Road to recovery: Managing an epidemic*

Simon Loertscher† Ellen V. Muir‡

This version: July 18, 2020 First version: 23 April 2020

Abstract

Without widespread immunization, the road to recovery from the current COVID-19 lockdowns will optimally follow a path that finds the difficult balance between the social and economic benefits of liberty and the toll from the disease. We provide an approach that combines epidemiology and economic models by taking as given the constraint that the maximum capacity of the healthcare system must not be exceeded. Treating the transmission rate as a decreasing function of the severity of the lockdown, we first determine the minimal lockdown that satisfies this constraint, using an epidemiological model with a homogeneous population to predict future demand for healthcare. Allowing for a heterogeneous population, we then derive the optimal lockdown policy under the assumption of homogeneous mixing and show that it is characterized by a bang-bang solution.

Keywords: COVID-19, SIR models, capacity constraints, managing an epidemic
JEL-Classification: H51, I18

*We are grateful for suggestions and comments by Gabriel Carroll, Chris Edmond, Ian Harper, Zi Yang Kang, Paul Milgrom, Martin Souchier, Gary Stoneham and seminar audiences at the COVID-19 Policy Hackathon organized by the Stanford Economic Association and the MIT Undergraduate Economic Association and at the University of Melbourne. Anand Bharadwaj provided excellent research assistance. Financial support by the June and Samuel Hordern Endowment is also gratefully acknowledged.

†Department of Economics & Centre for Market Design, Level 4, FBE Building, 111 Barry Street, University of Melbourne, Victoria 3010, Australia. Email: simonl@unimelb.edu.au.

‡Department of Economics, Stanford University. Email: evmuir@stanford.edu. Corresponding author.

1 Introduction

Without widespread immunization of the population (for example, due to the development of a vaccine), the road to recovery from the ongoing coronavirus-induced lockdowns will require sustained vigilance to ensure that the spread of the virus remains at a manageable level for a country's or region's healthcare system. At the same time, this recovery ought to start as soon as possible to limit the curtailing of liberty that these lockdowns impose, the mental and other health issues associated with social distancing and isolation, and to minimize the associated economic cost that already has little parallel in living memory. If eradication is impossible or possible only at tremendous costs, keeping the coronavirus pandemic under control without inducing economic and social hardship of a catastrophic scale thus requires finding a path through territory that is uncharted for both epidemiologists and economists. From a public health perspective, recovery requires the transition from a paradigm in which eradication of an epidemic is the goal to one in which the epidemic is *managed*. For economists, recovery requires ploughing a path through a system whose dynamics are non-linear.

In this paper, we show how this can be done by providing a methodology that permits return to some kind of normalcy while keeping the spread of the virus at a level that even at the peak of the endemic does not exceed the capacity constraint of the healthcare system. Specifically, we use a standard epidemiology model—a simple *SIR model*—to predict the peak of the epidemic while treating the rate of transmission as the variable that the policymaker can influence by choosing the severity of a lockdown. We treat as a hard constraint the capacity of the healthcare system, that is, the maximum number of COVID-19 patients that it can handle per period of time at the peak of the crisis.[1] Of course, this capacity constraint will need to be defined in such a way that patients with other—but no less severe—needs for care are still able to access treatment.[2]

The main contribution of this paper is to formulate an operational constraint that provides policymakers with guidance for how to manage an epidemic which is too costly to

[1]To fix ideas, throughout the paper we will talk about the capacity of the healthcare system as our binding constraint. However, the framework outlined in this paper can accommodate any constraint that can be expressed as a function of the number of COVID-19 cases that occur at the peak of the epidemic. This constraint can be interpreted as a normative criterion that society has to choose. For example, suppose that society viewed the number of deaths that would occur as a result of using the capacity of the healthcare system as a binding constraint as unacceptable. Then one could instead treat a fixed proportion of the healthcare system that is utilized at the height of the crisis as a binding constraint (such as requiring that the healthcare system never exceeds 80% capacity). An alternative, but mathematically equivalent, approach would be to place an upper bound on the number of deaths per day at the peak of the crisis, which may be in line with some of the concurrent debates (see, e.g., the New York Times article "The Cold Calculations America's Leaders Will Have to Make Before Reopening").

[2]In a public health catastrophe, this is not always the case; see, for example, this New York Times article: "The Pandemic's Hidden Victims: Sick or Dying, but Not From the Virus'.'

eradicate; to incorporate this constraint into a standard epidemiology model; and derive implications for the severity of the lockdown that is necessary to respect the constraint. Allowing for heterogeneity in the population, we also show that managing an epidemic subject to a capacity constraint involves a non-trivial economic optimization problem without requiring the policymaker to take a stance on the value of life because optimality requires satisfying the constraint at minimum economic cost. While the purpose of our model is to serve as a proof of concept that would need to be refined if applied, many of the key insights—such as the need to use epidemiology models to predict future healthcare demand and the non-trivial economic optimization problem when faced with a capacity constraint—will extend well beyond the confines of the specific setups we study.

The data required to apply the proposed approach are accurate estimates of the transmission rate, reliable measures of the spread of the disease, and the percentage of the infected persons that require treatment. Some of these data can and will need to be obtained using ongoing random sampling of the population. Importantly, for practical purposes what we propose can be implemented using continuous, real-time updating and adjustments. To the extent that the population is continuously and randomly sampled and tested, decision makers will have a reasonably clear real-time picture of the spread of the virus, which help mitigate problems with time-lags and inertia.

To convey a sense of the magnitude of the potential economic and social costs, consider the unemployment rate during the Great Depression in the U.S., then and now the world's largest economy, and the unemployment rates before and in the wake of the ongoing coronavirus-related lockdowns in the U.S. in 2020. The immediate consequences of the Great Depression were mass poverty and economic devastation, and at least indirectly, the rise of fascism in Europe. As Table 1 shows, the unemployment rate in the U.S. rose sharply from 5.2% in March 2020 to 19.5% in April 2020 as the nationwide lockdown hit the country and much of the rest of the world economy, and then steadily declined as the began to reopen. This steep incline and swift partial recovery reflects the peculiarity of the present economic downturn, which was *not* caused by a bad state of the economy. This is at the same time a source of hope and of concern: while the healthy underlying state of the economy at the onset may make for a relatively fast recovery, extended or repeated complete lockdowns can turn a public health shock into a deep and prolonged economic crisis. The firms workers could return to in May and June may simply go out of business after further or extended lockdowns. Thus, the problem of finding a smooth path to recovery is particularly salient.

Not surprisingly, there has been a recent upsurge of interest in SIR models in economics. Atkeson (2020) provides an introduction of this modeling approach to economics, which is standard in mathematical biology (see, e.g., Murray, 2002). Alvarez et al. (2020) apply an

Great D.	1929	1930	1931	1932	1933
Unemployment rate[3]	3.2	8.7	15.9	23.6	24.9

2020	Mar 20	Apr 20	May 20	Jun 20
Unemployment rate[4]	5.2	19.5	16.4	11.1

Table 1: Upper table: unemployment rates during the Great Depression in the U.S. Lower table: weekly unemployment filings (in thousands) in the U.S. in 2020.

optimal control approach to an SIR model to derive the optimal lockdown policy that trades off the cost of death against economic output. In contrast to Alvarez et al. (2020) and most of the recent macro literature, in our approach the optimal lockdown policy minimizes the economic impact of satisfying the capacity constraint of the healthcare system. This constraint is absent from the existing models yet highly relevant in reality—exceeding it can induce catastrophic health outcomes such as New York City or Lombardy have experienced earlier this year. For example, during the peak of the crisis, New York City had six times the number of deaths it would otherwise have.[5] While their paper does not account for capacity constraints, Akbarpour et al. (2020) use simulations in an agent-based model to study optimal policies at a level of detail that our model does not permit. There has also been a recent rise in more informal commentary and discussions of the problems at hand (see, e.g., Gilbert et al., 2020) and analyses of tradeoffs involving economics without explicitly embedding epidemiology models such as Hall et al. (2020) and Budish (2020).

