilon) \). This implies for
\(^{27}\)Note that Assumption 1 implies that \( -c'(\bar{b}^-) > 0 \) as \( -c'(\bar{b}^-) = 0 \) otherwise would imply that \( v''(t) < 0 \) which contradicts the convexity of \( v \).
all \( t \in (t^* - \epsilon, t^*) \) that

\[
(g^*)'(t) = -((g^*)'(t^*) - (g^*)'(t)) = - \int_t^{t^*} (g^*)''(r)dr \geq \delta(t^* - t) > 0.
\]

Similarly, we get that \((g^*)'(t) < 0\) for all \( t \in (t^*, t^* + \epsilon) \). It follows that \( t^* \) is a strict local maximum. Thus every local extremum of \( g^* \) in \((0, T)\) is a strict local maximum and hence it follows that \( g^* \) is strictly quasi concave. In particular, \( g^* \) is single peaked. For the sequel let \( t^* \in [0, T] \) be such that \( g^* \) is strictly increasing on \([0, t^*]\) and strictly decreasing on \([t^*, T]\). It follows from Proposition 2 that \( \beta^* \) is quasi-convex.

We next show that \( i^* \) is strictly log-concave on \([0, t^*]\). Since \( \beta^* \) is non-increasing on \([0, t^*]\) and \( s^* \) is strictly decreasing it follows that \( \frac{\partial}{\partial t} \log(i^*(t)) = \frac{(i^*)'(t)}{i^*(t)} = \beta^*(t)s^*(t) - \gamma \) is strictly decreasing on \([0, t^*]\). Consequently, \( i^* \) is strictly log-concave on \([0, t^*]\) and thus has at most one local maximum on \([0, t^*]\).

**Proof of Proposition 8.** Let \( c(\beta) = \delta(\overline{b} - \beta) \). It follows from (16) that any optimal control satisfies

\[
\beta^*(t) = \begin{cases} \overline{b} & \text{if } g^*(t) < \delta \\ \underline{b} & \text{if } g^*(t) > \delta. \end{cases}
\]

As argued in the proof of Proposition 7, \( g^* \) is strictly quasi-concave. If \( g^*(t) < \delta \) for all \( t \in [0, T] \) we set \( t^*_1 = t^*_2 = 0 \). If there exists \( \overline{t} \in [0, T] \) with \( g(\overline{t}) \geq \delta \) we set \( t^*_1 = \inf\{t \in [0, T] : g^*(t) \geq \delta\} \) and \( t^*_2 = \sup\{t \in [0, T] : g^*(t) \geq \delta\} \). This together with (42) implies (13).

Next note that \( \beta^* \) is almost everywhere equal to a non-increasing function on the interval \([0, t^*_2]\). This implies that \( \frac{\partial}{\partial t} \log(i^*(t)) = s^*(t)b^*(t) - \gamma \) is strictly decreasing and thus that \( i^* \) is strictly log-concave and admits at most one local maximum on \([0, t^*_2]\). It follows similarly that as \( \beta^* \) is almost everywhere equal to a constant function on \((t^*_2, T]\) that \( i^* \) has at most one local maximum on \([t^*_2, T]\). In the case \( t^*_2 > 0 \) there can never be a local maximum at \( t^*_2 \) as the left derivative is strictly less than the right derivative.

We finally show that \( i^* \) is strictly increasing on \([0, t^*_1]\). First assume that \( g^*(T) \geq \delta \). Then we have \( g^*(t) > g^*(T) \geq \delta \) for all \( t \in [0, T] \) and hence \( t^*_1 = 0 \) and there is nothing to show. For the remainder of the proof assume that \( g^*(T) < \delta \). Then we have \( \beta^*(T) = \overline{b} \) and hence for all \( t \in [0, t^*_1) \)

\[
\frac{(g^*)'(t)}{g^*(t)} = -\gamma + \frac{s^*(t)}{g^*(t)}[M(g^*(t)) - M(g^*(T))] - Q(i^*(t)) \leq -\gamma + \frac{s^*(t)}{g^*(t)}[g^*(t) - g^*(T)] \overline{b} \leq \frac{\delta s^*(t)}{i^*(t)} - \gamma = \frac{(i^*)'(t)}{i^*(t)}.
\]

Observe, that as \( g^* \) is strictly increasing on \([0, t^*_1]\) it follows that \( i^* \) is strictly increasing on \([0, t^*_1]\).  

The next lemma presents a basic result about the Lambert \( W \) function that we need for the subsequent results.

