Optimally Controlling an Epidemic

Martín Gonzalez-Eiras* Dirk Niepelt†

October 14, 2020

Abstract

We propose a flexible model of infectious dynamics with a single endogenous state variable and economic choices by individuals and a government. We analytically characterize equilibrium and optimal outcomes as well as static and dynamic externalities and we calibrate and simulate the model to inform about the ongoing COVID-19 pandemic. We find the following: (i) A lockdown is followed by its opposite—policies to stimulate activity beyond the privately optimal level. (ii) Social distancing has small welfare gains when governments lack instruments to stimulate activity. (iii) Future expected epidemiological changes substantially affect current policy. (iv) Re-infection risk may imply a more cautious steady state than optimal. (v) When a cure or vaccination arrives deterministically, optimal policy may be dis-continous, featuring a light lockdown when the arrival date exceeds a specific value, and a strict one otherwise.

JEL codes: I18

Keywords: Epidemic, lockdown, forced opening, SIR model, SIS model, SI model, logistic model, COVID-19.

1 Introduction

An epidemic generates economic tradeoffs. Since economic activity is associated with interaction among agents higher activity is conducive to a faster spreading of infections, with effects on public health, mortality, and future economic outcomes. The ongoing COVID-19 pandemic offers a prime example.[1] Effective instruments to manage the tradeoff include

*University of Copenhagen. Oster Farimagsgade 5, 1353 Copenhagen K, Denmark. E-mail: mge@alum.mit.edu

†Study Center Gerzensee; University of Bern; CEPR; CESifo. P.O. Box 21, CH-3115 Gerzensee, Switzerland. E-mail: dirk.niepelt@szgerzensee.ch. Web: www.niepelt.ch.

[1]It is sometimes argued that there is no conflict between "lives and livelihoods" in the context of COVID-19 because some countries better kept economic and health costs in check than others. This observation does not speak to the existence of tradeoffs. It might speak to the fact that the quality of government responses varied.

test-and-trace-and-quarantine schemes or mass vaccinations of the susceptible population. But these instruments only become available with delay, if they become available at all, and this leaves forced, non-targeted reductions in economic activity—lockdowns—as the primary means for governments to respond to an epidemic in its early phase.

In this paper, we propose a generic framework to analyze optimal lockdown policies and we derive a series of results that seem to have gone unnoticed in the recently burgeoning literature focusing on the intersection of epidemiology and economics.[2] Almost all models in that recent literature build on the classical epidemiological "SIR"-model developed by Kermack and McKendrick (1927).[3] But this model is just one among several "compartmental" epidemiological workhorse models, and certainly not the most relevant one when re-infection risk or other factors are of primary concern.[4] In addition, SIR-models of various flavours feature two endogenous epidemiological state variables; this makes it difficult to embed economic choices in those frameworks without sacrificing analytical tractability, transparency, and generality.

The generic epidemiological framework that serves as the basis for our economic analysis generalizes the "simple epidemiological model" (Bailey, 1975) and captures the essence of various models of infectious dynamics: These dynamics mainly depend on the population shares of two groups, those that have contracted the disease and those that have not yet but are susceptible. New infections are driven by the complementarity between the two groups, and cumulative infections therefore approximately follow a logistic law of motion. We show that a framework with a single endogenous state variable, the share of the population that has contracted the disease in the past, constitutes a tractable and accurate approximation of infectious dynamics in models with additional state variables. Moreover, it offers a flexible setting to analyze how economic choices and epidemiological dynamics interact and how optimal policy shapes them.[5]

We superimpose an economic layer with households and a government on this epidemiological base. We assume that households derive utility from their individual choices of economic activity, both positive because activity generates consumption and negative because activity involves effort. In addition, individual activity choices increase the risk of getting infected which is privately costly, for instance because an infection increases mortality risk or causes other direct and indirect negative consequences. Households are fully aware of the aggregate infection dynamics and the risks they face and they behave individually rationally.[6]

[2]Early papers in that recent literature are Atkeson (2020) and Eichenbaum et al. (2020). Within a couple of weeks many papers by other authors have followed, see for instance CEPR's *Covid Economics: Vetted and Real-Time Papers* series.

[3]There has been some effort to improve the modeling of the matching process that leads to infections in the SIR model. For example, Bisin and Moro (2020) take spatial aspects into account; they show that local interactions give rise to matching frictions and local herd immunity effects.

[4]Hethcote (1989; 2000) reviews workhorse epidemiological models.

[5]Focusing on the population shares of those that have contracted the disease in the past, and those that have not yet, our model does not explicitly distinguish between currently and previously infected individuals. This comes at a small cost as we discuss in section 2. It does not restrict the model's ability to represent societal costs of new infections or deaths.

[6]Goolsbee and Syverson (2020) document voluntary behavioral changes in the U.S. before states

Nevertheless, the privately optimal activity choices fail to fully internalize the social consequences, and this gives rise to "static" and "dynamic" externalities. The static externality arises because the private marginal costs of infection differ from the immediate social marginal costs, for example because health insurance reduces the marginal costs of a hospital visit that is borne by the patient. The dynamic externality arises because an individual household has a negligible effect on aggregate infection dynamics. In equilibrium, households therefore do not internalize that higher activity choices give rise to a faster change of the aggregate state variable with corresponding welfare consequences. While the private sector always ends up bearing the full social costs of infection (as opposed to the marginal costs) in equilibrium the static and dynamic externalities imply that the activity choices are distorted.

Specifically, the equilibrium activity level evolves inversely with new infections since household choices only reflect the economic consequences of activity as well as their private share of the immediate social costs of infection. At the outset of an epidemic, households therefore only take minor precautions but they drastically reduce activity when infections peak. In contrast, the activity level preferred by the benevolent government also reflects the dynamic welfare consequences of activity due to induced changes of the state variable.

The optimal government intervention in this environment aims at correcting the static and dynamic externalities. For most of the analysis we assume that the government has sufficient instruments at its disposal to effectively control household activity levels, that is, the government solves the social planner program. Depending on static and dynamic externalities the government therefore imposes a lockdown—it curtails activity below the privately optimal level (e.g., through stay-at-home-orders or non-essential business closures). Since the infection rate is increasing in the aggregate activity level a lockdown slows down infection dynamics and shifts peak infections into the future.

That infection dynamics give rise to externalities and call for corrective government action is well understood (Gersovitz and Hammer, 2004). More surprisingly, we find that this corrective action can go both ways. One of our results characterizes the optimal timing not only of lockdowns but also of "inverse lockdowns," namely interventions that aim at fostering or stimulating private sector activity. Arguably, such inverse lockdowns have widely been imposed in the context of the COVID-19 pandemic, for instance in the form of monetary easing, temporary sales tax reductions, or employment subsidies.[7] We find that inverse lockdowns are imposed even if individual households fail to fully internalize the immediate social marginal costs of activity, that is, even if the static externality is negative. The reason is that eventually, as the infection spreads, the dynamic externality becomes positive. As long as the static externality remains limited in size the total externality therefore becomes positive as well and the optimal government intervention aims at fostering activity.

Underlying the positive dynamic externality are capital gains from higher cumulative

introduced lockdown measures.

[7]Other inverse lockdown measures were planned but have not (yet) been implemented. For example, in the U.S. the director of the White House National Economic Council urged lawmakers in June 2020 to replace a USD 600-a-week lump-sum transfer with a "return-to-work bonus" (*The Wall Street Journal*, June 15, 2020).

infection rates. Early in the infection process, the value function of society is decreasing: Rising cumulative infection rates lower the value because they imply that peak infection rates and the associated health costs and reductions in economic activity are approaching quickly, pushing the present value of future suffering up. But after a point, the slope of the value function becomes positive because a higher cumulative infection rate means that society faces fewer infections in the future. This positive consequence of new infections at higher rates of cumulative infections is not internalized by the private sector and provides a rationale for the government to stimulate activity.

Calibrating our model to match COVID-19 infection data we estimate that the welfare effects of correcting the externalities are substantial. In our baseline scenario the pandemic induced welfare losses amount to roughly twenty five percent of lifetime consumption under laissez faire, but only twenty percent under the optimal policy. When activity (reduction) enters quadratically rather than linearly in the law of motion for infections then the welfare gain from optimally rather than not intervening increases. On the other hand when the government can only impose a lockdown but not an inverse lockdown, then the benefits of government intervention relative to laissez faire are much more subdued.

To capture virus mutations, medical enhancements, or improvements in the test-and-trace technology which change the infection rate conditional on the state we allow for stochastic changes of model parameters. We find that upside risk—the prospect of more easily manageable disease dynamics in the future—leads the government to more tightly, and for longer, restrict activity than in the baseline scenario.

When infection and subsequent recovery does not confer permanent immunity the calculus of optimal government intervention changes. For example, in the context of the COVID-19 pandemic medical scientists currently cannot rule out that individuals with antibodies may re-contract the disease.[8] We model this scenario under the assumption that a constant share of the population that had been infected in the past becomes susceptible again such that individuals may repeatedly get infected. With such re-infection risk, the economy converges to a steady state with strictly positive rather than zero new infections—the disease becomes endemic. Moreover, the steady state depends on policy.

We prove that, absent a static externality, the government's choice of activity level necessarily exceeds the activity level in equilibrium and the government's choice goes hand in hand with a higher cumulative infection rate than in equilibrium. When the static externality is sufficiently large or society cares little about the future the opposite result holds true.

Another key result of our analysis concerns the optimal policy when a cure or vaccine is expected to arrive deterministically (such that time constitutes a second state variable). In the context of the COVID-19 pandemic, for instance, scientists and policy makers seem to converge (at the time of this writing, in fall 2020) to the view that a vaccine will become widely available between the end of 2020 and summer 2021. Also, poorer countries are unlikely to be among the first recipients of a newly developed drug and instead might gain access with a delay at a pre-specified date.

We prove that in such a scenario the exact date at which the vaccine becomes available

[8]See, for example, Gudbjartsson et al. (2020) and To et al. (2020).

makes a big difference in the sense that the strictness of an optimal lockdown may be discontinuous in the sense that the optimal lockdown is light when the arrival date exceeds a specific value, and very strict otherwise. Intuitively, stricter restrictions on activity delay peak infections but further delay becomes increasingly costly as it requires an even more pronounced and extended lockdown. After a point, the costs of delaying peak infections until after the arrival date of the cure or vaccine becomes too high and the optimal policy refrains from strongly suppressing activity levels.

Related Literature Workhorse epidemiological models due to Kermack and McKendrick (1927) and Bailey (1975) are reviewed, e.g., in Hethcote (1989) and Hethcote (2000). To derive our generic epidemiological framework and to calibrate the model we rely on this literature; the mapping between different frameworks laid out in Gonzalez-Eiras and Niepelt (2020b); as well as on Atkeson (2020), Ferguson et al. (2020), Greenstone and Nigam (2020), Hall et al. (2020), and Russell et al. (2020).

Since mid March 2020 there has been an explosion of papers focusing on the intersection of epidemiological dynamics and economic cost-benefit analysis. Early contributions include Atkeson (2020) and Eichenbaum et al. (2020). Alvarez et al. (2020) compute the optimal lockdown policy when there is a rationale to flatten the infection curve in order to relax health care system capacity constraints. Farboodi et al. (2020) argue that in equilibrium and under the optimal policy the effective reproduction number always remains close to unity. Gersovitz and Hammer (2004), Bethune and Korinek (2020) and Jones et al. (2020) assess externalities. Kaplan et al. (2020) and Acemoglu et al. (2020) analyze the implications of heterogeneity (see the discussion below). Çenesiz and Guimaraẽs (2020) and Giannitsarou et al. (2020) analyze immunity loss and demographic dynamics.

Unlike our work these papers mostly focus on numerical analyses, with Toxvaerd (2020), Gonzalez-Eiras and Niepelt (2020a), Abel and Panageas (2020), and Miclo et al. (2020) constituting some notable exceptions.[9] Moreover, our paper distinguishes itself in several dimensions. First, we provide a flexible and tractable framework that nests many of the existing models or is closely related to them. Second, we provide several analytical results, unlike most contributions. Third, our transparent results identify key factors such as the capital gains due to infections, and they make clear that many results in the burgeoning literature may not be robust to minor changes in the posited epidemiological environment. Fourth, we consider a series of scenarios and comparisons: Laissez faire vs. optimal policy; lockdowns vs. forced openings; stochastic vs. deterministic arrival of a vaccine/cure; permanent vs. temporary immunity, generating either a disease free steady state or an endemic equilibrium; and, a stationary epidemiological environment vs. environments with changing characteristics, generating a motive for "precautionary" policy interventions.

[9]Toxvaerd (2020) who characterizes privately optimal social distancing; Gonzalez-Eiras and Niepelt (2020a) who characterize optimal lockdown policies; and Abel and Panageas (2020) who characterize the optimal steady state in a SIR model with vital dynamics. Miclo et al. (2020) derive the optimal policy under a capacity constraint.

Outline The remainder of the paper is organized as follows. We lay out the model in section 2 and present the conditions characterizing equilibrium and first best in section 3. The main analysis is contained in section 4. Section 5 concludes. Appendix A contains a discussion of the connections between different epidemiological models that we exploit to calibrate the model. All proofs are relegated to appendix B.

2 The Model

We consider an infinite horizon economy with households and a government. Time is continuous and indexed by $t \geq 0$.