The remainder of this paper is organized as follows. Section 2 describes the setups. Section 3 derives the dynamics of an epidemic and the lockdown necessary to keep it at a level that respects the capacity constraint in a homogeneous population model. In Section 4 we derive the optimal lockdown policy for a model with a heterogeneous population and show that the optimal policy takes a bang-bang form if mixing is homogeneous. Section 5 provides a discussion of possible and natural extensions of our baseline model. Section 6 concludes the paper. All proofs can be found in the appendix.

[5]Even though the absolute numbers of deaths from COVID19 look staggering in many countries, as percentages of the number of people who die under normal circumstances during the same period, they are typically well within the range of the normal seasonal variability in deaths due to the flu and other factors whose impact varies with the season. This further underlines the importance of capacity constraints: If these can be satisfied at all times, catastrophic health outcomes can be averted. And it is, of course, a compelling reason to explicitly model such constraints.

2 Setup

Consider a basic susceptible-infectious-recovered (SIR) model with a constant population of size N. This is a classic model in epidemiology (see, e.g., Murray, 2002, p.320), in which the population is divided into three compartments consisting of susceptible individuals, infected individuals and recovered individuals, respectively denoted by $S(t)$, $I(t)$ and $R(t)$ at time t. Note that because of the assumption of a constant population of size N, for all $t \geq 0$, we have

$$S(t) + I(t) + R(t) = N.$$

We let $N_1 = S(0)$, $N_2 = I(0)$ and $N_3 = R(0)$ and assume that only two types of transitions are possible: susceptible individuals can become infected and infected individuals recover. (As is standard, "recovered" simply means the individuals are no longer infectious, which occurs either because they gained immunity or died following infection.) We let β denote the average number of contacts per person per time and assume that we have a well-mixed or homogeneous population so that $I(t)/N$ is the fraction of contact occurrences that involve an infectious individual. The *rate of transition* between the susceptible compartment and the infectious compartment is thus given by $\beta I(t)/N$.

We denote by $\ell \in [0, 1]$ the severity of the lockdown, with $\ell = 0$ meaning no lockdown and $\ell = 1$ meaning complete lockdown. We assume that ℓ is the choice variable of the policymaker, and with regards to the epidemic, its impact is that it affects the transmission rate β as follows:

$$\beta(\ell) = \beta_0 + (1 - \ell)\beta_1,$$

where $\beta_0 \geq 0$ is a fixed component of the transmission rate, $\beta_1 > 0$ is a constant, and $\beta(\ell)$ makes the dependence of β on ℓ explicit.

We further assume that individuals recover at rate γ.[6] In SIR models, the parameter $R_0 = \beta/\gamma$ plays an important role in governing the dynamics of an epidemic. In this simple version whenever $R_0 N_1/N > 1$ the number of infected individuals will increase from time $t = 0$, resulting in an *epidemic*. If $R_0 N_1/N < 1$ then the number of infected individuals will decrease from time $t = 0$ and an epidemic does not occur (alternatively, we can think of the "peak" of the epidemic as occurring at time $t = 0$).

The proportion $\tau \in [0, 1]$ of those who are infected need *treatment*, so that, given $I(t)$

[6] For example, if D is the duration of infection then the rate of recovery is given by $\gamma = 1/D$.

and τ, the number of people requiring treatment at time t is

$$T(t) = \tau I(t).$$

Letting $K > 0$ denote the maximum capacity of the healthcare system to treat COVID-19 patients without reducing the care given to other patients in need, the constraint for managing the epidemic is, for all $t \geq 0$,

$$T(t) \leq K. \tag{1}$$

In the following section, we augment the epidemiology model by an economic production function to analyze tradeoffs involving economics. Specifically, we assume that GDP, denoted Y, is produced using labor L according to the production function $Y = L^\alpha$, where $\alpha \in (0,1)$ is a parameter that measures labor's productivity, which can be calibrated using labor's income share in national accounts data.[7] Letting $L_0 \geq 0$ denote the amount of labor that is not affected by the lockdown variable ℓ, the amount of labor that is productive given ℓ is

$$L(\ell) = L_0 + (1 - \ell)L_1,$$

where L_1 is the part of the labor that is affected by the lockdown variable ℓ. It follows that $L(0)$ is the pre-lockdown labor supply.

In Section 4, we will also consider a *heterogeneous agent* version of the model.[8]. Specifically, we assume that there is a continuum of types $\theta \in [\underline{\theta}, \overline{\theta}]$ in the population.[9] We denote by F the absolutely continuous distribution of types in the population and by $S(\theta, t)$, $I(\theta, t)$ and $R(\theta, t)$ the density of susceptible, infected and recovered individuals, respectively, of type θ at time t. The density of individuals $N(\theta)$ of type θ thus satisfies, for all $t \geq 0$,

$$N(\theta) = S(\theta, t) + I(\theta, t) + R(\theta, t).$$

We let $N_1(\theta) = S(\theta, 0)$, $N_2(\theta) = I(\theta, 0)$ and $N_3(\theta) = R(\theta, 0)$.

We assume that the type of any given individual is observable and that the policymaker can implement a type-dependent lockdown policy $\ell : [\underline{\theta}, \overline{\theta}] \to [0,1]$, where $\ell(\theta)$ denotes the severity of the lockdown for individuals of type θ, with $\ell(\theta) = 0$ meaning no lockdown and $\ell(\theta) = 1$ meaning complete lockdown. Similarly to the basic SIR model, we impose

[7]This corresponds to assuming a Cobb-Douglas production function with all input factors other than labor being fixed for the duration of the disease; see, for example, Jehle and Reny (2011).

[8]See, for example (Murray, 2002, p.320) for an age-structured SIR model

[9]One can think of the type of an agent as encompassing the relevant observable characteristics of that agent, such as age, gender, occupation and underlying health conditions.

a homogeneous mixing assumption. The transmission rate $\beta_\ell(\theta)$ for individuals of type θ under the lockdown policy ℓ is then given by

$$\beta_\ell(\theta) = \beta_0 + (1 - \ell(\theta))\beta_1.$$

For simplicity, we assume that the parameters β_0 and β_1 are independent of θ. Note that in the heterogeneous agent model we make the dependence of the evolution of the epidemic on the lockdown policy ℓ explicit by introducing a subscript ℓ to every variable in the model that depends on ℓ.

We assume that $\tau(\theta) \in [0, 1]$ is the proportion of infected individuals of type θ that require *treatment*. Given a lockdown policy ℓ, the total number of individuals $T_\ell(t)$ that require treatment at time t is given by

$$T_\ell(t) = \int_{\underline{\theta}}^{\overline{\theta}} \tau(\theta) I_\ell(\theta, t)\, dF(\theta).$$

Similarly to the homogeneous agent model, we augment the heterogeneous agent model by an economic production function to analyze tradeoffs involving economics. Specifically, if $L_\ell(\theta)$ denotes labor supplied by individuals of type θ under lockdown policy ℓ, then output produced by individuals of type θ is given by

$$Y_\ell(\theta) = A(\theta) L_\ell(\theta),$$

where $A(\theta)$ denotes the productivity of individuals of type θ. Total output Y_ℓ under lockdown policy ℓ is thus given by

$$Y_\ell = \int_{\underline{\theta}}^{\overline{\theta}} A(\theta) L_\ell(\theta)\, dF(\theta).$$

Letting $L_0 \geq 0$ denote the proportion of labor that is not affected by the lockdown variable ℓ, the amount of labor of type θ that is productive given ℓ is

$$L_\ell(\theta) = L_0 + (1 - \ell(\theta)) L_1.$$

For simplicity, we again assume that L_0 and L_1 are independent of θ.