---

\[\text{Note that if we have } g^*(t) < \delta \text{ for all } t \in [0, T] \text{ then by construction we have } t^*_1 = 0 \text{ and this claim clearly holds true.}\]
Lemma 18. Let \( s \in [0, 1) \) and \( \rho > 0 \). Then\(^{29}\) \( r = \rho W_0(-\frac{1}{\rho}e^{-\frac{s}{\rho}}) + 1 \) is the unique solution in \((0, 1)\) of the equation

\[
1 - r - se^{-\frac{r}{\rho}} = 0.
\] (43)

Proof. Note that (43) is equivalent to \( xe^x = \frac{-1}{\rho}e^{-\frac{1}{\rho}x} \) with \( x = \frac{W}{\rho} \). Since \( ze^{-z} \in [0, 1/e] \) for \( z \in [0, \infty) \) we have that \( -\frac{1}{\rho}e^{-\frac{1}{\rho}x} \in (-1/e, 0] \). There are thus two solutions to the equation \( xe^x = \frac{-1}{\rho}e^{-\frac{1}{\rho}x} \)

\[
\tau = W_{-1}\left(-\frac{1}{\rho}e^{-\frac{s}{\rho}}\right) < x = W_0\left(-\frac{1}{\rho}e^{-\frac{s}{\rho}}\right),
\]

where \( W_{-1} \) and \( W_0 \) denote the two real branches of the Lambert \( W \) function\(^{30}\). It follows that

\[
\tau = \rho W_{-1}\left(-\frac{1}{\rho}e^{-\frac{s}{\rho}}\right) + 1 < r = \rho W_0\left(-\frac{1}{\rho}e^{-\frac{s}{\rho}}\right) + 1
\]

are the only two solutions of (43).

If \( \rho > 1 \), then \( W_0\left(-\frac{1}{\rho}e^{-\frac{s}{\rho}}\right) = -\frac{1}{\rho} \) and hence the facts that \( s < 1 \) and that \( W_0 \) is increasing imply that

\[
r > \rho W_0\left(-\frac{1}{\rho}e^{-\frac{s}{\rho}}\right) + 1 = 0.
\]

If \( \rho \leq 1 \), then \( W_{-1}\left(-\frac{1}{\rho}e^{-\frac{s}{\rho}}\right) = -\frac{1}{\rho} \) and hence

\[
r > \rho W_0\left(-\frac{1}{\rho}e^{-\frac{s}{\rho}}\right) + 1 \geq \rho W_{-1}\left(-\frac{1}{\rho}e^{-\frac{s}{\rho}}\right) + 1 = 0.
\]

Moreover, we have \( r \leq \rho W_0(0) + 1 = 1 \). Consequently, we have \( r \in (0, 1] \).

To complete the proof we show that \( \tau < 0 \). If \( \rho \leq 1 \), then \( W_{-1}\left(-\frac{1}{\rho}e^{-\frac{s}{\rho}}\right) = -\frac{1}{\rho} \) and hence the facts that \( s < 1 \) and that \( W_{-1} \) is decreasing imply that

\[
\tau < \rho W_{-1}\left(-\frac{1}{\rho}e^{-\frac{s}{\rho}}\right) + 1 = 0.
\]

If \( \rho > 1 \), then \( W_0\left(-\frac{1}{\rho}e^{-\frac{s}{\rho}}\right) = -\frac{1}{\rho} \) and hence

\[
\tau < \rho W_{-1}\left(-\frac{1}{\rho}e^{-\frac{s}{\rho}}\right) + 1 \leq \rho W_0\left(-\frac{1}{\rho}e^{-\frac{s}{\rho}}\right) + 1 = 0.
\]

This completes the proof. \( \Box \)

We next consider a variant of the objective (4) with linear cost function \( v \) in the number of infected and vanishing cost function \( c \) (i.e., a reduction of the transmission rate does not incur costs) and infinite time

\(^{29}\)\(W_0: [-1/e, \infty) \to \mathbb{R}\) denotes the principal branch of the Lambert \( W \) function (also known as product logarithm).