2.1 Epidemiology

We adopt an epidemiological framework that is closely related to several canonical models in the epidemiological literature: The SIR model due to Kermack and McKendrick (1927), a modified SIR model and the simple epidemic model, the SI model, due to Bailey (1975), and the SIS model derived from it.[10] Our framework incorporates one endogenous state variable (rather than the two in the typical SIR model), possibly time as a second state variable (unlike SIR and SIS models), as well as economic activity (unlike SIR and SIS models).

2.1.1 Dynamics

At time t the population consists of $x(t)$ "pre-infection" (for short: "pre") households; $1 - \bar{y}$ "neutral" households; and $y(t) = \bar{y} - x(t)$ "post-infection" ("post") households. Members of the post group have been infected in the past, members of the pre group might be infected in the future, and members of the neutral group cannot be infected, for instance because they are immune. We allow for members of the post group to return to the pre group; this captures the fact that households might repeatedly be infected because they do not permanently develop immunity. Since we normalize the mass of the total population to unity, the mass of each population group also represents the group's population share,

$$0 \leq y(t) \leq \bar{y} \leq 1.$$

The initial population shares of the pre and post groups are given by $x(0) = \bar{y} - y(0) > 0$ and $y(0) > 0$.

While the health status of neutral households never changes post households transmit the disease to members of the pre group according to a logistic law of motion. Moreover, as mentioned before, some post households may return to the pool of pre households. As a consequence, the share of post households evolves according to the law of motion

$$\dot{y}(t) = g(a(t))\,\beta\,y(t)\,(\bar{y} - y(t)) - \gamma y(t), \tag{1}$$

[10] The "S," "I," and "R" in SIR, SI, and SIS stands for "susceptible," "infectious," and "removed," respectively. See Hethcote (1989) and Hethcote (2000) for an overview over epidemiological models of infectious diseases.

where a dot (as in $\dot{y}(t)$) denotes the time derivative. By definition, $\dot{x}(t) = -\dot{y}(t)$.

Variable $a(t)$ in the law of motion denotes an index of economic activity. We assume that higher activity fosters infections, that is function g is increasing and smooth. Parameters $\beta > 0$ and $\gamma \geq 0$ capture the biological characteristics of the disease. According to equation (1), the number of pre households that get infected depends on their number, $x(t) = \bar{y} - y(t)$; the infection rate, $g(a(t))\beta$; and the number of post households, $y(t)$. The total increase of post households equals the number of newly infected post households net of the share γ of post households who return to the pre group. When $\gamma = 0$, post households can never again be reinfected.

When $\gamma = 0$, the law of motion implies that for any $g(a(t))\beta > 0$ and as long as $y(0) > 0$, as we assume, the share of the post group is strictly increasing over time and converges to $\bar{y}$; conversely, the share of the pre group is strictly decreasing and converges to 0 in this case. When $\gamma > 0$ the dynamics need not be monotone unless $a(t)$ is constant.

When $a(t) = a$, equation (1) has the solution[11]

$$y(t) = \frac{\bar{y} - \frac{\gamma}{g(a)\beta}}{1 + \mathrm{e}^{-(g(a)\beta\bar{y}-\gamma)t}\left(\frac{\bar{y}}{y(0)} - 1 - \frac{\gamma}{g(a)\beta y(0)}\right)} \tag{2}$$

and $y(t)$ converges to zero or to $s(a) \equiv \bar{y} - \gamma/(g(a)\beta)$, depending on whether $g(a)\beta\bar{y}$ is smaller or larger than γ. Throughout, we consider the latter case; that is, we focus on the case where the long-run activity level a satisfies $s(a) > 0$. For future reference we note that solving equation (2) for t yields

$$t(y_0, y) = \ln\left(\frac{y(s(a) - y_0)}{y_0(s(a) - y)}\right) / (g(a)\beta\bar{y} - \gamma), \quad y_0 < y < s(a). \tag{3}$$

We assume that with Poisson arrival rate ν the disease and its consequences (described below) disappear and β drops to zero. This might be due, for example, to medical progress or the development of a vaccine. We also allow for the possibility that the disease deterministically disappears in finite time, at the future date T.

2.1.2 Costs of Infection

Infections impose costs, for example because the health care system requires resources, output is lost, utility foregone, or persons die. We represent such costs as a function of the number of transitions from the pre to the post group, that is, as a function of $\dot{y}(t) + \gamma y(t)$. We assume that the direct and indirect social costs associated with these transitions equals

$$\psi\, g(a(t))\, \beta\, y(t)\, (\bar{y} - y(t)), \tag{4}$$

where $\psi > 0$ may reflect expected health care costs per newly infected person, the statistical value of life, fatality rates, etc.[12] To save on notation, we sometimes use the

[11] We assume that $g(a)\beta\bar{y} \neq \gamma$. See for example Hethcote (1989) for the case of $g(a) = \bar{y} = 1$.

[12] Of course, the analysis could be generalized by letting ψ depend on the state, $\psi(y(t))$. In the interest of sharpening our results we opt for simplicity. Our analysis shows that constant costs per unit of new infections are sufficient to generate a motive for slowing down infections.

definition

$$\dot{y}^g(t) \equiv g(a(t))\,\beta\,y(t)\,(\bar{y} - y(t)) = \dot{y}(t) + \gamma y(t).$$

The superscript "g" denotes "gross" and indicates that the costs of infection are proportional to the gross flow from the pre to the post state rather than the net flow $\dot{y}(t)$.

2.1.3 Relation to SIR and SIS Models

The law of motion (1) nests a range of well known epidemiological models, augmented by an effect of economic activity on the infection rate (see also the discussion in appendix A). For example, under the restrictions $\gamma = 0$ and $\bar{y} = 1$ it corresponds to the SI model (Bailey, 1975) in which the number of transitions from the pre to the post group first increases and then decreases as $y(t)$ moves from near zero towards unity.

The law of motion subject to $\gamma = 0$ also corresponds to a special case of the canonical SIR model (Kermack and McKendrick, 1927) and the modified SIR model (Bailey, 1975). In their general form, both SIR models characterize the evolution of three population groups—susceptible, currently infected, and removed households—and therefore contain two endogenous state variables (the shares of two of the three groups). Susceptible households are infected by currently infected households, they remain infectious for a random time span, and eventually they join the pool of removed (recovered or deceased) households. These dynamics reduce to the law of motion (1) with $\gamma = 0$ when the currently infected and removed households are combined into a single group of post households.[13]

Importantly, blurring the distinction between currently infected and removed households does not undermine the model's capacity to represent societal costs of infection or death, for representing these costs does not require to explicitly account for the *stock* of currently infected households or the deceased population. It rather suffices to account for the *flow* of infections, that is the flow from the pre to the post infection state, and to associate costs with this flow.[14] In Gonzalez-Eiras and Niepelt (2020b) we offer a detailed discussion of the connection between SIR models and the law of motion (1). We establish theoretically and based on numerical examples that the law of motion allows to flexibly capture epidemiological dynamics very much in line with those in conventional SIR models.

While the canonical SIR model (Kermack and McKendrick, 1927) and the modified SIR model (Bailey, 1975) yield similar predictions for transition dynamics they differ with respect to their implications for the long-run share of the population that never gets infected; in the modified SIR model this share always equals zero. A hybrid model augments the modified SIR model with an additional parameter that allows to regulate the long-run population shares (Gonzalez-Eiras and Niepelt, 2020b). In the law of motion (1)

[13]Formally, the transition rate from currently infected to removed in the SIR model equals zero.

[14]Since the law of motion (1) does not explicitly account for deaths it does not account for changes in the population size and in implied population shares due to death. These effects are negligible when death rates are small as we assume. Blurring the distinction between currently infected and removed households also amounts to assuming that the groups are indistinguishable in terms of their economic characteristics (e.g., productivity). We view the consequences of this assumption as minor as well.

the parameter $\bar{y}$ plays the same role. A lower $\bar{y}$ implies a larger share of the population that gets never infected.

The law of motion (1) also is closely related to the SIS model in which households, once infected, randomly recover and return to the susceptible pool (rather than the removed pool as in SIR models) because infection does not confer immunity (Hethcote, 1989).[15] Our framework differs from the SIS model insofar as we represent infections in terms of flows, $\dot{y}^g(t)$, while in the SIS model $y(t)$ corresponds to the stock of currently infected persons. In the SIS model a higher value of γ decreases the steady-state stock y while in our framework the relationship between γ and the steady-state infection flow $\dot{y}^g = \gamma y$ is inverse-U-shaped. In an extended SIR model with loss of immunity (a SIR-S model) the effect of γ on y would be positive, contrary to the SIS model. We view such a positive effect of γ on (the costs of) infections as more plausible and we therefore restrict attention to small values of γ which guarantee that γy increases in γ.[16]

In conclusion, the law of motion (1) nests standard epidemiological models with one endogenous state variable and constitutes a flexible and tractable alternative to other standard models with two endogenous state variables (see also the discussion in appendix A). Relative to epidemiological models it introduces a role of economic activity on infections. We summarize our assumptions regarding the epidemiological part of the model as follows:

Assumption 1. Epidemiological dynamics are described by the law of motion (1) with $g(a)\beta\bar{y} \gg \gamma$ in steady state. Function g is smooth, strictly positive, and strictly increasing. The social costs of infection are given by (4).

2.2 Economics

2.2.1 Households

We assume that members of the pre, post, and neutral groups are homogeneous as far as their economic characteristics are concerned (see the discussion below on heterogeneity). A household i chooses the activity level $a_i(t)$ over time in order to maximize an intertemporal objective which accounts for the immediate economic effects of activity and for the costs of infection that the household bears. Households take aggregates as well as the law of motion (1) as given and discount the future at rate $\rho > 0$.

We represent the immediate economic effects by an indirect utility function, u, that depends on the individual choice, $a_i(t)$, and satisfies the following assumptions:

Assumption 2. Function u is smooth, twice differentiable, and strictly concave. It satisfies $\lim_{a\downarrow 0} u'(a) = \infty$ and reaches a maximum at $a^\star \in (0, \infty)$, $u'(a^\star) = 0$.

We represent the costs of infection that the household internalizes by the product of two terms: the social costs given in (4) and a factor ξ that depends on the household's as

[15] The SIS model has $\bar{y} = 1$.

[16] Without loss of generality, let $g(a) = 1$ and $\bar{y} = 1$. Both in the SIS model and our framework the steady-state values satisfy $\beta(1-y) = \gamma$. In the SIS model we therefore have $dy/d\gamma = -1/\beta$ while in our framework, $d\dot{y}^g/d\gamma = d(\gamma y)/d\gamma = y - \gamma/\beta$. In a SIR-S model the steady-state values satisfy $\beta xy = cy = \gamma z$ (see appendix A), and thus $dy/d\gamma > 0$.

well as the aggregate activity level,

$$\xi(a_i(t), a(t))\, \psi \dot{y}^g(t). \tag{5}$$

Let ξ_{a_i} denote the partial derivative of ξ with respect to $a_i(t)$. We make the following assumption regarding ξ:

Assumption 3. Function ξ is smooth, twice differentiable, and convex in $a_i(t)$. Moreover, $\xi(a(t), a(t)) \equiv 1$ and $\xi_{a_i}(a_i(t), a(t))$ is homogeneous.

Convexity implies that the marginal cost of infection perceived by an individual household is nondecreasing in the household's activity level. The condition $\xi(a(t), a(t)) \equiv 1$ imposes the natural restriction that the total private costs borne by households (but not necessarily their marginal costs) sum to the social costs. The homogeneity assumption is made for tractability reasons.

Assumption 3 allows for a variety of cost specifications. To give one example, suppose that $\xi(a_i(t), a(t)) = \alpha(a_i(t))^\omega/(a(t))^\omega + (1-\alpha)$ for $\omega \geq 1$. Since $a_i(t) = a(t)$ in equilibrium, private costs sum to social costs and the equilibrium share of the social marginal costs that the household internalizes equals $\alpha\omega/a(t)$. Depending on α this share may be much smaller than unity, representing for instance the consequences of health insurance and moral hazard. Nevertheless, for $\omega > 1$, it is strictly convex from the perspective of an individual household, capturing for example congestion effects and rivalry in the consumption of health care services.

2.2.2 Government

We assume that policy makers have instruments to control economic activity along the activity-infection margin, for instance by imposing social distancing measures, closure of non-essential businesses, or other lockdown measures. If so desired, the government may also stimulate activity. Using these instruments the government faces the same program as a social planner.

2.3 Discussion

Heterogeneity and Multiple Policy Instruments In the interest of tractability and transparency our setup abstracts from heterogeneity. This has obvious limitations, some more important than others.

One dimension of heterogeneity concerns differences in activity level between individuals (which we do not allow for). This dimension appears of second order in the context of COVID-19 where it has been documented that many infected individuals remain asymptomatic or show only mild symptoms.[17] As long as tests are only selectively applied, as is still the case at the time of this writing, many infected individuals therefore necessarily behave like persons who have not been infected. Our modeling assumption of merging

[17] See also the discussion in Farboodi et al. (2020, p. 39).

(relative to a SIR structure) the susceptible, infected, and recovered groups as far as behavior is concerned reflects this.