3 Homogeneous agent model

We now analyze the dynamics of an epidemic and then derive the minimal lockdown policy ℓ_K necessary to satisfy the constraint K at all times in the model with homogeneous agents.

3.1 Dynamics of an epidemic

The dynamics of an epidemic in our simple homogeneous agent SIR model are governed by the following system of non-linear differential equations:

$$\frac{dS(t)}{dt} = -\frac{\beta I(t)S(t)}{N}, \quad \frac{dI(t)}{dt} = \frac{\beta I(t)S(t)}{N} - \gamma I(t) \quad \text{and} \quad \frac{dR(t)}{dt} = \gamma I(t),$$

with initial conditions $S(0) = N_1$, $I(0) = N_2$ and $R(0) = N_3$. Harko et al. (2014) provided an analytic solution to this system of equations by parameterizing time t by a parameter u. In particular, introducing the integration constants

$$S_0 = N_1 \mathrm{e}^{\frac{\beta N_3}{\gamma N}}, \quad u_0 = \mathrm{e}^{-\frac{\beta N_3}{\gamma N}}, \quad \text{and} \quad C_1 = -\beta$$

we have

$$\begin{aligned} t(u) &= \int_{u_0}^{u} \frac{1}{\xi(C_1 - \gamma \log(\xi) + S_0 \beta \xi / N)} d\xi, \\ I(u) &= N - S_0 u + \frac{\gamma \log(u) N}{\beta}, \\ R(u) &= -\frac{\gamma \log(u) N}{\beta}. \end{aligned}$$

Notice that when $u = u_0$ we have $t = 0$ and that u decreases as t increases.[10] We can then back out $S(u)$ using $S(u) + I(u) + R(u) = N$.

The basic dynamics of an epidemic are as follows. As susceptible individuals become infected and then recover, the stock of susceptible individuals decreases over time and the stock of recovered individuals increases over time. The number of infected individuals initially increases before reaching an epidemic peak and then gradually decreasing. The number of infected individuals stops increasing once the population of susceptible individuals is sufficiently small. An example of a typical epidemic path is shown in Figure 1. Note that unless stated otherwise all figures are drawn for the parameterization $N_1 = 0.999$, $N_2 = 0.001$,

[10] In the limit as $t \to \infty$ we have $u \to -\gamma W \left(-\mathrm{e}^{\frac{C_1}{\gamma}} \beta S_0 / (\gamma N)\right) N / (\beta S_0)$, where W is the product log function.

$N_3 = 0$, and $\gamma = 1/18$; Figure 1 assumes $\beta = 3.1/18$.[11]

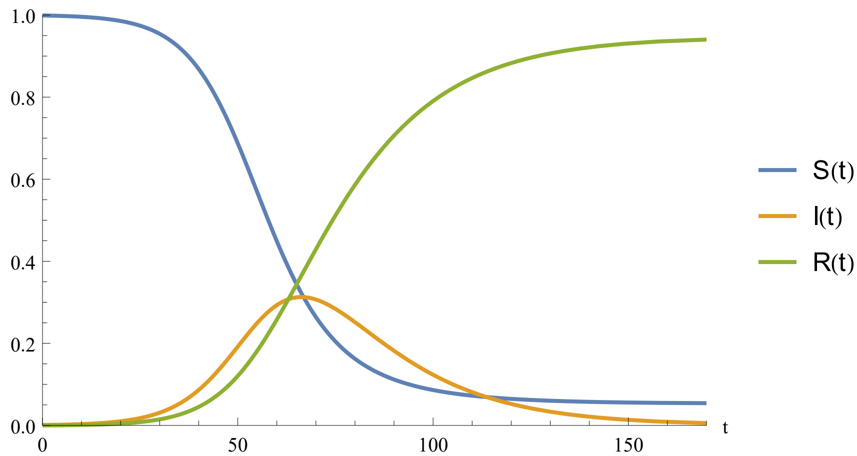

Figure 1: The evolution of a typical epidemic.

Assuming $R_0 N_1/N = \beta N_1/(\gamma N) > 1$ (so that the peak of the epidemic does not occur at $t = 0$), the maximal number of infected individuals $I^*(\beta)$ during the epidemic is characterized by

$$I^*(\beta) = N - S_0 u_{\max} + \frac{\gamma \log(u_{\max}) N}{\beta},$$

where

$$u_{\max} = \frac{\gamma N}{\beta S_0}.$$

Notice that we have

$$\frac{d\, I^*(\beta)}{d\, \beta} = -\frac{\gamma \log\left(\frac{\gamma N}{\beta N_1}\right) N}{\beta^2} > 0, \tag{2}$$

where the inequality follows from the fact that $\log(\gamma N/(\beta N_1)) < 0$ since by assumption $\gamma N/(\beta N_1) < 1$.

3.2 Capacity constraints

Notice that the parameters β_0, β_1, γ and τ, as well as the initial conditions, impose restrictions on the lower feasible bound for K. Specifically, denote by $I_\ell(t)$ the number of infectious

[11] We normalize the size of the population to 1 and assume that initially 0.1% of the population is infected. Following Wang et al. (2020), we take $\gamma = 1/18$, which reflects an average disease duration of 18 days (and so the appropriate interpretation of the time scale t is then also in days). We also set $\beta = \beta_0$ so that $R_0 = 3.1$, which is in line with estimates from Wuhan, China prior to the introduction of strict lockdown measures.

at time t given policy $\ell \in [0,1]$, by

$$I^*_\ell = I^*(\beta(\ell))$$

the maximum number of infected individuals given ℓ, and by

$$T^*_\ell = \tau I^*_\ell$$

the maximum number of people needing treatment per time given policy ℓ. From (2) we have that I^*_ℓ and T^*_ℓ are continuously decreasing in ℓ since $\beta(\ell)$ is decreasing in ℓ. From this and continuity it follows that K is feasible if and only if

$$K \in [T^*_1, T^*_0]. \tag{3}$$

If $K < T^*_1$, then the capacity constraint is so tight that it can never be satisfied at the peak of the epidemic, not even with the most severe lockdown policy. If $K > T^*_0$, then no lockdown is required to satisfy the constraint.

Conversely, for any K satisfying (3) there is a *minimal lockdown policy*, denoted ℓ_K, that satisfies the constraint that the number of individuals requiring treatment at time t never exceeds K. Formally,

$$\ell_K := \min\{\ell \in [0,1] \mid T^*_\ell \leq K\}.$$

Because T^*_ℓ is a decreasing functions of ℓ, it follows that ℓ_K is a decreasing function of K. Intuitively, as the capacity constraint K increases, the severity of the required lockdown decreases.

As for policy implications, this means that, all else equal, states or countries with larger capacities can afford less stringent lockdowns. For a given lockdown policy the transmission rate parameter β can also vary substantively between states and countries as the value of this parameter varies with factors such as population density and household composition. Since the maximum number of patients requiring treatment is given by $\tau I^*(\beta)$ and $I^*(\beta)$ is increasing in β (see (2)), it follows that, all else equal, states or countries with larger transmission rates require more stringent lockdowns. Formally, compare two regions, each with capacity K, with transmission rates parameterized by (β_0, β_1) and $(\hat{\beta}_0, \hat{\beta}_1)$ satisfying $\hat{\beta}_i \geq \beta_i$ for $i = 0, 1$, where at least one of these inequalities is strict. Denoting the respective minimal lockdown policies by ℓ_K and $\hat{\ell}_K$, we then have

$$\ell_K < \hat{\ell}_K.$$

In other words, regions with lower transmission rates can afford slacker lockdown policies as is illustrated in Panel (a) of Figure 2. This figure uses the same parameters values as Figure 1 but with $(\beta_0, \beta_1) = (0.5/18, 2.6/18)$ and $(\hat{\beta}_0, \hat{\beta}_1) = (0.6/18, 2.6/18)$.