\(^{30}\)\(W_0\) is characterized as follows: For each \( z \in [-1/e, \infty) \) the number \( W_0(z) \) is the unique \( w \in [-1, \infty) \) such that \( we^w = z \). \( W_{-1} \) is characterized as follows: For each \( z \in [-1/e, 0) \) the number \( W_{-1}(z) \) is the unique \( w \in (-\infty, -1] \) such that \( we^w = z \). In particular, \( W_{-1}(z) \leq W_0(z) \) for all \( z \in [-1/e, 0) \).
horizon $T = \infty$. That is, we seek to minimize

$$\mathcal{J}(\beta) = \nu \int_0^\infty i(t) dt$$

for some $\nu > 0$. Note that $\mathcal{J}$ is proportional to the long-run total number of people getting infected throughout the epidemic (which is equal to the number of people that recover or die from the disease). Indeed, the third compartment $r$ of the SIR model satisfies $r'(t) = \gamma i(t)$ and as $r(0) = 0$ we obtain

$$\mathcal{J}(\beta) = \nu \gamma \int_0^\infty r'(t) dt = \nu \gamma \left( \lim_{t \to \infty} r(t) \right).$$

Hence, minimizing $J$ is equivalent to minimizing $\lim_{t \to \infty} r(t)$. Note that this limit always exists since $r$ is non-decreasing and bounded from above by $1$.

We next restrict the set of admissible controls $\mathcal{B}$. To this end, we fix a time horizon $t_2 > 0$ and only consider strategies $\beta$ with $\beta(t) = \tilde{\beta}$ for all $t \geq t_2$. This means that the planner is not allowed to enact any measures reducing the transmission rate after time $t_2$. Prior to time $t_2$ the planner is allowed to enact maximal measures on a time interval $(t_1, t_2)$, where the time $t_1$ can be chosen by the planner. Mathematically, the strategies hence take the form $\beta(t) = \tilde{\beta}$ for all $t \in (t_1, t_2)$ and $\beta(t) = \tilde{\beta}$ for all $t \in [0, t_1] \cup [t_2, \infty)$ \(^{31}\). The planner hence only optimizes the beginning $t_1 \in [0, t_2]$ of SD measures in order to minimize $\mathcal{J}$. We denote such strategies by $\beta_{t_1, t_2}$. For given $t_2 > 0$ the set of all $\beta_{t_1, t_2}$, $t_1 \in [0, t_2]$, is denoted by $\mathcal{B}_{t_2}$. To highlight the dependence of the state variables $s$, $i$ and $r$ associated to $\beta_{t_1, t_2} \in \mathcal{B}_{t_2}$ we sometimes write in the sequel $s_{t_1, t_2}, i_{t_1, t_2}, r_{t_1, t_2} : [0, \infty) \to [0, 1]$.

In the sequel we provide sufficient conditions that ensure that enacting SD over the maximal time interval $[0, t_2]$ is not optimal to minimize the long-run total number of infected.

**Proposition 19.** Let $t_2 > 0$. Minimizing $\mathcal{J}$ over $\mathcal{B}_{t_2}$ is equivalent to maximizing the number of people that recover or die during the SD period. More precisely, it holds for $t_1, \tilde{t}_1 \in [0, t_2]$ that $\mathcal{J}(\beta_{t_1, t_2}) < \mathcal{J}(\beta_{\tilde{t}_1, t_2})$ if and only if $r_{t_1, t_2}(t_2) - r_{t_1, t_2}(t_1) > r_{\tilde{t}_1, t_2}(t_2) - r_{\tilde{t}_1, t_2}(\tilde{t}_1)$.

**Proof.** We show that $\mathcal{J}(\beta_{t_1, t_2})$ is a decreasing function in $r_{t_1, t_2}(t_2) - r_{t_1, t_2}(t_1)$. To simplify notation we fix $t_1 \in [0, t_2]$ and omit the index $(t_1, t_2)$ in this proof. Recall that by (45) minimizing $\mathcal{J}$ is equivalent to minimizing the long-time limit $r = \lim_{t \to \infty} r(t)$. We first provide a characterization of $r$. Since the fraction of susceptibles satisfies $s'(t) = -\tilde{\beta} i(t) s(t)$ for all $t \geq t_2$, we obtain that

$$s(t) = s(t_2) e^{-\tilde{\beta} \int_{t_2}^{t} i(u) du} = s(t_2) e^{-\frac{\tilde{\beta}}{\gamma} \int_{t_2}^{t} r'(u) du} = s(t_2) e^{-\frac{\tilde{\beta}}{\gamma} (r(t) - r(t_2))}.$$  \hspace{1cm} (46)