Other dimensions of heterogeneity concern the costs of reduced activity (which might differ across groups) and infection dynamics (e.g., the presence of "super spreaders"). Kaplan et al. (2020) analyze how optimal policy should reflect the fact that a lockdown imposes differential costs on households. They find that economic exposure to the pandemic and lockdowns is highly correlated with financial vulnerability implying very uneven losses across the population. Acemoglu et al. (2020) analyze fatality rates by age group and they assess the welfare losses due to a government's inability to properly target lockdown measures. If targeting is possible the optimal policy implies a strict and long lockdown for the most vulnerable groups only. Our analysis neither reflects cross-sectional conflict nor does it allow for targeted policy interventions; that is, it implicitly assumes that individuals perfectly insure each other and the government can only affect average activity.

Herd Immunity We assume that absent the exogenous arrival of a cure, infections only stop when y reaches $\bar{y}$, at least as long as $g(a(t)) > 0$. While exogeneity of $\bar{y}$ is restrictive the model does not completely rule out to think about strategies to reach herd immunity. Specifically, slowing down infections in the model is a strategy to more likely reach herd immunity with fewer infections.

"Flattening the Curve" Our analysis stipulates that the costs of infection are proportional to $\dot{y}(t)$. This implies that absent time discounting and without the prospect for a cure there is no incentive to flatten the curve (when setting $g(a) = 0$ is excessively costly). We consider this feature of the model to be useful. As others have shown flattening-the-curve considerations specifically arise in situations where hospital capacities are limited. In the context of COVID-19 this has been the case in some countries but not in others. Our analysis focuses on tradeoffs that arise independently of such capacity constraints. For an analysis of optimal policy with a specific focus on capacity constraints, see for example Miclo et al. (2020).

2.4 Functional Form Assumptions and Calibration

To sharpen analytical results we sometimes impose the preference assumption

$$u(a) = \sigma(\ln(a) - a + 1),\ \sigma > 0,$$

implying $a^\star = 1$ and $u(a^\star) = 0$. We use this functional form because it is flexible and convenient and yields a tractable first-order condition.[18] Unless otherwise noted, we let $\sigma = 1$. Similarly, we sometimes impose the assumption

$$g(a) = a^n, n \in \mathbb{N},$$

[18] See also Çenesiz and Guimaraẽs (2020) or Farboodi et al. (2020).

for the law of motion. This specification allows for both constant and increasing returns to scale as far as the effect of activity on infectious dynamics is concerned.[19]

Throughout the paper we use numerical simulations to illustrate our results. The simulations make use of the functional form assumptions described above and are based on parameter values that we calibrate to match properties of the COVID-19 pandemic. Our unit of time is a day and $t = 0$ corresponds to mid March 2020. Accordingly, we set $\rho = -\ln(0.95)/365$ (five percent annual discount rate) and $\nu = 1/(365 * 1.5)$ (one-and-a-half years expected duration until discovery of a vaccine).[20]

In appendix A we describe in detail how we calibrate the remaining parameters, using information about parameter values in the canonical SIR model and theoretical results connecting SIR models and the logistic model.[21] This yields $y(0) = 0.1893 \cdot 10^{-3}$, $\beta = 0.9660 \cdot 10^{-1}$ (corresponding to an infection rate in the SIR model (at normal activity level) of 0.1333), and $\bar{y} = 0.75$. We set $\gamma = 0$ except when we analyze the consequences of lack of immunity.

To calibrate the derivative $\xi_{a_i}(a, a)$ and the social cost parameter ψ we use estimates of expected health care and mortality costs as well as households' willingness to pay to eliminate COVID-19 induced mortality risk.[22] We assume that households fully internalize mortality risk but not the social marginal costs of health care implying a social cost parameter $\psi/\sigma = 193.4$ and an internalization rate (the ratio of private and social marginal costs) of 0.8266. In the model, the internalization rate equals ζ/n where $\zeta \equiv \xi_{a_i}(1, 1)$. Table 1 summarizes the baseline calibration. The parameter σ only scales values; we normalize it to one.

3 Equilibrium and First Best

Let $U^\star \equiv u(a^\star)/\rho$ denote the value of the representative household when first-best activity is chosen permanently and no costs of infection occur. The latter value is attained once a vaccine is developed or all households have gained immunity because $y(t) \approx \bar{y}$ and $\gamma = 0$.

We represent the equilibrium conditions as well as the government's optimality conditions recursively. The state in the program of the government or an individual household is given by (y, t). When $T = \infty$ such that there is no deterministic terminal date of the disease then the state only includes y.

[19] While in the canonical SIR model a doubling of population shares implies a quadrupling of infections this is not the case in a modified SIR model (Gonzalez-Eiras and Niepelt, 2020b). The epidemiological evidence on constant versus increasing returns to scale ("frequency dependence" versus "density dependence") is mixed (Hethcote, 1989). Acemoglu et al. (2020) allow for both cases. Farboodi et al. (2020) opt in their baseline for the quadratic formulation, taking as a reference the distinction made by Diamond and Maskin (1979).

[20] The probability of recovery until time t equals $1 - \mathrm{e}^{-\nu t}$; the expected duration until discovery thus equals $\int_{t=0}^{\infty} t\nu \mathrm{e}^{-\nu t} dt = \nu^{-1}$. See, e.g., Alvarez et al. (2020).

[21] We rely on parameter estimates by Atkeson (2020), Ferguson et al. (2020), Greenstone and Nigam (2020), Hall et al. (2020), and Russell et al. (2020).

[22] We rely on parameter estimates by Bartsch et al. (2020), Hall et al. (2020), and Menachemi et al. (2020).

Parameter	Value
ρ	$0.1405 \cdot 10^{-3}$
ν	$0.1826 \cdot 10^{-2}$
$y(0)$	$0.1893 \cdot 10^{-3}$
β	$0.9660 \cdot 10^{-1}$
$\bar{y}$	0.7500
ψ/σ	$0.1934 \cdot 10^{3}$
ζ/n	0.8266
σ	1.0000

Table 1: Baseline calibration. See the text and appendix A for explanations.

3.1 Decentralized Equilibrium

Letting U denote the household's value function, the HJB equation reads

$$(\rho + \nu)U(y,t) = \max_{a_i} u(a_i) - \xi(a_i, a(y,t))\,\psi\dot{y}^g(y,t) + U_t(y,t) + \dot{y}(y,t)U_y(y,t) + \nu U^\star$$

subject to (1) where value function subscripts denote partial derivatives. The left-hand side of the HJB equation represents the risk-adjusted required return on the household's value and the right-hand side represents the dividend and capital gains components of the return. The dividend component contains the immediate economic benefit of activity net of the private costs of infection. The capital gains component reflects the change in the value due to changes in the state, as well as the risk that the disease is eradicated in which case the household chooses the activity level $a^\star$ and attains $U^\star$.

The first-order condition with respect to a_i yields

$$u'(a_i) = \xi_{a_i}(a_i, a(y,t))\,\psi\dot{y}^g(y,t) \text{ s.t. } (1).$$

Since individual and aggregate choices coincide in equilibrium the activity level in the decentralized equilibrium satisfies

$$u'(a(y,t)) = \xi_{a_i}(a(y,t), a(y,t))\,\psi g(a(y,t))\beta y(\bar{y} - y), \tag{6}$$

where we substitute equation (1).

3.2 First Best

As discussed earlier, the government effectively faces the program of a social planner. Unlike an individual household the social planner accounts for the fact that individual and aggregate activity choices coincide in equilibrium. As a consequence the planner internalizes the consequences of economic activity for infection dynamics.

Letting V denote the value function of the government/social planner, its HJB equation reads

$$(\rho+\nu)V(y,t)=\max_a u(a)-\psi\dot{y}^g(y,t)+V_t(y,t)+\dot{y}(y,t)V_y(y,t)+\nu U^\star$$

subject to (1). Recall that the derivative of $\dot{y}^g(y,t)$ and $\dot{y}(y,t)$ with respect a equals $g'(a)\beta y(\bar{y}-y)$. The activity level chosen by the government thus solves

$$u'(a(y,t))=g'(a(y,t))\beta y(\bar{y}-y)\left(\psi-V_y(y,t)\right). \tag{7}$$

4 Analysis

We now turn to the analysis of equilibrium and optimal policy.

4.1 Social Planner Allocation

If the initial value of y equals zero the law of motion (1) implies that there will be no infections. Accordingly, the government's optimal choice is to engage in the ideal level of economic activity, $a^\star$, such that $V(0,t)=U^\star$. The same holds true for $y=\bar{y}$ but only if $\gamma=0$ such that there will be no loss of immunity and subsequent renewed infections: if $\gamma=0$ then $V(\bar{y},t)=U^\star$.

For any other value of y the law of motion in combination with $g'>0$ (assumption 1) and the equilibrium requirement $a(y,t)>0$ (assumption 2) implies that the government faces current and future costs of infection as well as, possibly, losses due to reduced economic activity. Since the discount rate is finite this implies that $V(y,t)<U^\star$ for all $y\in(0,\bar{y})$, and also $V(\bar{y},t)<U^\star$ if $\gamma>0$. Since V is continuous, as established in the following lemma, $V(y,t)$ is decreasing in a neighborhood of $y=0$.

Lemma 1. Under assumptions 1 and 2, $V(0,t)=U^\star$ and $V(y,t)<U^\star$ for all $y\in(0,\bar{y})$. Moreover, if $\gamma=0$, then $V(\bar{y},t)=U^\star$; if $\gamma>0$, then $V(\bar{y},t)<U^\star$; and V is continuous.

Figure 1 illustrates additional properties of the solution of the government's problem under the assumption that there is no re-infection risk ($\gamma=0$) and focusing on the time autonomous case ($T=\infty$ such that time is not a state variable). When $\gamma=0$ the appropriate terminal condition to solve the government's HJB equation is $\lim_{y\to\bar{y}}V(y)=U^\star$ (and parallel for the decentralized equilibrium discussed below). The top left panel of the figure displays the government's value function, the top right panel the government's choice of activity level, and the left panel in the middle row infections, all as functions of y.[23] We discuss the other panels later.

[23]The figure is drawn using the baseline calibration introduced earlier. Unless otherwise noted, we use that calibration throughout when presenting numerical examples. When we plot functions of the state or of time it is always understood that these relationships apply before a cure for the disease arrives. (Afterwards $a(t)$ would equal unity even if $y<\bar{y}$, etc.) Also, for simplicity we omit qualifications of this type when stating our formal results.

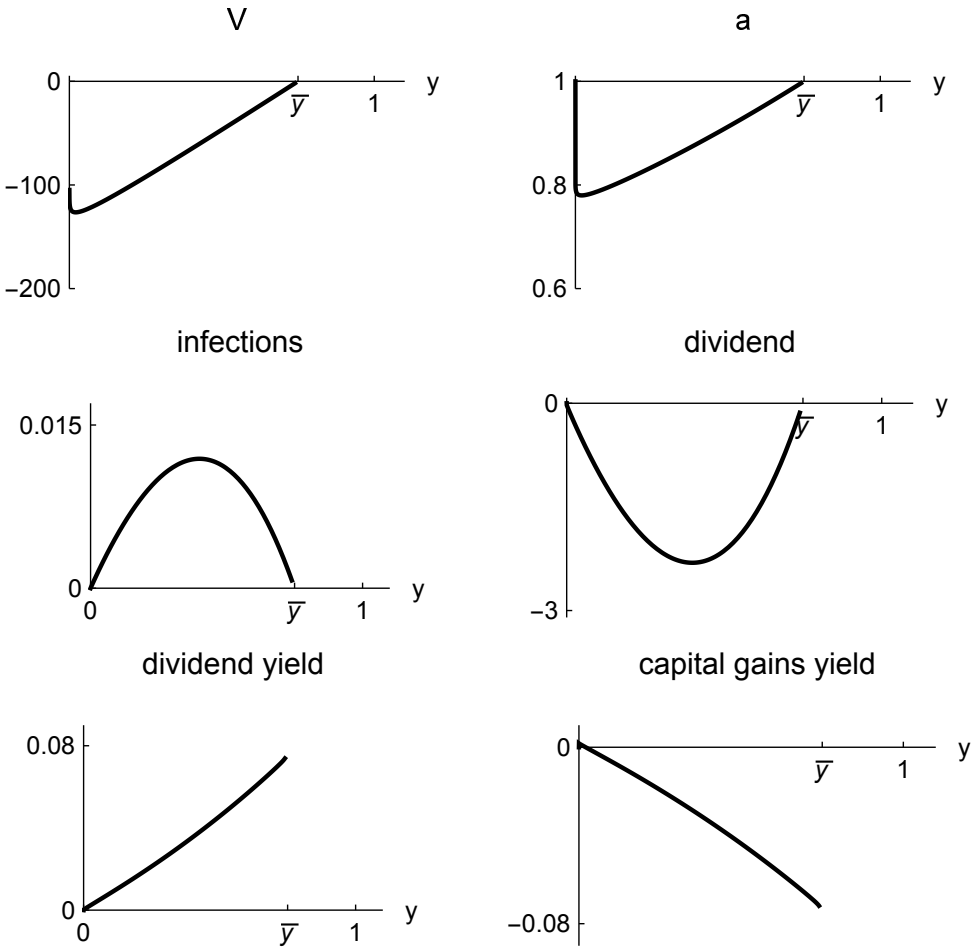

Figure 1: Value function, activity level, infections, and other outcomes in the government's program.