(a) A range of β schedules (b) A range of N_2 values values

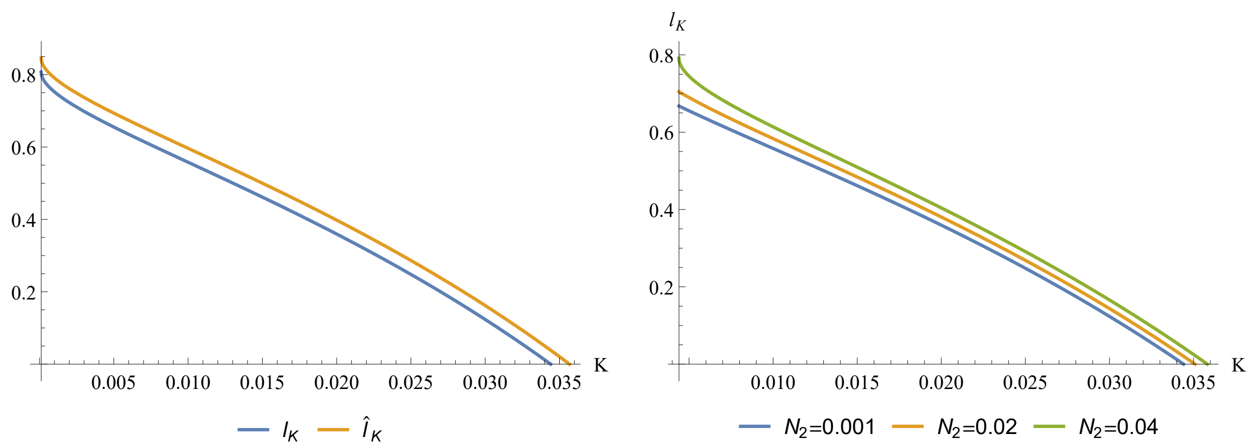

Figure 2: Panel (a) illustrates that a higher schedule of β values necessitates a more severe lockdown for a given K value. Panel (b) illustrates the relationship between ℓ_K and K for a range of N_2 values. As was shown analytically, for a given N_2 value, the severity of the lockdown ℓ_K decreases as the capacity K of the healthcare system increases. This figure also shows that a more severe lockdown is required if a higher proportion N_2 of the population is initially infected.

The relationship between lockdown and capacity We now look in slightly more detail at the relationship between K and ℓ_K. Panel (b) of Figure 2 plots this relationship, assuming $\beta(\ell) = 0.5\gamma + 2.6\gamma(1 - \ell)$ and $\tau = 0.11$.[12] As before we set $N_3 = 0$ and $\gamma = 1/18$ but we now create plots for three different values of N_2: $N_2 = 0.001$ (in which case $N_1 = 0.999$), $N_2 = 0.02$ (in which case $N_1 = 0.98$) and $N_1 = 0.04$ (in which case $N_1 = 0.96$). Panel (b) of Figure 2 shows that the lockdown policy ℓ needed to achieve a given K increases in

[12]Following Wang et al. (2020), we use $R_0 = 3.1$ with no lockdown and $R_0 = 0.5$ under the strictest possible lockdown. The first of these R_0 values is an estimate for Wuhan, China prior to any policy interventions by the Chinese government. The second of these R_0 values is an estimate for Wuhan, China under the strictest lockdown measures implemented by the government. We use $\tau = 0.11$, which is consistent with data from New York which showed that around 11% of confirmed coronavirus patients were hospitalized at the peak in hospitalizations (Feuer, 2020). Note that τ is not the rate of hospitalization (i.e. the proportion of coronavirus patients that are hospitalized at some point over the course of their illness) but rather the proportion of coronavirus patients that are hospitalized at any given point in time. An alternative approach would be to include a separate compartment in the model for hospitalizations, the main advantage being that this would produce a time-lag between the peak in the number of infected individuals and the peak in the number of hospitalized individuals (which is consistent with what we observe in data). We ensure that the binding constraint on the healthcare system is not violated due to this time-lag effect by calibrating the model using the proportion of the population that is hospitalized at the peak in hospitalizations.

the proportion of the population that is initially infected. This figure also shows how the proportion of the population that requires treatment at the height of the pandemic, for a given lockdown policy ℓ, increases in the proportion of individuals N_2 that is initially infected. Consequently, for a given cap K, a more severe lockdown is required as N_2 increases. This result highlights the high cost of a delayed policy response.[13]

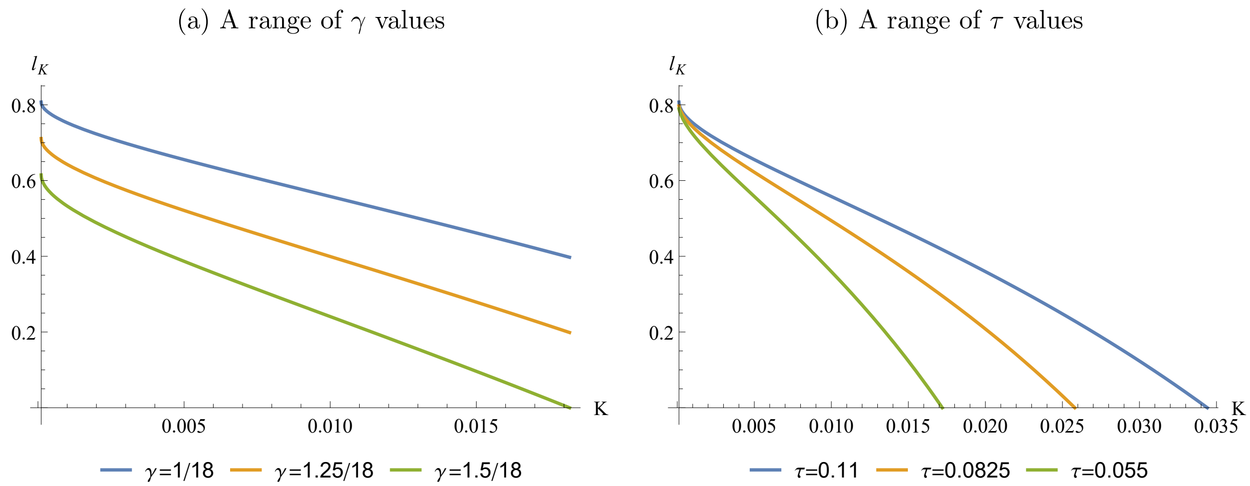

Figure 3: The severity of the required lockdown decreases as γ increases and as τ decreases.

Figure 3 provides some additional comparative statics showing how the severity of the required lockdown decreases as γ increases and as τ decreases. Panel (a) uses the same parameter values as Panel (b) of Figure 2 but with $N_2 = 0.001$ (in which case $N_2 = 0.999$ since we set $N_3 = 0$) and $\gamma = 1/18$, $1.25/18$ and $1.5/18$. Panel (b) also uses the same parameter values as Panel (b) of Figure 2 but with $N_2 = 0.001$ (in which case $N_2 = 0.999$ since we set $N_3 = 0$) and $\tau = 0.11$, 0.0875 and 0.055. One interpretation of these comparative statics is that as superior treatments become available, individuals both require less overall treatment and recover from the disease more quickly and hence a less severe lockdown is required.