This and the fact that $s(t) + i(t) + r(t) = 1$ for all $t \geq 0$ implies that

$$r'(t) = \gamma (1 - r(t) - s(t)) = \gamma \left(1 - r(t) - s(t_2) e^{-\frac{\tilde{\beta}}{\gamma} (r(t) - r(t_2))} \right)$$

for all $t \geq t_2$. Consequently, the long-time limit of $r = \lim_{t \to \infty} r(t)$ satisfies

$$1 - r - s(t_2) e^{-\frac{\tilde{\beta}}{\gamma} (r(r(t_2))} = 0.$$

\(^{31}\)These strategies thus coincide with the interval strategies defined in (13) in the case $T = \infty$.
By Lemma 18 the unique solution $r \in (0, 1]$ is given by
\[
r = \frac{\gamma}{b} W_0 \left( -\frac{b}{\gamma} e^{-\frac{b}{\gamma} s(t_2)e^{\frac{b}{\gamma} r(t_2)}} \right) + 1.
\] (47)

Since $W_0$ is increasing it follows that $r$ is minimized if $s(t_2)e^{\frac{b}{\gamma} r(t_2)}$ is maximized.

Next, we take a strategy $b \in B_{t_2}$ which is characterized by the beginning $t_1$ of SD measures. Similarly to (46) we obtain that
\[
s(t_2) = s(t_1)e^{-\frac{b}{\gamma} (r(t_2) - r(t_1))}
\] and
\[
s(t_1) = s(0)e^{-\frac{b}{\gamma} (r(t_1) - r(0))}.
\]
This implies that
\[
s(t_2)e^{\frac{b}{\gamma} r(t_2)} = s(0)e^{\frac{b}{\gamma} r(0)} e^{\left(\frac{b}{\gamma} - \frac{b}{\gamma}\right)(r(t_2) - r(t_1))}.
\]
This together with (47) and (45) implies that $J$ is a decreasing function in $r(t_2) - r(t_1)$. This completes the proof.

**Proof of Proposition 9.** By the assumption of the proposition
\[
s(0)b > e^{b/\gamma}, \quad \text{and} \quad t_2(1 - s(0)) [b s(0) e^{-b/\gamma} - \gamma] > 1 - s(0) e^{-b/\gamma}.
\] (48)

Since $t_2 > 0$ is fixed in this result, we drop $t_2$ from the index in the sequel. By Proposition 19 we have to show that there exists $t_1^* > 0$ such that
\[
r_{t_1}(t_2) - r_{t_1}(t_1) > r_0(t_2) - r(0)
\]
for all $t_1 \in (0, t_1^*]$. To this end introduce for all $t_1 \in [0, t_2]$ the function $\tau_{t_1} : [0, 1] \to [0, 1]$ satisfying
\[
\tau_{t_1}(t) = r_{t_1}((t_2 - t_1)t + t_1) - r_{t_1}(t_1)
\]
for all $t \in [0, 1]$. We thus have to show that there exists $t_1^* > 0$ such that
\[
\tau_{t_1}(1) > \tau_{t_0}(1)
\]
for all $t_1 \in (0, t_1^*]$. Note that for all $t_1 \in [0, t_2]$ it holds that $r_{t_1}(0) = 0$ and
\[
\tau_{t_1}'(t) = (t_2 - t_1)r_{t_1}'((t_2 - t_1)t + t_1) = \gamma(t_2 - t_1) r_{t_1}'((t_2 - t_1)t + t_1)
\]
\[
= \gamma(t_2 - t_1) (1 - r_{t_1}((t_2 - t_1)t + t_1) - s_{t_1}((t_2 - t_1)t + t_1))
\]
\[
= \gamma(t_2 - t_1) (1 - r_{t_1}(t_1) - \tau_{t_1}(t) - s_{t_1}((t_2 - t_1)t + t_1)), \quad t \in [0, 1].
\] (49)
Moreover, note that for all $t_1 \in [0, t_2], t \in [t_1, t_2]$ we have

$$s_{t_1}(t) = s(0)e^{-\int_0^t \beta_{t_1}(s) \gamma_{t_1}(s) ds} = s(0)e^{-\int_0^t \overline{b}_{t_1}(s) ds - \int_0^t \gamma_{t_1}(s) ds} = s(0)e^{-\overline{b}_t r_{t_1}(s) ds - \gamma r_{t_1}(s) ds} = s(0)e^{-\overline{b}_t r_{t_1}(s) ds - \gamma r_{t_1}(s) ds}.$$ 