We note several general features. First, there exists a $y^{\min} \in (0, \bar{y})$ such that V attains its global minimum at $y^{\min}$ with $V(y^{\min}) < U^{\star}$. This follows directly from lemma 1. Clearly, higher costs of infection (ψ) lower the government's value function. The following lemma further characterizes $y^{\min}$:

Lemma 2. Under assumptions 1 and 2 and if $\gamma = 0$ and $T = \infty$, V has a unique minimum at $y^{\min} \leq \bar{y}/2$. Parameter changes that imply a higher (lower) $y^{\min}$ also imply less (more) pronounced convexity of V around $y^{\min}$.

Second, the optimal choice of activity is not symmetric although the function $\beta y(\bar{y}-y)$ is symmetric around the point $y = \bar{y}/2$. The asymmetry follows from the fact that the government's first-order condition does not only account for the effect of activity on infections and thus costs of infection (this alone would yield a symmetric policy function) but also for the effect of activity on the continuation value, and this effect is not symmetric. For example, in our benchmark specification with $u(a) = \sigma(\ln(a) - a + 1)$ and $g(a) = a$, the government's first-order condition (7) reads $\sigma\left(1/a(y) - 1\right) = \beta y(\bar{y} - y)(\psi - V'(y))$,

which reduces to

$$a(y) = \frac{\sigma}{\sigma + \beta y(\bar{y} - y)(\psi - V'(y))}. \tag{8}$$

Note that this implies $(\rho + \nu)V(y) = \sigma \ln(a(y))$.[24]

Compare this to the decentralized equilibrium condition, equation (6). Using $\xi_{a_i}(a, a) = \xi_{a_i}(1, 1)/a = \zeta/a$ (which we establish in the proof of proposition 1 below) this condition reduces to

$$a(y) = \frac{\sigma}{\sigma + \beta y(\bar{y} - y)\zeta\psi}. \tag{9}$$

Note that the equilibrium choice of activity is symmetric around $\bar{y}/2$.

Third, to gain intuition for the shape of V consider first the path of infections displayed on the left in the middle row of figure 1. Infections are hump shaped because the logistic function is S shaped; the fact that $a(y)$ varies with the state does not fundamentally alter that result. As discussed earlier the hump shaped path of infections is consistent with the predictions of many epidemiological frameworks. Since the costs of infection are increasing in the number of infections they are hump shaped as well.

Consider next the dividend component of the government's return on the right-hand side of the government's HJB equation, $u(a(y)) - \psi \dot{y}^g(y) + \nu U^\star$. This dividend component is displayed on the right-hand side of the middle row of figure 1. It reflects the costs of infection as well as the losses from reduced activity. The HJB equation implies that the capital gains component of the return, $\dot{y}(y)V'(y)$, equals the required return on the government's value, $(\rho + \nu)V(y)$, net of the dividend component. Equivalently, the dividend yield and the capital gains yield add to $\rho + \nu$. The bottom row of figure 1 displays the two yields. The dividend yield is always positive because both the dividend and the value are negative. In contrast, the capital gains yield changes sign at $y^{\min}$. Since capital gains are initially negative ($V'(y) < 0$) and $V(y) < 0$ the capital gains yield is initially positive; once capital gains become positive the yield turns negative.

Over most of the state space the slope of V is relatively constant. Since infections are hump shaped, however, capital gains $\dot{y}(y)V'(y)$ are hump shaped as well. When many infections occur, costs of infection and losses from low activity quickly materialize. Accordingly, the value function, which discounts the future costs and losses, swiftly increases; the capital gains are large and dividends low. In contrast, capital gains are negative for small values of y. Infections are still low in this part of the state space but they accelerate, and as a consequence, dividends are not yet strongly depressed but the period when they will be is approaching quickly. The value function which discounts the future costs and losses reflects this.

4.2 Externalities

Individual households do not perceive their own choice of activity to affect the aggregate activity level. As a consequence they do not fully internalize the welfare consequences of higher activity. This gives rise to two externalities: a "static" one related to the

[24] Using equation (8), we have $(\rho+\nu)V(y) = \sigma(\ln(a(y))+1) - a(y)\{\sigma + \beta y(\bar{y}-y)(\psi - V'(y))\} + \nu U^\star = \sigma(\ln(a(y))+1) - \sigma + \nu U^\star = \sigma \ln(a(y))$.

contemporaneous costs of infection, and a "dynamic" one related to the effect of activity on the state and the continuation value. We analyze these externalities in turn.

A comparison of the first-order conditions (6) and (7) reveals two differences between the marginal costs of activity perceived by an individual household and the government. Subtracting the right-hand side of (7) from the right-hand side of (6) and evaluating terms at a common activity level yields $\mathcal{S}(a,y)+\mathcal{D}(a,y,t)$ with

$$\begin{aligned}\mathcal{S}(a,y) &\equiv \psi\dot{y}^g(y)\left(\xi_{a_i}(a,a)-\frac{g'(a)}{g(a)}\right),\\ \mathcal{D}(a,y,t) &\equiv \dot{y}^g(y)\frac{g'(a)}{g(a)}V_y(y,t).\end{aligned}$$

$\mathcal{S}(a,y)$ represents the static externality which arises because households do not fully internalize the effects of their choice of activity on the costs of infection. $\mathcal{D}(a,y,t)$ represents the dynamic externality which arises because households do not internalize that higher activity increases the infection rate and thus the continuation value.

The following proposition characterizes the externalities:

Proposition 1. Under assumptions 1 and 3 and if $g(a)=a^n$, the total externality equals

$$\frac{\dot{y}^g(y)}{a}n\left(\psi(\zeta/n-1)+V_y(y,t)\right). \tag{10}$$

The static externality is negative when $\zeta \leq n$.

The total externality is proportional to infections, $\dot{y}^g(y)$, because infections drive the costs of infection and change the state variable, which in turn affects the continuation value. The factor of proportionality contains two terms, $\psi(\zeta/n-1)$ and $V_y(y,t)$. The former reflects the fact that with $\zeta<n$, households do not fully internalize the negative consequences of their actions for the costs of infection; that is, there is a negative static externality. The latter factor represents the effect of economic activity on the continuation value, due to a higher infection rate. Individual households do not internalize this effect at all, giving rise to the dynamic externality.

Figure 2 illustrates the consequences of the externalities. As in figure 1, we focus on the time autonomous case without re-infections and we let $n=1$; as discussed previously, we assume $\zeta/n=0.8266$, that is, households internalize eighty three percent of the social costs of infection. The solid lines in the figure represent the outcomes implemented by the government and correspond to the schedules in figure 2; the dashed lines represent the equilibrium outcomes.

The externalities lower the value in equilibrium relative to the situation with government intervention, pushing the dashed line in the left panel below the solid one. More interestingly, the activity levels displayed in the right panel differ; early on, the government chooses a lower activity level than in equilibrium, later on the opposite holds true. The driving force behind the reversal is the capital gains component which lies at the source of the dynamic externality. For $y \geq y^{\min}$ the capital gains component is positive but only the government internalizes the capital gains when choosing the activity level.

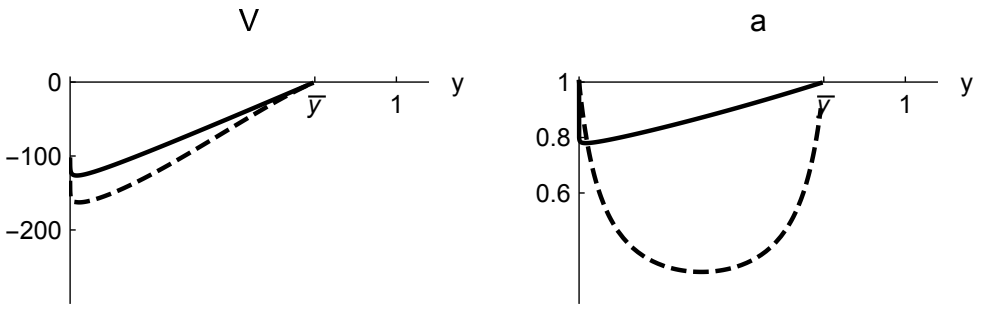

Figure 2: Value function and activity level in the government's program (solid) and in equilibrium (dashed).

When this effect is sufficiently strong to compensate for the negative static externality (due to $\zeta \leq n$) then the equilibrium activity level falls short of the level chosen by the government. In the figure, the capital gains component (and dynamic externality) equals zero at $y = y^{\min} \approx 0.0207$. At $y \approx 0.0252$ the total externality equals zero and the activity levels chosen by the government and in equilibrium coincide. For higher values of the state the total externality is positive, that is, equilibrium activity is too *low*.

Figure 3 illustrates how the corrective government interventions shape infections and activity over time. In contrast to figures 1 and 2 the horizontal axis now depicts time (in days), not the level of the endogenous state.

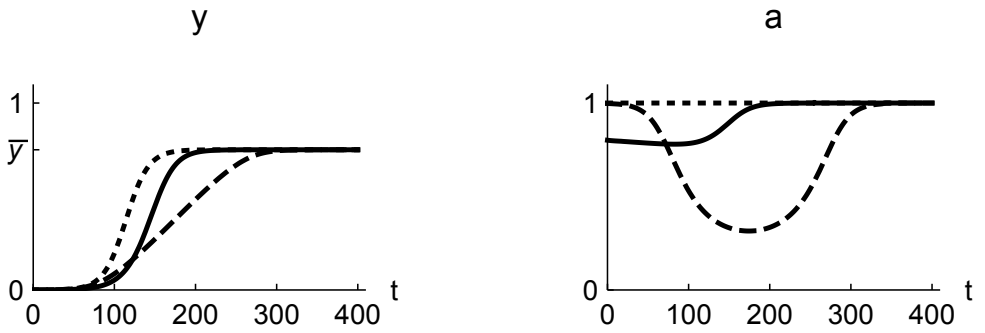

Figure 3: Infections and activity level in the government's program (solid), in equilibrium (dashed), and with no intervention (dotted).

For an extended period of time the government forces households to reduce activity by about twenty one percent. In equilibrium, households would reduce activity to the level imposed by the government only after approximately eighty days. The early government intervention delays infections relative to the situation without adjustments of activity and it also suppresses infections relative to equilibrium, but only modestly so. After approximately 130 days the government increases activity and infections increase as quickly as they would have a month or so earlier with no adjustments. In equilibrium, in contrast, activity now falls strongly because the product $y(t)(\bar{y} - y(t))$ approaches its maximum and individuals do not internalize the positive dynamic externality.

Evaluated at $y(0)$, the share of the U.S. population infected in mid March 2020, we

find $U(y(0)) \approx -145.8$ and $V(y_0) \approx -112.9$. Solving

$$\begin{aligned} \frac{\sigma}{\rho+\nu}(\ln(a^\star(1-\phi^u)) - a^\star + 1) &= -145.8, \\ \frac{\sigma}{\rho+\nu}(\ln(a^\star(1-\phi^v)) - a^\star + 1) &= -112.9 \end{aligned}$$

yields $\phi^u = 0.2493$ and $\phi^v = 0.1992$. Compared with the equilibrium outcome, the welfare gains due to optimal government intervention therefore amount to the equivalent of roughly five percent of lifetime consumption.

4.3 Lockdowns and Inverse Lockdowns

We refer to a "lockdown" as a situation where the government wishes to depress economic activity below the level chosen in equilibrium. Conversely, we refer to an "inverse lockdown" as a situation where the government wishes to stimulate economic activity beyond the level chosen in equilibrium. Instruments to implement a lockdown include for instance stay-at-home-orders, social distancing rules, business closures, or school closures. Instruments to implement an inverse lockdown may take the form of stimulation measures like monetary easing, temporary sales tax reductions, employment subsidies, or a "return-to-work bonus."

Our preceding analysis of externalities has direct implications for lockdowns and inverse lockdowns. We have the following result:

Proposition 2. Under assumptions 1 and 3 and if $g(a) = a^n$, the economy is in lockdown when

$$\psi(\zeta/n - 1) + V_y(y,t) < 0,$$

and in inverse lockdown when the reverse inequality holds.

Intuitively, starting from $a = a^\star$, there is a second-order loss of reducing activity but a first-order gain from slowing down infections. Unless there are zero infection dynamics both individuals and the government choose activity levels below $a^\star$. If the static externality (represented by the term $\psi(\zeta/n-1)$) and the dynamic externality (represented by the term $V_y(y,t)$) combined are negative then the government perceives a larger first-order gain from lowering activity. In this case the government imposes a lockdown in order to correct the distorted individual choices. Note that the condition in proposition 2 is not directly affected by the parameters γ or ν which determine the rate of re-infections or the arrival rate of a permanent cure. These parameters matter only indirectly because they affect the value function.

Turning to the timing of lockdowns and inverse lockdowns, recall that infections increase the government's continuation value—infections induce capital gains—once y reaches the value $y^{\min}$. Let y^0 denote the value of y (if it exists) at which the total externality equals zero, $\psi(\zeta/n-1) + V_y(y^0,t) = 0$; and let $V_y{}^{\max}$ denote the maximum value of $V_y(y,t)$, both along the path implemented by the government. We have the following result:

Proposition 3. Under assumptions 1 and 3 and if $g(a) = a^n$ and $\zeta < n$, lockdowns occur as follows:

i. Starting from small y, the government immediately imposes a lockdown;

ii. if $\gamma = 0$ and $V_y^{\max} > \psi(1 - \zeta/n)$ then the government also imposes an inverse lockdown;

iii. if V is locally convex at $y = y^0$ then an inverse lockdown immediately follows the lockdown.