Economic impact By mapping the severity of the lockdown to gross domestic product (GDP), one can trace out the relationship between the capacity constraint and economic

[13]For example, when San Francisco issued a shelter-in-place order on March 16, 2020 its number of per capita confirmed coronavirus cases was comparable to that of New York City. By the time New York City was subject to a shelter-in-place order on March 23, 2020 the number of per capita confirmed coronavirus cases was greater than that of San Francisco by roughly an order of magnitude.

output. Substituting $L(\ell)$ into the production function yields output

$$Y(\ell) = (L_0 + (1-\ell)L_1)^\alpha \,.$$

It follows that, given the minimal lockdown policy ℓ_K for the constraint K, output, denoted Y_K, is given by

$$Y_K = Y(\ell_K).$$

Because $Y(\ell)$ decreases in ℓ and ℓ_K decreases in K, it follows that

$$\frac{dY_K}{dK} > 0.$$

A plot illustrating how Y_K increases in K for a given set of parameters can be found in Figure 4, which assumes $L_0 = 1/3$, $L_1 = 2/3$ and $\alpha = 1/3$ (and as before $N_1 = 0.001, 0.02, 0.04$, $N_3 = 0$, $\beta(\ell) = 0.5\gamma + 2.6\gamma(1-\ell)$, $\gamma = 1/18$, $\tau = 0.11$).[14] This plot shows that longer delay in the initial policy response—which leads to a higher number of infected individuals in the population prior to any lockdown intervention—results in policy makers facing a more severe economic impact of the pandemic in order to satisfy the binding constraint K.

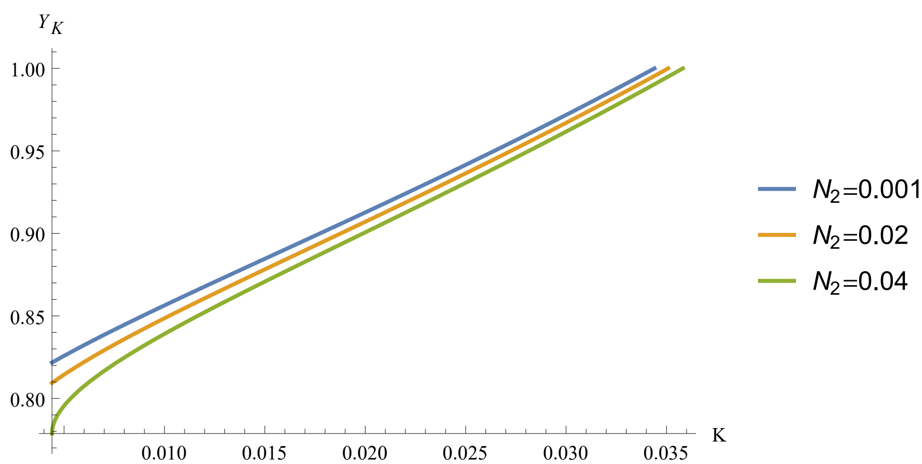

Figure 4: Output Y_K increases in K and decreases in N_2.

[14]We use a standard value of $\alpha = 1/3$ and setting $L_0 = 1/3$ corresponds to a labor force in which one third of workers are essential. While we do not pursue this here, our model also lends itself to the possibility of relating the costs and benefits of a lockdown to the statistical value of life such as DALY (disability-adjusted life year) or QALY (quality-adjusted life year) that are widely used in public health debates. See Hall et al. (2020) for a framework that considers the tradeoff between consumption and COVID-19 deaths. By expressing the value of a life in terms of years of per capita consumption this approach allows those authors to derive an upper bound on the level of consumption a utilitarian society would be willing to forgo in order to avoid COVID-19 deaths.

3.3 Sensitivity analysis and confidence intervals

The very nature of contagious diseases is that, inherently, their dynamics are non-linear. Large part of the need to to use epidemiology models arises from the need to predict future spread and, in our setting, healthcare demand as a policy that only adapts to concurrent data without accounting for states in the future will fail to satisfy the capacity constraints. Of course, as with any predictive model, predictions are subject to uncertainty and errors that can result both from model misspecification and from uncertainty about parameters within the model. We now briefly discuss how the latter can be accounted for within the homogeneous population model. (In Section 5, we then discuss modeling alternatives, none of which will invalidate the basic point that combining economics and epidemiology models is required to find a sustainable path through the pandemic.)

Up to this point we have treated the parameters of the model as given and known. As just mentioned, in practice, there may be considerable uncertainty and measurement error associated with these parameter values, implying that the predictions of the model are not deterministic. We now discuss how the distribution of the predicted peak of the epidemic can be derived from the, by assumption, known distributions of the uncertain parameters R_0, τ, N_1, N_2 and N_3.

After some tedious algebra and imposing the normalization $N = 1$, the proportion of the population requiring treatment at the peak of the epidemic is given by

$$T^*(R_0, \tau, N_1, N_3) = \tau \left(1 - N_3 - \frac{\log(R_0) + \log(N_1) + 1}{R_0}\right).$$

The density of T^* is thus

$$f_{T^*}(y) = \int_{\mathbb{R}^5} f_{R_0}(x_1) f_\tau(x_2) f_{N_1}(x_3) f_{N_3}(x_4) \delta(y - T^*(x_1, x_2, x_3, x_4))\, dx_1\, dx_2\, dx_3\, dx_4,$$

where δ denotes the Dirac delta function and f_X denotes the distribution of the random variable X. From here one can construct a confidence interval for the maximum number of individuals requiring treatment at the peak of the epidemic.

For example, suppose that R_0, N_1 and N_3 are known parameters and that $\tau \sim \mathrm{N}(0.11, 0.01)$. That is, τ is normally distributed with mean 0.11 and standard deviation 0.01. Then $T^*(R_0, \tau, N_1, N_3)$ is normally distributed with mean $\mu(R_0, N_1, N_2) = T^*(R_0, 0.11, N_1, N_3)$ and standard deviation

$$\sigma(R_0, N_1, N_2) = 0.01 \left(1 - N_3 - \frac{\log(R_0) + \log(N_1) + 1}{R_0}\right).$$

A 95% confidence interval for the value of T^* is then given by $[\mu-1.96\sigma, \mu+1.96\sigma]$. Therefore, if we use $T = \mu(R_0, N_1, N_2) + 1.96\sigma(R_0, N_1, N_2)$ then we can say that with 97.5% confidence, the constraint K will not be violated at the peak of the epidemic. An illustration is provided in Figure 5. This figure uses precisely the same parameters as those shown in Panel (b) of Figure 2 and sets $N_2 = 0.001$ (in which case $N_1 = 0.999$).

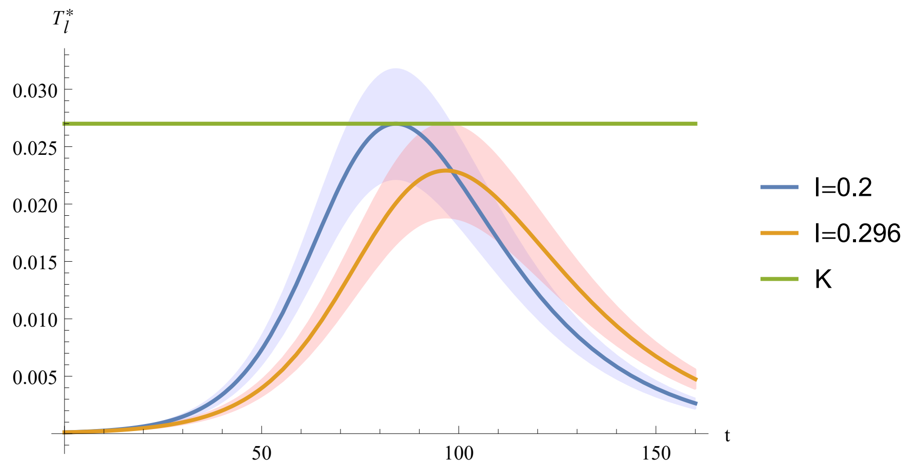

Figure 5: Assume $K = 0.0270$. If τ is deterministic and equal to 0.11, we have $\ell_K = 0.2$. In contrast, if τ is normally distributed with mean 0.11 and a standard deviation of 0.01, then $\ell = 0.296$ is required to satisfy the constraint with a probability of 0.975.