This together with (49) implies that for all $t_1 \in [0, t_2], t \in [0, 1]$ we have $\tau_{t_1}(t) = f(t_1, \overline{\tau}_{t_1}(t))$, where

$$f(t_1, r) = \gamma(t_2 - t_1) \left(1 - r_{t_1}(t_1) - r - s(0)e^{-\overline{b}_t (r_{t_1}(t_1) - r(0))/\gamma e^{-br/\gamma}}\right)$$

for $t_1 \in [0, t_2], r \in [0, 1]$. Hence, the fact that for every $t_1 \in [0, t_2]$, the function $\tau_{t_1} : [0, 1] \rightarrow [0, 1]$ is strictly increasing and satisfies the initial condition $\tau_{t_1}(0) = 0$ implies that

$$\int_0^{\tau_{t_1}(1)} \frac{1}{f(t_1, z)} dz = \int_0^1 \frac{\tau_{t_1}'(t)}{f(t_1, \tau_{t_1}(t))} dt = 1. \quad (50)$$

Next note that for every $t_1 \in [0, t_2]$ the strategies $\beta_{t_1}$ and $\beta_{t_2}$ are both equal to $\overline{b}$ on $[0, t_1]$ and hence for all $s \in [0, t_1]$ we have $i_{t_1}(s) = i_{t_2}(s)$, This implies that

$$\frac{d}{dt_1} r_{t_1}(t_1) = \frac{d}{dt_1} \left(r(0) + \gamma \int_0^{t_1} i_{t_1}(s) ds\right) = \frac{d}{dt_1} \left(r(0) + \gamma \int_0^{t_1} i_{t_2}(s) ds\right) = \gamma i_{t_2}(t_1) = \gamma i_{t_1}(t_1).$$

This implies for all $t_1 \in [0, t_2], r \in [0, 1]$ that

$$\frac{d}{dt_1} f(t_1, r) \bigg|_{t_1=0} = -\gamma \left(1 - r_{t_1}(t_1) - r - s(0)e^{-\overline{b}_t (r_{t_1}(t_1) - r(0))/\gamma e^{-br/\gamma}}\right)$$

$$+ \gamma(t_2 - t_1)i_{t_1}(t_1) \left(\overline{b}s(0)e^{-\overline{b}_t (r_{t_1}(t_1) - r(0))/\gamma e^{-br/\gamma}} - \gamma\right).$$

Using $r(0) = 0$ and the assumptions (48) we hence obtain for all $t_1 \in [0, t_2], r \in [0, 1]$ that

$$\frac{d}{dt_1} f(t_1, r) \bigg|_{t_1=0} = -\gamma \left(1 - r - s(0)e^{-br/\gamma}\right) + \gamma t_2(1 - s(0)) \left(\overline{b}s(0)e^{-br/\gamma} - \gamma\right)$$

$$\geq -\gamma \left(1 - s(0)e^{-br/\gamma}\right) + \gamma t_2(1 - s(0)) \left(\overline{b}s(0)e^{-br/\gamma} - \gamma\right)$$

$$> -\gamma \left(1 - s(0)e^{-br/\gamma}\right) + \gamma \left(1 - s(0)e^{-br/\gamma}\right) = 0.$$  

From this we obtain that there exists $t^*_1 > 0$ such that for all $t_1 \in (0, t^*_1]$ we have

$$\int_0^{\tau_{t_1}(1)} \frac{1}{f(t_1, z)} dz < \int_0^{\tau_0(1)} \frac{1}{f(0, z)} dz$$

and $f(t_1, z) > 0$ for all $z \in [0, \tau_0(1)]$. This together with (50) implies for all $t_1 \in (0, t^*_1]$ that

$$\int_0^{\tau_{t_1}(1)} \frac{1}{f(t_1, z)} dz = 1 > \int_0^{\tau_0(1)} \frac{1}{f(t_1, z)} dz$$

and hence $\tau_{t_1}(1) > \tau_0(1)$. This completes the proof. 

\[\square\]
References


Dewatripont, Mathias, Michel Goldman, Eric Muraille, and Jean-Philippe Platteau, “Rapidly identifying workers who are immune to COVID-19 and virus-free is a priority for restarting the economy,” VOX CEPR Policy Portal, 2020, 23.


Liu, Y and A Gayle, “At. 2020. Reproductive Number of C.-19 is Higher Compared to S. Coronavirus, Wilder-Smith, A., & Rocklöv, J. The reproductive number of COVID-19 is higher compared to SARS coronavirus,” Journal of Travel Medicine, 2020, 27 (2).