The first part of proposition 3 is consistent with the fact that during the ongoing COVID-19 pandemic many governments imposed lockdown measures early on. The last part, in contrast, which concerns the reversal from a lockdown to an inverse lockdown, appears more surprising. It might partly explain stimulus measures such as temporary sales tax reductions or employment subsidies; and it suggests quite different policy interventions in the future.

Note that the basic intuition underlying the reversal result is general: Since an epidemic generates costs the value function during the transition is lower than after the transition; that is, at some point, society experiences capital gains. These capital gains arise due to the change of an aggregate state variable (or many such state variables) which an individual takes as given; that is, the capital gains are not internalized by individuals. As long as the capital gains are sufficiently large to outweigh negative static externalities the reversal result thus follows.

4.4 Quadratic Effect of Activity on Infections

When $n = 2$ reductions in activity suppress infections more strongly and this alters the tradeoffs perceived both by the government and individual decision makers. Equations (8) and (9) therefore no longer apply. Instead, the optimal and equilibrium activity levels now are the (positive) solutions of quadratic equations, and the value functions U and V change accordingly.

Figures 4 and 5 illustrate the outcomes when $n = 2$. Both $y^{\min}$ and the value of y at which the total externality equals zero increase (to approximately 0.0281 and 0.0343 respectively). The government reduces activity substantially more strongly than when $n = 1$, in contrast to individuals in equilibrium. Accordingly, the transition dynamics in the two cases are more similar than when $n = 1$.

The welfare costs of the pandemic under laissez faire are nearly unchanged, $\phi^u \approx 0.2484$, relative to the case with $n = 1$. In contrast, ϕ^v falls more strongly to approximately 0.1848, implying that the welfare gains from intervention increase to more than six percent.

4.5 Higher Costs of Infections

Returning to the case with $n = 1$, we consider the consequences of higher costs of infection. Doubling the cost parameter ψ we find that the values of $y^{\min}$ and of y at the point where the total externality equals zero now are given by approximately 0.0234 and 0.0285,

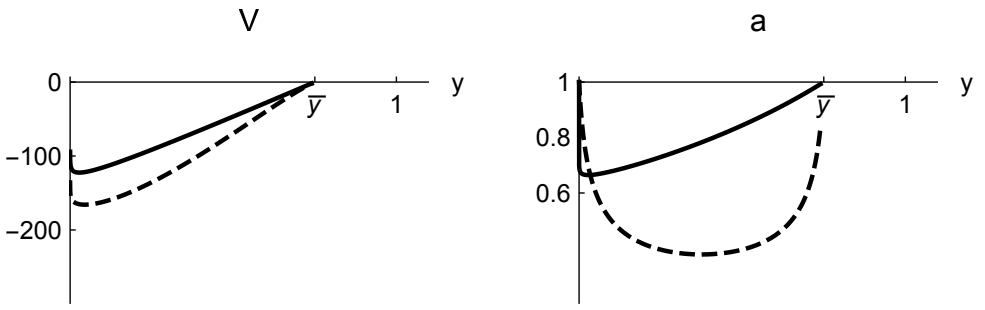

Figure 4: Value function and activity level in the government's program (solid) and in equilibrium (dashed) when $n = 2$.

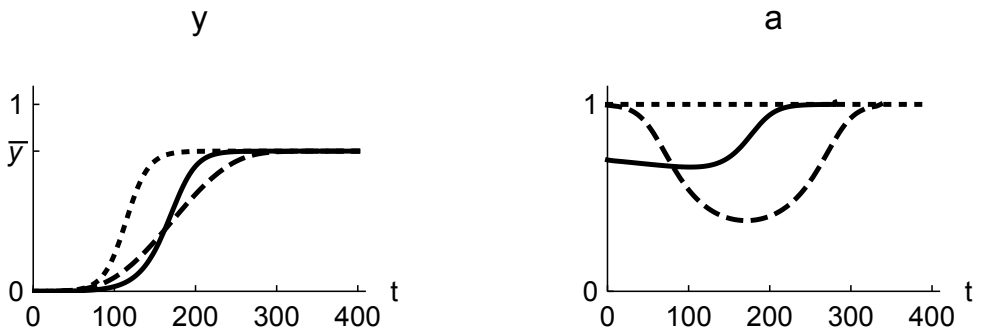

Figure 5: Infections and activity level in the government's program (solid), in equilibrium (dashed), and with no intervention (dotted) when $n = 2$.

respectively. Not surprisingly, both the government and individuals reduce activity much more strongly during the transition than in the baseline case. As a consequence, transition dynamics slow down.

The welfare costs of the pandemic are very large, both under laissez faire and under the optimal policy: $\phi^u \approx 0.4530$ and $\phi^v \approx 0.3502$. The welfare gains from intervention thus increase to more than thirteen percent of lifetime consumption.

4.6 Constraints on Policy Instruments

Constraints on policy instruments can prevent governments from implementing the first-best allocation. What are the consequences? Arguably the most relevant example of constrained instruments arises in a situation where the government can curtail economic activity (impose a lockdown) but lacks the powers to correct households' unwillingness to reengage after a lockdown (impose an inverse lockdown). The relationship beetween the value functions of the social planner, the government, and individuals in equilibrium differs in this case from the situation considered so far: In the range where the social planner imposes an inverse lockdown (high y values) the government's value function coincides with the *equilibrium* value function; and in the range where the social planner imposes a lockdown (low y values) the government's value function lies between the value function of the social planner and the equilibrium value function.

To solve for the government's value function subject to the instrument constraints we impose value matching and smooth pasting conditions. That is, we first solve for U and then find V and an endogenous value of the state, $\hat{y}$ say, satisfying the conditions

$$V(\hat{y}) = U(\hat{y}) \quad \text{and} \quad V'(\hat{y}) = U'(\hat{y}).$$

Figures 6 and 7 illustrate the results.

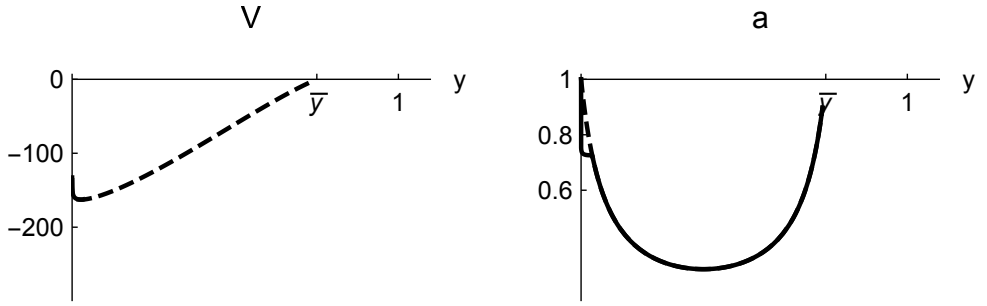

Figure 6: Value function and activity level in the government's program (solid) and in equilibrium (dashed) when an inverse lockdown is not feasible.

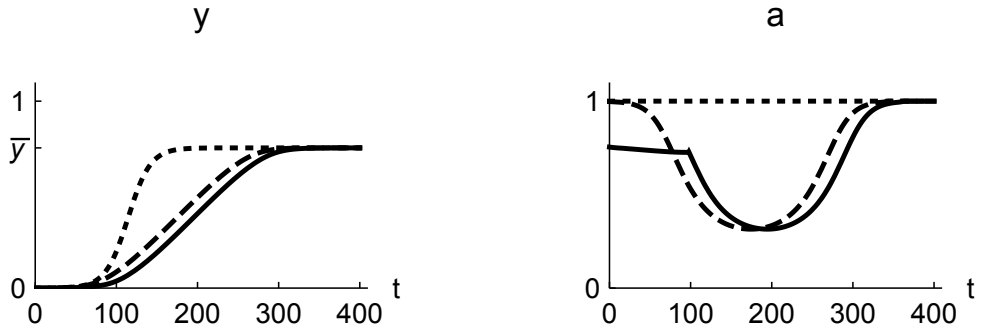

Figure 7: Infections and activity level in the government's program (solid), in equilibrium (dashed), and with no intervention (dotted) when an inverse lockdown is not feasible.

We find that the constraints imply a stricter but slightly shorter lockdown than in the baseline scenario. The activity level changes abruptly at $y \approx 0.0338$ when the lockdown ends. Thereafter, the dynamics under the optimal government policy resemble those in equilibrium, with a time lag. By definition, the welfare gains due to government intervention are smaller than when the government is unconstrained. More interestingly, the welfare gains from government intervention become rather small—the difference between ϕ^u and ϕ^v diminishes substantially ($\phi^u \approx 0.2493$ and $\phi^v \approx 0.2458$).

4.7 Precautionary Motives

Next, we ask how expected changes in the (epidemiological) environment affect optimal policy. Scenarios where such a question might arise include settings where policy makers

anticipate virus mutations; new medical treatments; improved test, trace, and quarantine strategies; improved implementability of lockdown restrictions; or, in contrast, increased political resistance against such restrictions.

We consider a one-time, permanent reduction in β by eighty percent (reflecting, e.g., improved test, trace, and quarantine strategies) that is expected to materialize stochastically, with arrival rate μ. We assume that $\mu = 1/120$, reflecting an expected duration of four months until the regime changes.

The left panel of figure 8 depicts the government's value function when β assumes the baseline value but may randomly fall to the lower value (solid line); when β permanently assumes the baseline value (dashed line); and when β permanently assumes the lower value (dotted line). The right panel depicts the activity level chosen by the government in the same three cases. Not surprisingly, the value function when regime change is possible is higher than in the baseline case reflecting the upside risk. Our central result is that the prospect of this upside risk leads the government to wait longer before relaxing the lockdown, and to substantially reduce activity levels until β falls: Rather than at $y \approx 0.0252$ (baseline case) the government switches to stimulative measures only at $y \approx 0.0517$; and early in the pandemic the government reduces activity to just about sixty percent. That is, expecting a better trade-off between lives and livelihoods in the future, the government imposes a significantly more aggressive lockdown initially.

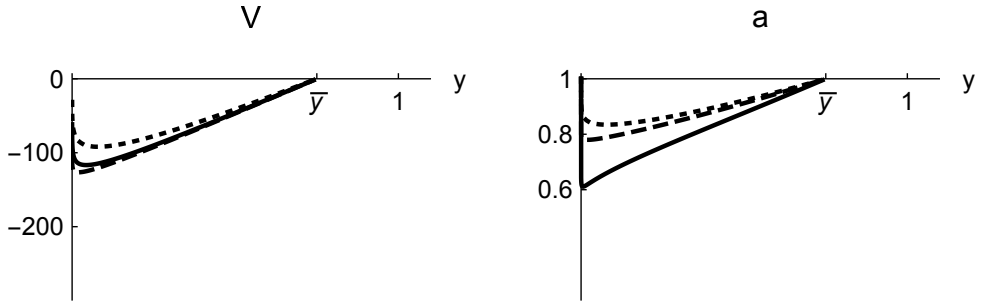

Figure 8: Value function and activity level in the government's program when β assumes the baseline value but may randomly fall (solid), permanently assumes the baseline value (dashed), or permanently assumes the lower value (dotted).

We also find that the upside risk reduces the welfare losses in equilibrium and under the optimal government intervention ($\phi^u \approx 0.1678$ and $\phi^v \approx 0.1438$); the welfare gains from optimal government intervention fall compared with the baseline case.

4.8 Lack of Immunity

Common sense intuition regarding plausible policy interventions to slow down infections often follows under the implicit assumption that persons who recover are protected from re-infection such that herd immunity will eventually be attained. This implicit assumption seems also present in much of the contemporary policy discussion surrounding COVID-19. But medical scientists cannot rule out yet that immunity wanes and individuals with antibodies may re-contract the disease (see the discussion in the introduction).

To study the implications of lack of immunity we consider the case with $\gamma > 0$ such that persons who have undergone infection in the past may join the pre-group again and subsequently recontract the disease. Accordingly, the steady-state value of y is given by $s(a) \equiv \bar{y} - \gamma/(a\beta)$; due to $\gamma > 0$ it lies strictly below $\bar{y}$.

We proceed in two steps. First, we analyze the steady state in decentralized equilibrium and under the optimal policy. Subsequently, we consider transition dynamics. In decentralized equilibrium, the steady state (a, y) is characterized by the law of motion and the first-order condition, equations (1) and (9):

$$\begin{aligned} \gamma &= a\beta(\bar{y} - y), \\ a &= \frac{\sigma}{\sigma + \beta y(\bar{y} - y)\zeta\psi}. \end{aligned}$$

This system has a unique positive solution,

$$\begin{aligned} a &= \frac{\sigma - \gamma\psi\bar{y}\zeta}{2\sigma} + \frac{\varphi}{2\beta\sigma}, \\ y &= \frac{\sigma + \gamma\psi\bar{y}\zeta}{2\gamma\psi\zeta} + \frac{\varphi}{2\beta\gamma\psi\zeta} \end{aligned}$$

with $\varphi \equiv \sqrt{\beta}\sqrt{4\gamma^2\psi\sigma\zeta + \beta(\sigma - \gamma\psi\bar{y}\zeta)^2}$. Differentiating this solution implies that in a neighborhood of $\gamma = 0$ (and $a = 1$), a is U-shaped; y decreases; and steady-state infections $\dot{y}^g = \gamma y$ are inverse-U-shaped in γ. For the reasons discussed in subsection 2.1.3 only the increasing segment of this inverse-U-shaped relationship is relevant.