An alternative to the approach adopted here would be to perform a conditional worst-case analysis. That is, if one had estimates of moments of a particular parameter distribution (such as its mean and variance), then one could compute confidence intervals with respect to the worst-case distribution.

4 Heterogeneous agent model

We now turn our attention to the heterogeneous agent model. Without loss of generality, one can normalize $N(\theta) = 1$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$. This in turn implies that

$$N = \int_{\underline{\theta}}^{\overline{\theta}} (S(\theta,t) + I(\theta,t) + R(\theta,t))\, dF(\theta) = 1.$$

Under our assumption of homogeneous mixing among all type cohorts, the time t rate of transition between the compartment of susceptible individuals of type θ and the compartment of infected individuals of type θ is

$$\int_{\underline{\theta}}^{\overline{\theta}} \beta_\ell(y) I_\ell(y,t)\, dF(y).$$

The dynamics of an epidemic in this SIR model are then governed by the following system of non-linear differential equations

$$\begin{aligned}
\frac{dS_\ell(\theta,t)}{dt} &= -\int_{\underline{\theta}}^{\overline{\theta}} \beta_\ell(y) I_\ell(y,t)\, dF(y) S_\ell(\theta,t), \\
\frac{dI_\ell(\theta,t)}{dt} &= \int_{\underline{\theta}}^{\overline{\theta}} \beta_\ell(y) I_\ell(y,t)\, dF(y) S_\ell(\theta,t) - \gamma I_\ell(a,t), \\
\frac{dR_\ell(\theta,t)}{dt} &= \gamma I_\ell(\theta,t),
\end{aligned}$$

with initial conditions $S_\ell(\theta,0) = N_1(\theta)$, $I_\ell(\theta,0) = N_2(\theta)$ and $R_\ell(\theta,0) = N_3(\theta)$, where $N_1(\theta) + N_2(\theta) + N_3(\theta) = 1$. Letting $S_\ell(t) = \int_{\underline{\theta}}^{\overline{\theta}} S_\ell(\theta,t)\, d\theta$, $I_\ell(t) = \int_{\underline{a}}^{\overline{\theta}} I_\ell(\theta,t)\, d\theta$ and $R_\ell(t) = \int_{\underline{\theta}}^{\overline{\theta}} R_\ell(\theta,t)\, d\theta$ and integrating this system of differential equations yields

$$\begin{aligned}
\frac{dS_\ell(t)}{dt} &= -\int_{\underline{a}}^{\overline{a}} \beta_\ell(y) I_\ell(y,t)\, dF(y) S_\ell(t), \\
\frac{dI_\ell(t)}{dt} &= \int_{\underline{\theta}}^{\overline{\theta}} \beta_\ell(y) I_\ell(y,t)\, dF(y) S_\ell(t) - \gamma I_\ell(t), \\
\frac{dR_\ell(t)}{dt} &= \gamma I_\ell(t).
\end{aligned}$$

Thus, under homogeneous mixing this model reduces to a simple homogeneous agent SIR model with a transmission rate $\beta = \int_{\underline{\theta}}^{\overline{\theta}} \beta_\ell(y) I_\ell(y,t)\, dF(y)$ that depends on the type-dependent lockdown policy.

The social planner then selects the lockdown policy ℓ that maximizes output Y_ℓ subject to the constraint that total hospitalizations not exceed K at any point during the epidemic. That is, for all $t \geq 0$,

$$T_\ell(t) \leq K.$$

To ensure that we have an interesting problem, we assume that this constraint is violated under the lockdown policy $\ell(\theta) \equiv 0$ and is slack under lockdown policy $\ell(\theta) \equiv 1$.

Proposition 1. *Assume $R(\theta,0) = k_0$ and $I(\theta,0)/S(\theta,0) = k_1$, where $k_0, k_1 > 0$ are constants, and for all $\theta \in [\underline{\theta}, \overline{\theta}]$, $\frac{dA(\theta)}{d\theta} < 0$. Then the optimal policy is bang-bang, that is, there* *is an* $\theta^* \in [\underline{\theta}, \overline{\theta}]$ *such that*

$$\ell^*(\theta) = \begin{cases} 0, & \theta \leq \theta^* \\ 1, & \theta > \theta^* \end{cases}.$$

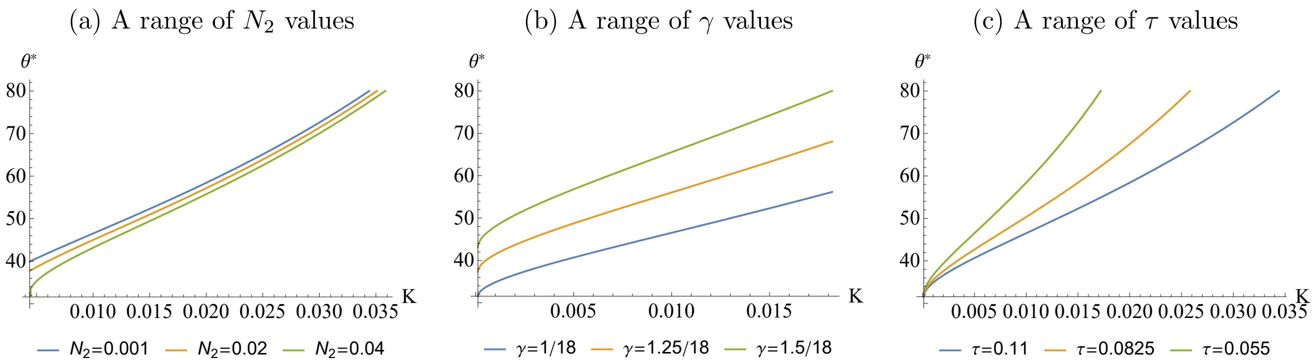

Figure 6: The age cutoff θ^* associated with the optimal bang-bang lockdown policy increases in K and γ and decreases in N_2 and τ. These figures use the same parameter values as Figures 2 and 3 where types are distributed uniformly over the interval $[20, 80]$ and τ is constant across types.

Having a bang-bang solution is not only analytically convenient but also useful for practice: even though the planner might want to consider a continuum of lockdown policies, which in practice would be difficult, in this case it is without loss of generality to only consider only minimal and maximal lockdown policies across type cohorts. Figure 6 indicates that similar comparative illustrated in Figures 2 and 3 hold for the cutoff type θ^* that characterizes the optimal lockdown policy under the heterogeneous agent model.

Interestingly, the heterogeneous agent model induces a non-trivial economic optimization problem that does *not* require taking a stance on how economic activity is traded off against the number of deaths caused by the disease. Indeed, the optimal lockdown policy in this setup maximizes economic output subject to a given capacity constraint. In the sense, the dollars-death tradeoff is not the starting point of the analysis but rather a result of the analysis.[15] Of course, many of the specific policy implications derived from our setup need not carry over to richer models but the basic feature that models with heterogeneous agents and a capacity constraint induce an economic optimization problem without specifying the value of life remains valid.

[15]While we require the capacity constraint not be exceeded at the peak of the epidemic, one could, at the cost of increasing the complexity of the model by incorporating additional dimensions of heterogeneity, augment the model with additional constraint such as requiring the total number of hospitalizations or deaths not to exceed a given threshold.