The steady state in the government's program solves the conditions

$$\begin{aligned} \gamma &= a\beta(\bar{y} - y), \\ a &= \frac{\sigma}{\sigma + \beta y(\bar{y} - y)(\psi - V'(y))}, \\ (\rho + \nu)V'(y) &= -a\beta(\bar{y} - 2y)(\psi - V'(y)) - \gamma V'(y), \end{aligned}$$

where the first equation follows from equation (1); the second from the first-order condition (8); and the last from the envelope condition (using $\dot{y}(t) = 0$). Eliminating y and $V'(y)$ yields the equation

$$(\rho + \nu + \gamma)\gamma\psi = a(1 - a)\beta\sigma\left(\frac{\rho + \nu}{a\beta\bar{y} - \gamma} + 1\right), \tag{11}$$

which characterizes the government's choice of a. The ratio in parentheses is positive ($s(a) > 0$) and $a \in (0, 1)$ such that both the left- and the right-hand side of equation (11) are strictly increasing in γ. Moreover, the right-hand side is non-monotone in a. Equation (11) thus generally admits multiple solutions. Of course, only one of them constitutes the outcome implemented by the government. In fact, under the restriction that γ is small such that the model is well specified the solution of equation (11) necessarily yields a government choice for a which is decreasing in γ.

Proposition 4. Under assumptions 1–3 and if $\gamma > 0$, $T = \infty$, $u(a) = \sigma(\ln(a) - a + 1)$, and $g(a) = a$, there is a unique steady state in decentralized equilibrium. For small values of γ the government chooses a higher activity level (and higher y) than in decentralized equilibrium if $(\rho + \nu)(1 - \zeta) < \zeta\beta\bar{y}$.

That is, absent a static externality ($\zeta = 1$) the government's choice of a necessarily exceeds the equilibrium level of a, and more concern about the future (low ρ and/or ν) as well as a high infection rate (β) and a high $\bar{y}$ makes the same result more likely. When the static negative externality is maximal ($\zeta = 0$), in contrast, then the steady-state equilibrium level of a necessarily is suboptimally high.

Intuitively, on the one hand, the equilibrium level of a tends to exceed the optimal level when individuals do not fully internalize the costs of infection ($\zeta < 1$). On the other hand, the parameters β, $\bar{y}$, ρ, and ν increase the capital gains in the government's problem ($V'(y)$) which reduce the net costs of infection internalized by the government.

Turning to the dynamic analysis note first that to solve the HJB equations we need to impose different boundary conditions than previously where $\gamma = 0$: Since $\bar{y}$ is no longer a rest point the conditions $\lim_{y\to\bar{y}} V(y) = U^\star$ etc. no longer apply. Instead, we impose the new boundary condition that the value function evaluated at the steady-state y-value (characterized above) equals the capitalized value of steady-state utility flows net of steady-state flow costs of infection. That is, for example, we solve the government's HJB equation subject to the boundary condition

$$(\rho + \nu)V(y) = u(a) - \psi a\beta y(\bar{y} - y),$$

where a is optimal.

In the quantitative analysis, we consider small values for γ, either $\gamma = 0.001$ or $\gamma = 0.005$ (i.e., immunity is lost on average after 1000 or 200 days, respectively). We find that the effects are nevertheless substantial or even very large. For $\gamma = 0.001$ the steady-state values of y and a under the optimal policy are given by $y \approx 0.7396$ and $a \approx 0.9942$ while in equilibrium, they equal $y \approx 0.7383$ and $a \approx 0.8820$. Moreover, $\phi^u \approx 0.3257$ and $\phi^v \approx 0.2769$. When we let $\gamma = 0.005$ the steady-state values of y and a under the optimal policy are given by $y \approx 0.6942$ and $a \approx 0.9271$ while in equilibrium, $y \approx 0.6434$ and a collapses to approximately 0.4857. Moreover, $\phi^u \approx 0.5587$ and $\phi^v \approx 0.4756$.

4.9 Deterministic Arrival of a Cure

Finally, we consider the case when a cure arrives deterministically, at date $T < \infty$, rather than stochastically ($\nu = 0$; this also implies $\gamma = 0$).[25] Such an environment might represent a scenario with several promising candidates for a vaccine which need to undergo final trials but at least one of them is always certain to work; or the situation in a poor country where limited financial means and medical infrastructure force the government to wait until the cure becomes freely available.

[25] Farboodi et al. (2020) briefly consider the case with a deterministic vaccine arrival but they do not find our result that optimal policy may be dis-continuous in T.

With a deterministic arrival date, time becomes a second state variable and the government's problem may be formulated more conveniently using the Hamiltonian

$$\mathcal{H}(a(t), y(t), t) = u(a(t)) - \psi g(a(t))\beta y(t)(\bar{y} - y(t)) + \mu(t)g(a(t))\beta y(t)(\bar{y} - y(t))$$

with $\mu(t)$ denoting the (present value) multiplier attached to the law of motion. An optimal plan satisfies the boundary condition $\mu(T) = 0$ as well as the conditions $\mathcal{H}_a(a(t), y(t), t) = 0$ and $\mathcal{H}_y(a(t), y(t), t) = -\dot{\mu}(t) + \rho\mu(t)$.[26] This implies

$$\begin{aligned} u'(a(t)) &= g'(a(t))\beta y(t)(\bar{y} - y(t))\left(\psi - \mu(t)\right), \\ \dot{\mu}(t) &= g(a(t))\beta(\bar{y} - 2y(t))\left(\psi - \mu(t)\right) + \rho\mu(t), \\ \dot{y}(t) &= g(a(t))\beta y(t)(\bar{y} - y(t)), \end{aligned}$$

where the last condition represents the law of motion.

As it turns out the government's optimal policy may not be continuous in T. To prove this result we consider the special case with $\rho = 0$ which allows for a sharper characterization:

Proposition 5. Under assumption 1 and $\gamma = \rho = \nu = 0$ and $T < \infty$, the government's optimal choice satisfies $a(t) = a$. If

$$\ln\left(\frac{\bar{y}}{y(0)} - 1\right) > \frac{\underline{a} - 1}{1 + \ln(\underline{a}) - \underline{a}}, \quad \underline{a} \equiv \frac{1}{1 + \beta\frac{\bar{y}^2}{4}\psi/\sigma}, \tag{12}$$

then $\exists\ T^* < \hat{T} \equiv \frac{\ln(\bar{y}/y(0)-1)}{\underline{a}\beta\bar{y}}$ such that a is dis-continuous at T^*.

Proposition 5 implies that the government's program reduces to the choice of a *constant* activity level. Moreover, when condition (12) is satisfied then, starting from a low value of T, the optimal policy responds to an increase in T by both reducing activity and allowing for a higher stock $y(T)$ when the cure arrives. Eventually, the marginal costs of curtailing economic activity become so large that it is no longer worth to impose a severe lockdown in order to keep $y(T)$ in check. At that point it becomes optimal instead to choose a much higher activity level and to "accept" higher infection numbers. When condition (12) is not satisfied, in contrast, i.e. when the initial level of infections is sufficiently high, there are smaller gains from saving lives and the optimal policy calls for a high level of activity, close to unity, and a high level of $y(T)$, regardless of T.[27]

Figure 9 displays the contour lines of the government's objective function under our baseline calibration. The figure shows that, as T increases, the optimal activity level first falls before eventually dis-continuously jumping close to unity.

[26] The multiplier $\mu(t)$ corresponds to the partial derivative $W_y(y(t), t)$ such that $\dot{\mu}(t) = W_{yt}(y(t), t) + W_{yy}(y(t), t)g(a(t))\beta(\bar{y} - 2y(t))$.

[27] The government's program does not satisfy Arrow's second order conditions because the maximized Hamiltonian fails to be concave in y for all $t \in [0, t]$. Accordingly, the first-order conditions are not sufficient for a global maximum.

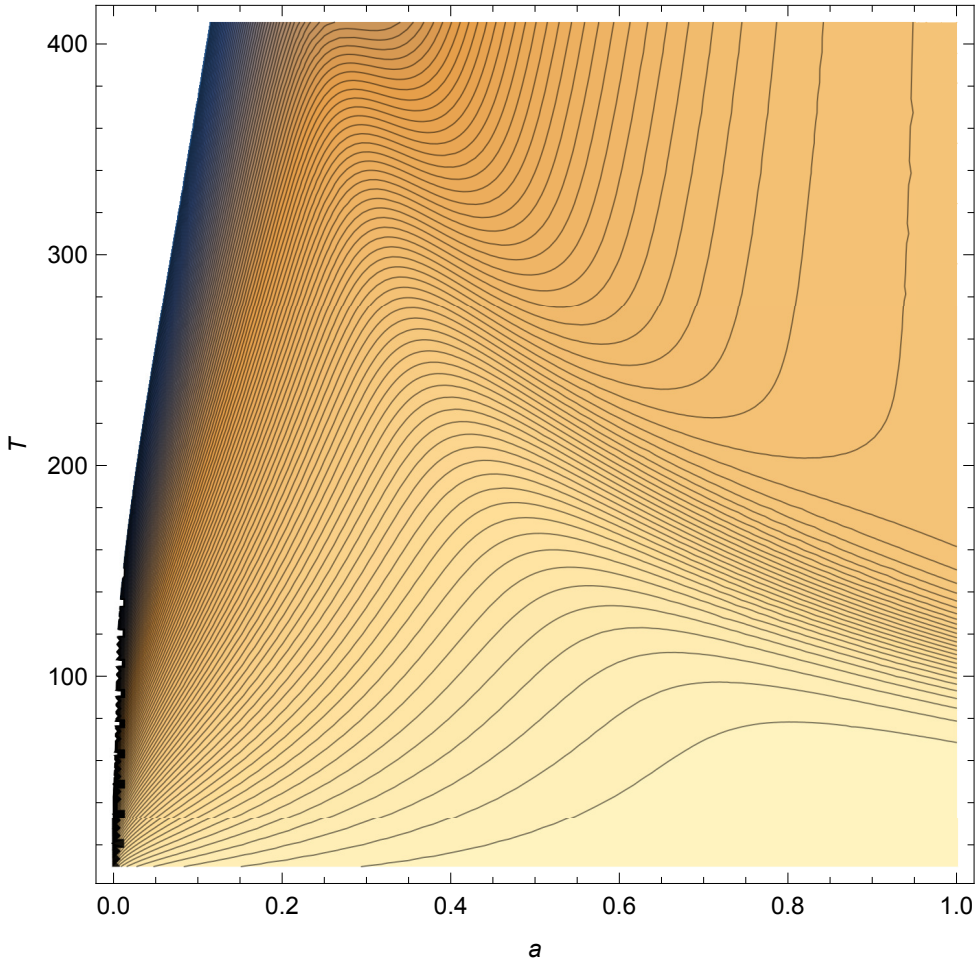

Figure 9: Contours of the government's objective function when T is finite.

5 Conclusion

We have developed a flexible model of infectious dynamics with a single endogenous state variable and economic choices by individuals and a government that can be used to inform about the ongoing COVID-19 pandemic. The model allows us to explore the welfare gains from government intervention relative to the outcome under laissez faire where households adjust their behavior for fear of infection but do not internalize static and dynamic externalities.

We find several novel results. First, a lockdown is generally followed by its opposite—policies to stimulate activity beyond the privately optimal level. Initially, the total externality from activity is always negative, giving incentives to reduce activity. But eventually, infections generate capital gains for society and if these are strong enough they outweigh the negative static externalities. This intuition extends to richer models with many state variables.

Second, social distancing has small welfare gains when governments lack instruments to stimulate activity to internalize the capital gains. Third, prospects of a more manage-

able epidemiological environment, e.g. due to more efficient test-and-trace-and-quarantine methods or improved medical treatments, imply that governments should impose a stricter lockdown, for longer.

Fourth, re-infection risk may imply a more cautious steady state than optimal if static externalities are small. Even if the virus becomes endemic there are capital gains that the planner internalizes, implying higher activity levels. Finally, when a cure or vaccination arrives deterministically, optimal policy might be dis-continous in the sense that the lockdown is light when the arrival date exceeds a specific value, but strict otherwise. The postponement of the arrival data increases the cost of lost livelihoods to a point at which the government is willing to accept higher infection levels and turn a blind eye on the epidemic.

References

Abel, A. B. and Panageas, S. (2020). Optimal management of a pandemic in the short run and the long run, *Working Paper 27742*, NBER, Cambridge, Massachusetts.

Acemoglu, D., Chernozhukov, V., Werning, I. and Whinston, M. D. (2020). Optimal targeted lockdowns in a multi-group SIR model, *Working Paper 27102*, NBER, Cambridge, Massachusetts.

Alvarez, F., Argente, D. and Lippi, F. (2020). A simple planning problem for COVID-19 lockdown, *Working Paper 26981*, NBER, Cambridge, Massachusetts.

Atkeson, A. (2020). What will be the economic impact of COVID-19 in the US? rough estimates of disease scenarios, *Working Paper 26867*, NBER, Cambridge, Massachusetts.

Bailey, N. T. J. (1975). *The Mathematical Theory of Infectious Diseases and its Applications*, 2 edn, Hafner Press, New York.

Bartsch, S. M., Ferguson, M. C., McKinnell, J. A., O'Shea, K. J., Wedlock, P. T., Siegmund, S. S. and Lee, B. Y. (2020). The potential health care costs and resource use associated with COVID-19 in the United States, *Health Affairs* **forthcoming**.

Bethune, Z. A. and Korinek, A. (2020). Covid-19 infection externalities: Trading off lives vs. livelihoods, *Working Paper 27009*, NBER, Cambridge, Massachusetts.

Bisin, A. and Moro, A. (2020). Learning epidemiology by doing: The empirical implications of a spatial-SIR model with behavioral responses, *Working Paper 27590*, NBER, Cambridge, Massachusetts.

Çenesiz, A. and Guimaraẽs, L. (2020). COVID-19: What if immunity wanes? arXiv 2008.03283.

Diamond, P. A. and Maskin, E. (1979). An equilibrium analysis of search and breach of contract, i: Steady states, *The Bell Journal of Economics* **10**(1): 282–316.

Eichenbaum, M. S., Rebelo, S. and Trabandt, M. (2020). The macroeconomics of epidemics, *Working Paper 26882*, NBER, Cambridge, Massachusetts.

Farboodi, M., Jarosch, G. and Shimer, R. (2020). Internal and external effects of social distancing in a pandemic, *Working Paper 27059*, NBER, Cambridge, Massachusetts.

Ferguson, N. M., Laydon, D., Nedjati-Gilani, G. and collaborators (2020). Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand, *Report 9*, Imperial College, London.

Gersovitz, M. and Hammer, J. S. (2004). The economical control of infectious diseases, *Economic Journal* **114**: 1–27.

Giannitsarou, C., Kissler, S. and Toxvaerd, F. (2020). Waning immunity and the second wave: Some projections for SARS-CoV-2, *Discussion Paper 14852*, CEPR, London, UK.

Gonzalez-Eiras, M. and Niepelt, D. (2020a). On the optimal 'lockdown' during an epidemic, *Covid Economics* **7**: 68–87.

Gonzalez-Eiras, M. and Niepelt, D. (2020b). Tractable epidemiological models for economic analysis, *Discussion Paper 14791*, CEPR.

Goolsbee, A. and Syverson, C. (2020). Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020, *Working Paper 27432*, NBER, Cambridge, Massachusetts.

Greenstone, M. and Nigam, V. (2020). Does social distancing matter?, *Covid Economics* **7**: 1–23.

Gudbjartsson, D. F., Norddahl, G. L., Melsted, P., Gunnarsdottir, K., Holm, H., Eythorsson, E., Arnthorsson, A. O., Helgason, D., Bjarnadottir, K., Ingvarsson, R. F., Thorsteinsdottir, B., Kristjansdottir, S., Birgisdottir, K., Kristinsdottir, A. M., Sigurdsson, M. I., Arnadottir, G. A., Ivarsdottir, E. V., Andresdottir, M., Jonsson, F., Agustsdottir, A. B., Berglund, J., Eiriksdottir, B., Fridriksdottir, R., Gardarsdottir, E. E., Gottfredsson, M., Gretarsdottir, O. S., Gudmundsdottir, S., Gudmundsson, K. R., Gunnarsdottir, T. R., Gylfason, A., Helgason, A., Jensson, B. O., Jonasdottir, A., Jonsson, H., Kristjansson, T., Kristinsson, K. G., Magnusdottir, D. N., Magnusson, O. T., Olafsdottir, L. B., Rognvaldsson, S., le Roux, L., Sigmundsdottir, G., Sigurdsson, A., Sveinbjornsson, G., Sveinsdottir, K. E., Sveinsdottir, M., Thorarensen, E. A., Thorbjornsson, B., Thordardottir, M., Saemundsdottir, J., Kristjansson, S. H., Josefsdottir, K. S., Masson, G., Georgsson, G., Kristjansson, M., Moller, A., Palsson, R., Gudnason, T., Thorsteinsdottir, U., Jonsdottir, I., Sulem, P. and Stefansson, K. (2020). Humoral immune response to SARS-CoV-2 in Iceland, *The New England Journal of Medicine* **forthcoming**.

Hall, R. E., Jones, C. I. and Klenow, P. J. (2020). Trading off consumption and covid-19 deaths, *Working Paper 27340*, NBER, Cambridge, Massachusetts.

Hethcote, H. W. (1989). Three basic epidemiological models, *in* L. Gross, T. G. Hallam and S. A. Levin (eds), *Applied Mathematical Ecology*, Springer, Berlin, pp. 119–144.

Hethcote, H. W. (2000). The mathematics of infectious diseases, *SIAM Review* **42**(4): 599–653.

Jones, C. J., Philippon, T. and Venkateswaran, V. (2020). Optimal mitigation policies in a pandemic: Social distancing and working from home, *Working Paper 26984*, NBER, Cambridge, Massachusetts.

Kaplan, G., Moll, B. and Violante, G. L. (2020). The great lockdown and the big stimulus: Tracing the pandemic possibility frontier for the U.S., *Working Paper 27794*, NBER, Cambridge, Massachusetts.

Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics, *Proceedings of the Royal Society, Series A* **115**(772): 700–721.

Menachemi, N., Yiannoutsos, C. T., Dixon, B. E., Duszynski, T. J., Fadel, W. F., Wools-Kaloustian, K. K., Unruh Needleman, N., Box, K., Caine, V., Norwood, C., Weaver, L. and Halverson, P. K. (2020). Population point prevalence of SARS-CoV-2 infection based on a statewide random sample—Indiana, *Morbidity and Mortality Weekly Report, April 25–29* **69**: 960–964.

Miclo, L., Spiro, D. and Weibull, J. W. (2020). Optimal epidemic suppression under an ICU constraint. Unpublished, Stockholm School of Economics.

Russell, T. W., Hellewell, J., Jarvis, C. I., van Zandvoort, K., Abbott, S., Ratnayake, R., CMMID COVID-19 working group, Flasche, S., Eggo, R. M. and Kucharski, A. J. (2020). Estimating the infection and case fatality ratio for COVID-19 using age-adjusted data from the outbreak on the Diamond Princess cruise ship, *Eurosurveillance* **25**(12).

To, K. K.-W., Hung, I. F.-N., Ip, J. D., Chu, A. W.-H., Chan, W.-M., Tam, A. R., Fong, C. H.-Y., Yuan, S., Tsoi, H.-W., Ng, A. C.-K., Lee, L. L.-Y., Wan, P., Tso, E., To, W.-K., Tsang, D., Chan, K.-H., Huang, J.-D., Kok, K.-H., Cheng, V. C.-C. and Yuen, K.-Y. (2020). COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing, *Clinical Infectious Diseases* **forthcoming**.

Toxvaerd, F. (2020). Equilibrium social distancing, *Working Paper 2020/08*, Cambridge Institute for New Economic Thinking, Cambridge.

A Calibration Strategy

A.1 Law of Motion (1)

In this appendix, we describe how we use information about parameter values in the canonical SIR model to deduce parameter values for the law of motion (1).[28]

A.1.1 Canonical SIR Model

The canonical SIR model due to Kermack and McKendrick (1927) specifies laws of motion for the population shares of three groups: the "susceptible," the "infected" or "infectives," and the "removed." Their respective population shares at time $t \geq 0$ are denoted by $x(t)$, $y(t)$, and $z(t)$, respectively, where $x(t) + y(t) + z(t) = 1$.[29] We normalize the mass of the total population at time $t = 0$ to unity.

At time $t = 0$ the population consists of $x(0)$ susceptible persons and a few infected persons, $y(0)$. There are no removed persons at this time, $z(0) = 0$. In each instant after time $t = 0$, infected persons transmit the disease to members of the susceptible group and a share of the infected either dies or recovers and develops immunity. Formally,

$$\dot{x}(t) = -b(t)x(t)y(t), \tag{13}$$
$$\dot{y}(t) = -\dot{x}(t) - (c^d + c^r)y(t), \tag{14}$$
$$\dot{z}(t) = (c^d + c^r)y(t). \tag{15}$$

Here, $b(t)$ denotes a possibly time-varying infection rate. The extent to which susceptible persons are infected depends on their number, $x(t)$; the infection rate, $b(t)$; and the population share of infected persons. The number of infected persons increases one-to-one with the susceptible persons that get infected, while a share $c \equiv c^d + c^r$ of the infected population dies or recovers; the coefficients c^d and c^r parameterize the flow into death and recovery, respectively.

Consider the case where $b(t)$ is constant at value b. Inspection of equations (13) and (14) reveals that for $bx(0) > c$ the share of infected persons increases until it reaches a maximum when $x(t) = c/b$; thereafter, the share declines. Intuitively, when $x(0)$ falls short of c/b (the "herd immunity level") then there are fewer new infections of susceptible persons than outflows from the infected pool due to recoveries and death. As is well known (e.g., Theorem 2.1 in Hethcote, 2000), $x(\infty)$ falls short of the herd immunity level unless $x(0) = c/b = x(\infty)$ and $y(0) = 0$.[30]

In the SIR-S model a share γ of the removed population loses immunity and moves

[28] See Gonzalez-Eiras and Niepelt (2020b) for a broader and more detailed discussion, also with respect to other SIR models.

[29] We follow the notation introduced by Kermack and McKendrick (1927).

[30] Note also, from equation (14), that at the beginning of an epidemic with $x(t) \approx 1$ and $z(t) \approx 0$, b approximately equals the growth rate of the number of persons who are or were infected, $\frac{\dot{y}(t)+\dot{z}(t)}{y(t)+z(t)} = b\frac{x(t)y(t)}{y(t)+z(t)} \approx b$.

back to the susceptible pool. Accordingly, the dynamic system is given by

$$\begin{aligned}\dot{x}(t) &= -b(t)x(t)y(t) + \gamma z(t),\\ \dot{y}(t) &= b(t)x(t)y(t) - cy(t),\\ \dot{z}(t) &= cy(t) - \gamma z(t).\end{aligned}$$

In steady state this reduces to

$$\gamma z = bxy = cy.$$

Calibration We measure time in days and use information about the spread of COVID-19 in the United States to calibrate the model. We associate time $t = 0$ with mid March 2020, the date around which public health authorities considered to impose restrictions. We assume that at this time, $z(0)$ equalled practically nil.

Following Atkeson (2020) and the sources cited therein we assume that the flow rate from the infected to the removed population equals $c = 1/18$, corresponding to an exponentially distributed infection duration that averages 18 days.[31]

From Russell et al. (2020), Greenstone and Nigam (2020), and the sources cited therein we infer that the inverse of the infection fatality rate, c/c^d, lies in the range $[100, 200]$.

To calibrate $y(0)$ we use data on COVID-19 deaths in mid March 2020 as well as information about c^d and c. The number of deaths on March 16 equalled 23.[32] Based on equation (15) we infer that the initial share of the infected population in mid March, $y(0)$, equalled $1.8933 \cdot 10^{-4}$.[33] This compares to a reported case count of 4507, corresponding to a population share of $1.3745 \cdot 10^{-5}$.[34]

Finally, to calibrate b we rely on information in Ferguson et al. (2020) who argue that the "basic reproduction number" $\mathcal{R}_0 = b/c$ for COVID-19 equals approximately 2.4 which implies $b = 0.1333$. When we simulate the canonical SIR model subject to these parameter values we find that infections peak after roughly 114.34 days, on 7 July 2020. We use this date below.

A.1.2 Logistic Model

The law of motion (1) (subject to $\gamma = 0$) follows from the canonical SIR model (and other related models, see the discussion in Gonzalez-Eiras and Niepelt, 2020b) by letting

[31] Note that $\int_0^\infty c\mathrm{e}^{-ct}t\ dt = 1/c$.

[32] See `https://github.com/nytimes/covid-19-data/blob/master/us.csv`. Regressing the full set of March data on an exponential trend generates a similar point estimate for March 16.

[33] We have $y(0) \cdot (\text{US population}) = (\text{new deaths})/c^d = (\text{new deaths})/c \cdot c/c^d$. We use US population = 328 million, new deaths = 23, and $c/c^d = 150$.

[34] See `https://github.com/nytimes/covid-19-data/blob/master/us.csv`. The reported number corresponds to the cumulative case count but there are very few removed cases at the time. Common estimates of the extent of underreporting suggest a factor of ten, in line with our results; see, e.g., `https://www.medrxiv.org/content/10.1101/2020.03.14.20036178v2.full.pdf+html` or `https://fondazionecerm.it/wp-content/uploads/2020/03/Using-a-delay-adjusted-case-fatality-ratio-to-estimate-under-reporting-_-CMMID-Repository.pdf` or `https://www.npr.org/sections/coronavirus-live-updates/2020/06/25/883520249/cdc-at-least-20-million-americans-have-had-coronavirus-heres-who-s-at-highest-ri`.

$c^d = c^r = 0$ such that $z(t) = 0$, and by letting $b(t) = g(a(t))\beta$. Variable $y(t)$ now has the interpretation of the stock of persons who underwent infection in the past, not of the number of currently infected persons. For $g(a(t)) = 1$ the law of motion (1) (subject to $\gamma = 0$) implies a logistic path for $y(t)$ that converges to $\bar{y}$,

$$y(t) = \frac{\bar{y}}{1 + \mathrm{e}^{-\beta\bar{y}t}(\bar{y}/y(0) - 1)}.$$

Here, $\bar{y}$ has the interpretation of $1 - x(\infty)$ where $x(\infty)$ denotes the long-run share of susceptible individuals in the canonical SIR model who do not contract the disease.