Policy-dependent mixing Another notable feature of Proposition 1 is that it does not include an assumption concerning how τ varies with θ. This is a direct consequence of the homogeneous mixing assumption. However, in practice, one might also expect a lockdown policy to impact how the type cohorts mix. In this case, the structure of the optimal policy will also depend on how hospitalization rates vary across types. We now relax the assumption of homogeneous mixing and consider type-dependent mixing. Motivated by the form of the optimal policy under homogeneous mixing and for purposes of tractability we restrict attention to lockdown policies such that $\ell(\theta) \in \{0, 1\}$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$. We further assume that only types subject to the same lockdown policy mix (and mix in a homogeneous fashion). We then have the following proposition.

Proposition 2. *Assume that for all* $\theta \in [\underline{\theta}, \overline{\theta}]$ *we have* $\frac{dA(\theta)}{d\theta} < 0$, $\frac{d\tau(\theta)}{d\theta} > 0$, $R(\theta, 0) = k_0$ *and* $I(\theta, 0)/S(\theta, 0) = k_1$, *where* $k_0, k_1 > 0$ *are constants. Then under policy-dependent mixing with* $\ell(\theta) \in \{0, 1\}$ *for all* $\theta \in [\underline{\theta}, \overline{\theta}]$ *the optimal policy is monotone. That is, there exists a* $\theta^*_{PD} \in [\underline{\theta}, \overline{\theta}]$ *such that*

$$\ell^*(\theta) = \begin{cases} 0, & \theta \leq \theta^*_{PD} \\ 1, & \theta > \theta^*_{PD} \end{cases}.$$

Moreover, $\theta^*_{PD} > \theta^*$.

Suppose, for illustrative purposes, that agent types correspond to age cohorts and that younger age cohorts are more productive than older age cohorts. Then Proposition 1 shows that under homogeneous mixing the optimal lockdown policy allows younger individuals who are more productive to return to work, while older individuals are subject to a strict lockdown in order to combat the spread of the epidemic. Proposition 2 shows that if older age cohorts are also more vulnerable and are more likely to require hospitalization, then the optimal lockdown policy under policy-dependent mixing takes a similar form. Moreover, a bang-bang lockdown policy is more effective, in the sense that a smaller proportion of the population is subjected to a lockdown (and hence economic output is higher) with policy-dependent mixing than without it.

5 Discussion

We now provide a brief discussion of natural alternatives to and extensions of our modeling approach and then conclude the section with a short discussion of dynamically adjusting policies.

5.1 Epidemiology modelling alternatives

The model we have analyzed is one with a constant population. A first natural extension would be to include births and deaths, which are sometimes also referred to as *vital dynamics*. Moreover, one could include additional compartments—such as having a compartment for exposed individuals and another for hospitalized individuals—or additional transitions by allowing, for example, for the possibility that recovered individuals become susceptible. However, as long as recovered individuals retain immunity until the introduction of a vaccine, it would not be necessary to account for the possibility of recovered individuals becoming susceptible again in models that are designed to deal with the current crisis.[16] Similarly, assuming a constant population is probably a reasonable approximation for the problem at hand, where a relatively short time period is relevant and the death rate for individuals that participate in the workforce is very low.[17] Accounting for the impact of births (including immigration), deaths (including those from coronavirus) and aging (i.e. people moving into and out of the labor force) on the dynamics of the epidemic as well as the size of the labor force would greatly complicate the model without substantively changing its broad predictions.

5.2 Agent-based modeling

Continuum SIR models, such as the ones analyzed here, provide good approximations for large populations. However, for smaller populations or more refined targets—such as ensuring that ICU beds do not run out or targeting a relatively narrow death rate band—this family of models does not necessarily provide a good approximation. For these kinds of applications, models using agent-based simulations are more appropriate tools. In simulation-based approaches, it is relatively straightforward to include extensions such as age-structuring the population and labor force because there is no need to seek analytic solutions. A recent first step in this direction is taken by Akbarpour et al. (2020), whose model does, however, not account for capacity constraints of the healthcare system. An important next step would therefore be to incorporate capacity constraints into an agent-based model.

[16] Accounting for individuals transition from recovered back to susceptible is important, for example, when studying the seasonal dynamics of diseases such as influenza.

[17] Verity et al. (2020, Table 1 (last column)) report dramatic differences in age-specific fatality-to-infected ratios. For example, for the group of individuals 60 years old and older, this ratio is 3.28% while for the cohort of individuals in their 20s, it is 0.0309%, that is, the individuals in their 20s are roughly 100 times less likely to die when infected than those past 60. Similarly, Williamson et al. (2020, p.2) report that the "overall cumulative incidence of death 90 days after study start was $< 0.01\%$ in those aged 18-39 years, rising to 0.67% and 0.44% in men and women respectively aged ≥ 80 years"; see also their Figure 2.

5.3 Dynamically adjusting policies

Another natural extension of the theoretical analysis would be to allow the lockdown policy to vary over time and solve the control problem that pins down the optimal dynamic lockdown policy. For example, following the peak of the epidemic a less stringent lockdown would suffice in order to ensure that the capacity constraint remains satisfied. This analysis would also serve to shed light on issues such as whether a shorter and more severe lockdown would allow for a higher overall level of economic output without violating the capacity of the healthcare system, relative to a less severe but more protracted lockdown.

Similarly, in practice, new and arguably more accurate parameter estimates will probably be obtained as the epidemic progresses, which will allow policy makers to adjust policies dynamically.

6 Conclusions

This time *is* different.[18] The cause of the economic downturn (a pandemic rather than the burst of a financial bubble or any other structural issue with the economy), its scope (universal, hitting all countries more or less within the same quarter) and magnitude (record increases in unemployment filings in the United States) are unprecedented. While there are good reasons to be confident that, informed by the in-depth analyses of past mistakes, the policy response to a severe economic downturn will be better and swifter than at the onset and during the Great Depression, the unparalleled nature of the current shock makes recovery a perilous and winding road. Although policy makers may be ready to act swiftly, the ongoing virulence of the disease may prevent them from so doing. Without widespread immunization, return to normalcy would be difficult if not impossible even if there were no inertia in rebooting economies that have come to a standstill. We will have to find the path to recovery by learning on the go, and learning quickly.

This paper proposes a pathway to recovery. It suggests a shift in thinking from extinguishing an epidemic in its entirety at all costs, to instead thoughtfully managing it with regards to the tradeoffs between managing the pandemic and economic output. This can be achieved by targeting an appropriately chosen constraint, such as the capacity of the healthcare system. Implementing this approach requires that we build new models that combine economic and epidemiological data and collect accurate data concerning the course of the

[18]This sentence is the title of the New York Times bestseller by Reinhart and Rogoff (2009), where it is used to explain why financial crises occur—because decision makers, in the lead-up to a financial crisis, trend to ignore important precedents. Here and now, however, it seems an accurate description of the current COVID-19 crisis.

epidemic through widespread testing.

Arguments have been put forth that policy makers should first take care of the public health aspect of the pandemic and only tackle the economic fallout once the health crisis has been dealt with. Generally speaking, it is not clear what it means to only turn to the economic aspects down the track nor whether the two dimensions can be really separated. Of course, it may be that an effective vaccine is just around the corner. In this case, the present paper and much of the related research will be of little use for the current pandemic, and for most countries and regions, catastrophic health outcomes like New York City or Lombardy experienced in the first half of 2020 will have been averted. However, it may also that developing an effective vaccine proves difficult or even elusive, in which case the health crisis and the economic crisis cannot be separated. Our approach provides a way of formalizing the notion of dealing with the health crisis first—avoid health catastrophes by satisfying the capacity constraints at all times—while minimizing the economic fallout of satisfying these constraints.

References

Akbarpour, M., C. Cook, A. Marzuoli, S. Mongey, A. Nagaraj, M. Saccarola, P. Tebaldi, S. Vasserman, and H. Yang (2020): "Socioeconomic Network Heterogeneity and Pandemic Policy Response," Working paper.