We have the following standard result:

Proposition 6. Consider the law of motion (1) with $\gamma = 0$, $g(a(t)) = 1$, and $y(0) < \bar{y} \leq 1$. Then, $\dot{y}(t)$ reaches a maximum at

$$t = \ln\left(\frac{\bar{y} - y(0)}{y(0)}\right)/(\beta\bar{y}).$$

Proof. Solving $\ddot{y}(t) = 0$ (or $y(t) = \bar{y} - y(t)$) for t yields the result. □

Calibration Following Hall et al. (2020) who in turn rely on Ferguson et al. (2020) we assume that 75 percent of the population would contract the disease eventually in the absence of any mitigation measures, $\bar{y} = 0.75$. Moreover, we use proposition 6 to infer the value of β that corresponds to the b value in the canonical SIR model such that both models predict peak infections at the same date. That is, we choose β such that the t value in proposition 6 corresponds to 7 July 2020. Using $y(0) = 1.8933 \cdot 10^{-4}$ and $\bar{y} = 0.75$, this yields $\beta = 0.9660 \cdot 10^{-1}$.

A.2 Costs of Infection

In this appendix, we discuss the calibration of the parameters representing private and social costs of infection. Recall from proposition 1 that $\xi_{a_i}(a, a) = \xi_{a_i}(1, 1)/a = \zeta/a$ such that the private marginal costs equal $\xi_{a_i}(a, a)\psi\dot{y}^g(y) = \zeta\psi\dot{y}^g(y)/a$ while the social marginal costs equal $\psi d\dot{y}^g(y)/da = \psi n\dot{y}^g(y)/a$. Accordingly, the ratio of private and social marginal costs equals ζ/n.

We calibrate ζ/n based on U.S. estimates of hospitalization costs and the value of life by Bartsch et al. (2020) and Hall et al. (2020), respectively. Bartsch et al. (2020) estimate direct medical costs including follow up expenses (over a year) of $1.25 trillion under the assumption that eighty percent of the U.S. population are infected. This translates into conditional per-capita costs of about $4,764 (eighty percent of 328 million persons). Hall et al. (2020) assess the value of life at $270,000 per year. With an average remaining life expectancy of 14.5 years every life lost to COVID-19 thus costs $3,915,000. Menachemi et al. (2020) estimate an infection fatality rate of 0.58 percent implying a conditional expected cost of dying from COVID-19 of $22,707 per infected individual. Under the assumption that individuals fully internalize mortality risk but not marginal social medical costs we conclude that $\zeta/n = 22,707/(22,707 + 4,764) \approx 0.8266$.

To calibrate ψ based on the dollar amount \$22,707 + \$4,764 we use Hall et al.'s (2020) estimate according to which households would be willing to sacrifice 28 percent of consumption to eliminate all COVID-19 related mortality risk over one year (neglecting other costs). Let $1-\phi = 0.28$ denote this share. In the model the utility cost of sacrificing the share $1-\phi$ of consumption during N days equals[35]

$$N \cdot \sigma \left\{(1 + \ln(a^\star) - a^\star) - (1 + \ln(a^\star\phi) - a^\star)\right\} = -N\sigma \ln(\phi).$$

Suppose that, effectively, all infections occur during the first N days of the epidemic such that $\int_0^N \dot{y}(t)dt \approx \bar{y}$ and the total (undiscounted) mortality costs amount to $\hat{\psi}\bar{y}$. We conclude that

$$\hat{\psi} \approx -N\sigma \ln(\phi)/\bar{y}.$$

With $N = 365$ and $\bar{y} = 0.75$ this implies social costs due to mortality risk of $\hat{\psi} \approx \sigma 159.9$. Adding medical costs we arrive at an estimate for ψ of $\psi = \sigma\hat{\psi}n/\zeta \approx \sigma 193.4$.

[35] We neglect time discounting as do Hall et al. (2020). Note that only the benefit of economic activity ("consumption") not the cost associated with it ("labor supply") is reduced by the fraction $1-\phi$.

B Proofs

B.1 Proof of Lemma 1

Proof. To prove continuity at $y = 0$, we consider without loss of generality the case of $\nu = 0$. Consider y_0 and y_1 with $0 < y_0 < y_1 \leq s(a^\star)$ where $s(a^\star)$ denotes the steady-state level of y when activity equals $a^\star$ (see equation (2)). Let $\mathcal{C}(a^\star, y)$ denote the costs of infection when activity equals $a^\star$ and the state is given by (y, t). Define

$$\tilde{V}(y_0, t_0) = \int_0^{t(y_0,y_1)} \mathrm{e}^{-\rho\,\tau} u(a^\star) d\tau - \int_{y_0}^{y_1} \mathrm{e}^{-\rho\, t(y_0,y)} \mathcal{C}(a^\star, y) dy + \mathrm{e}^{-\rho\, t(y_0,y_1)} V(y_1, t_0 + t(y_0, y_1)).$$

Recall that $t(y_0, y)$ denotes the time span over which the state variable moves from y_0 to y. $\tilde{V}(y_0, t_0)$ represents the value at time t_0 conditional on $y = y_0$ when activity is fixed at level $a^\star$ until $y = y_1$ is reached, from which point on activity is chosen optimally. By construction, $\tilde{V}(y_0, t_0)$ constitutes a lower bound for $V(y_0, t_0)$ such that $\tilde{V}(y_0, t_0) \leq V(y_0, t_0) \leq U^\star$. Note that $\mathcal{C}(a^\star, y)$ is bounded for $y \in [0, \bar{y}]$. Also, from equation (1), $\lim_{y\downarrow 0} \mathcal{C}(a^\star, y) = 0$. Finally, from equation (3), $\lim_{y_0\downarrow 0} t(y_0, y) = \infty$. It follows that $\lim_{y_0\downarrow 0} \tilde{V}(y_0, t_0) = U^\star$ and therefore $\lim_{y_0\downarrow 0} V(y_0, t_0) = U^\star$. □

B.2 Proof of Lemma 2

Proof. From the government's HJB equation, the envelope condition satisfies

$$(\rho + \nu)V'(y) = -g(a(y))\beta \left[(\bar{y} - 2y)(\psi - V'(y)) - y(\bar{y} - y)V''(y)\right].$$

Let $\hat{y}$ denote a point where V reaches a local minimum or maximum. Evaluated at $\hat{y}$ the HJB equation reduces to

$$(\bar{y} - 2\hat{y})\psi = \hat{y}(\bar{y} - \hat{y})V''(\hat{y}). \tag{16}$$

To see that V has a unique minimum suppose to the contrary that there exist multiple local minima. Consider two neighboring minima at, say, y^a and y^c with $y^a < y^c$. Then there must exist a local maximum at some y^b with $y^a < y^b < y^c$. Note that $V''(y^a) > 0$, $V''(y^b) < 0$, and $V''(y^c) > 0$. Evaluating equation (16) at y^a, y^b, and y^c the right-hand side of equation (16) thus changes signs twice while the left-hand side changes signs at most once. We have thus arrived at a contradiction which proves that V has a unique minimum, $y^{\min}$.

To see that $y^{\min} \leq \bar{y}/2$ suppose to the contrary that $y^{\min} > \bar{y}/2$. Since the minimum is unique we have $V'(\bar{y}/2) < 0$ and thus, from the envelope condition, $(\rho + \nu)V'(\bar{y}/2) = g(a(\bar{y}/2))\beta(\bar{y}/2)^2 V''(\bar{y}/2)$ such that $V''(\bar{y}/2) < 0$. Since the minimum lies to the right of $\bar{y}/2$ there must exist an inflection point, say y^i, with $\bar{y}/2 < y^i < y^{\min}$, $V'(y^i) < 0$, and $V''(y^i) = 0$. But evaluated at y^i the envelope condition implies

$$\underbrace{(\rho + \nu)V'(y^i)}_{<0} = -\underbrace{g(a(y^i))\beta}_{>0}\underbrace{(\bar{y} - 2y^i)}_{<0}\underbrace{(\psi - V'(y^i))}_{>0},$$

which yields a contradiction. We conclude that $y^{\min} \leq \bar{y}/2$.

From equation (16), $V''(\hat{y}) = (\bar{y} - 2\hat{y})\psi/\hat{y}/(\bar{y} - \hat{y})$ which is strictly decreasing in $\hat{y}$. This proves the last claim. □

B.3 Proof of Proposition 1

Proof. From assumption 3, $\xi_{a_i}(a_i, a)$ is homogenous of degree m say. Accordingly, $\xi(a_i, a)$ is homogeneous of degree $m + 1$. Since $\xi(\lambda a, \lambda a) = 1$ for all $\lambda \neq 0$, m equals -1: $\xi_{a_i}(a, a) = \lambda \xi_{a_i}(\lambda a, \lambda a)$. Letting $\lambda = 1/a$ implies $\xi_{a_i}(a, a) = \xi_{a_i}(1, 1)/a \equiv \zeta/a$ and yields the first result. The derivations also imply that $\zeta - n$ is proportional to the static externality, establishing the second result. □

B.4 Proof of Proposition 2

Proof. The result follows directly from equation (10) in proposition 1. □

B.5 Proof of Proposition 3

Proof. Part i. follows because the value function is decreasing in a neighborhood of $y = 0$ (from lemma 1), implying that the static and dynamic externalities both are negative such that the government imposes a lockdown. Part ii. follows from the fact that $V_y(y, t)$ eventually is positive. For "small" $1-\zeta/n$ the total externality therefore eventually turns positive. As a consequence, the government imposes an inverse lockdown if and when the economy reaches the relevant part of the state space (requiring the condition on γ). If V is locally convex then the total externality switches signs at y^0, establishing part iii. □

B.6 Proof of Proposition 4

Proof. Uniqueness of equilibrium follows by construction. Differentiating the equilibrium solution with respect to γ implies that in a neighborhood of $\gamma = 0$ (and $a = 1$), $da/d\gamma|_{\gamma=0} = -\psi\bar{y}\zeta/\sigma$. For the government's solution, equation (11) implies $da/d\gamma|_{\gamma=0} = -\psi(\rho + \nu)\bar{y}/(\sigma(\rho + \nu + \beta\bar{y}))$. With $\gamma = 0$ the two steady-state values are identical. We conclude that for small values of γ the government chooses a higher steady-state activity level (and, from equation (1), a higher y) than in decentralized equilibrium if

$$(\rho + \nu)(1 - \zeta) < \zeta\beta\bar{y}.$$

□

B.7 Proof of Proposition 5

Proof. Differentiating the first-order condition, $\mathcal{H}_a(a(t), y(t), t) = 0$, with respect to time implies

$$\dot{a}(t)\frac{d}{d\,a(t)}\frac{u'(a(t))}{g'(a(t))} = \beta(\bar{y} - 2y(t))\dot{y}(t)\,(\psi - \mu(t)) - \beta y(t)(\bar{y} - y(t))\dot{\mu}(t).$$

The right-hand side of this condition collapses to zero when we substitute for $\dot{y}(t)$ from the law of motion and for $\dot{\mu}(t)$ from the condition $\mathcal{H}_y(a(t), y(t), t) = -\dot{\mu}(t)$. We conclude that $\dot{a}(t) = 0$.

The solution is characterized by the following system of equations in two unknowns, a and $y_T \equiv y(T)$ (for simplicity we set $\sigma = 1$):

$$a = \frac{1}{1 + \beta y_T(\bar{y} - y_T)\psi}, \tag{17}$$

$$y_T = \frac{\bar{y}}{1 + e^{-a\beta\bar{y}T}\left(\frac{\bar{y}}{y(0)} - 1\right)}. \tag{18}$$

Here, we have used the government's first-order condition at T and the fact that $\mu(T) = 0$. Note that equation (17) implies a negative relation between a and y_T if $y_T < \bar{y}/2$, and a positive relation if $y_T > \bar{y}/2$.

For sufficiently small (y_0, T), y_T necessarily lies below $\bar{y}/2$ even if $a = a^\star$. A small increase in T then reduces a and increases y_T. (From equation (17), $y_T < \bar{y}/2$ implies $\text{sign}(\frac{da}{dT}) \neq \text{sign}(\frac{dy_T}{dT})$; moreover, an increase in T cannot lead to higher activity and lower infections.) We claim that at some point, responding in this way cannot be optimal any longer. Suppose otherwise and consider the value for T at which y_T would equal $\bar{y}/2$ and correspondingly, activity would equal $a = \underline{a} \equiv \frac{1}{1+\beta\frac{\bar{y}^2}{4}\psi}$. From equation (18) this is given by

$$\hat{T} = \frac{\ln(\bar{y}/y(0) - 1)}{\underline{a}\beta\bar{y}}.$$

This policy would yield the welfare

$$(1 + \ln(\underline{a}) - \underline{a})\hat{T} - \psi\bar{y}/2,$$

which under condition (12) is lower than $-\psi\bar{y}$. In contrast, following the policy $a = a^\star$ generates welfare $-\psi\bar{y}$. We have thus arrived at a contradiction, which proves the proposition. □