Alvarez, F. E., D. Argente, and F. Lippi (2020): "A Simple Planning Problem for Covid-19 Lockdown," NBER Working Paper # 26981.

Atkeson, A. (2020): "What Will Be the Economic Impact of Covid-19 in the US? Rough Estimates of Disease Scenarios," NBER Working Paper # 26867.

Budish, E. (2020): "$R < 1$ as an Economic Constraint: Can We "Expand the Frontier" in the Fight Against Covid-19?" Note.

Feuer, A. (2020): "Stark Death Toll, but Cautious Optimism in N.Y. Over Hospitalizations," *New York Times*.

Gilbert, M., M. Dewatripont, E. Muraille, J.-P. Platteau, and M. Goldman (2020): "Preparing for a responsible lockdown exit strategy," *Nature Medicine*, Published Online, April 14.

Hall, R. E., C. I. Jones, and P. J. Klenow (2020): "Trading Off Consumption and COVID-19 Deaths," Working paper.

Harko, T., F. S. N. Lobo, and M. K. Mak (2014): "Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates," *Applied Mathematics and Computation*, 236, 184–194.

Jehle, G. and P. Reny (2011): *Advanced Microeconomic Theory*, Addison-Wesley, 3rd ed.

Murray, J. D. (2002): *Mathematical Biology: I. An Introduction*, Berlin: Springer, 3rd ed.

Reinhart, C. and K. Rogoff (2009): *This time is different*, Princeton, New Jersey: Princeton University Press.

Verity, R., L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, ..., and N. M. Ferguson (2020): "Estimates of the severity of coronavirus disease 2019: a model-based analysis," *The Lancet*, Published Online, March 30.

Wang, H., Z. Wang, Y. Dong, R. Chang, C. Xu, X. Yu, ..., and Y. Cai (2020): "Phase-adjusted estimation of the number of coronavirus disease 2019 cases in Wuhan, China," *Cell Discovery*, 6.

Williamson, E. J., A. J. Walker, K. Bhaskaran, S. Bacon, C. Bates, . Caroline E. Morton, and B. Goldacre (2020): "Open SAFELY: factors associated with COVID-19 death in 17 million patients," *Nature*, 8 July.

Appendix

A Proofs

A.1 Proof of Proposition 1

Before proving Proposition 1 we prove a useful lemma. In particular, in the proof of Proposition 1 we will see that the planner's optimization problem is equivalent to maximizing output Y_ℓ subject to the constraint that the population average transmission rate $\int_{\underline{\theta}}^{\overline{\theta}} \beta_\ell(\theta)\, dF(\theta)$ does not exceed $\overline{\beta}(K) > 0$. We therefore start by proving that this equivalent optimization problem has a bang-bang solution. Note that in order to have an interesting problem, we assume that

$$\beta_0 < \overline{\beta}(K) < \beta_0 + \beta_1.$$

This implies that the constraint is violated under the lockdown policy $\ell(\theta) \equiv 0$ and satisfied under the lockdown policy $\ell(\theta) \equiv 1$.

Lemma A.1. *Assume that $\frac{dA(\theta)}{d\theta} < 0$ for every $\theta \in [\underline{\theta}, \overline{\theta}]$ and suppose that the planner maximizes output Y_ℓ subject to the constraint that the population average transmission rate $\int_{\underline{\theta}}^{\overline{\theta}} \beta_\ell(\theta)\, dF(\theta)$ does not exceed $\overline{\beta}(K) \in (\beta_0, \beta_0 + \beta_1)$. Then the optimal policy is bang-bang. That is, there exists $\theta^* \in [\underline{\theta}, \overline{\theta}]$ such that*

$$\ell^*(\theta) = \begin{cases} 0, & \theta \le \theta^* \\ 1, & \theta > \theta^* \end{cases}.$$

Proof. The social planner solves

$$\max_{\ell(\cdot)} \int_{\underline{\theta}}^{\overline{\theta}} A(\theta) L_\ell(\theta)\, dF(\theta) \quad \text{s.t.} \quad \int_{\underline{\theta}}^{\overline{\theta}} \beta_\ell(\theta)\, dF(\theta) \le \overline{\beta}(K).$$

Let $\mu^* \ge 0$ be the solution value of the Lagrange multiplier. By assumption the constraint is binding so $\mu^* > 0$. Substituting $L_\ell(\theta) = L_0 + (1 - \ell(\theta))L_1$, it follows that $\ell^*(\theta)$ solves

$$\max_{\ell(\cdot)} \int [A(\theta)(L_0 + (1 - \ell(\theta))L_1) - \mu^*(\beta_0 + (1 - \ell(\theta))\beta_1)]\, dF(\theta) + \mu^* \overline{\beta}(K).$$

We thus have

$$\ell^*(\theta) = \begin{cases} 0, & A(\theta)L_1 \le \mu^* \beta_1 \\ 1, & A(\theta)L_1 > \mu^* \beta_1 \end{cases}$$

and combining this last expression with the fact that $\frac{dA(\theta)}{d\theta} < 0$ shows that the optimal policy has the desired bang-bang form. □

We are now ready to prove Proposition 1.

Proof. The social planner solves

$$\max_{\ell(\cdot)} \int_{\underline{\theta}}^{\overline{\theta}} A(\theta)L_\ell(\theta)\, dF(\theta) \quad \text{s.t.} \quad \int_{\underline{\theta}}^{\overline{\theta}} \tau(\theta)I_\ell(\theta, t_{\max}(\ell))\, dF(\theta) \leq K,$$

where $t_{\max}(\ell)$ denotes the time at which the maximum number of individuals are hospitalized under the policy ℓ. Next, recall that we introduced the normalization $N(\theta) = 1$ and assume that for all $\theta \in [\underline{\theta}, \overline{\theta}]$ we have the initial conditions $R(\theta, 0) = 0$ and $I(\theta, 0)/S(\theta, 0) = k$, where $k > 0$ is a constant. Then this implies that for all $t \geq 0$ and any $\theta, \theta' \in [\underline{\theta}, \overline{\theta}]$ we have

$$S(\theta, t) = S(\theta', t), \quad I(\theta, t) = I(\theta', t) \quad \text{and} \quad R(\theta, t) = R(\theta', t).$$

This implies that the system of differential equations governing the dynamics of our SIR model can be rewritten

$$\frac{dS(t)}{dt} = -\int_{\underline{\theta}}^{\overline{\theta}} \beta_\ell(y)\, dF(y)I(t)S(t),$$

$$\frac{dI(t)}{dt} = \int_{\underline{\theta}}^{\overline{\theta}} \beta_\ell(y)\, dF(y)I(t)S(t) - \gamma I(t),$$

$$\frac{dR(t)}{dt} = \gamma I(t).$$

Thus, the evolution of the epidemic now depends on the lockdown policy ℓ only through the transmission rate $\beta = \int_{\underline{\theta}}^{\overline{\theta}} \beta_\ell(y)\, dF(y)$. Furthermore, the constraint can be rewritten

$$\tau I_\ell(t_{\max}(\ell)) \leq K,$$

where $\tau = \int_{\underline{\theta}}^{\overline{\theta}} \tau(\theta)\, dF(\theta)$. Thus, it follows from our previous analysis that the constraint is satisfied provided $\int_{\underline{\theta}}^{\overline{\theta}} \beta_\ell(y)\, dF(y) \leq \overline{\beta}(K)$, where the upper bound $\overline{\beta}$ on the transmission rate depends on K. Thus, the social planner's problem can be rewritten

$$\max_{\ell(\cdot)} \int_{\underline{\theta}}^{\overline{\theta}} A(\theta)L_\ell(\theta)\, dF(\theta)$$

$$\text{s.t.} \quad \int_{\underline{\theta}}^{\overline{\theta}} \beta_\ell(\theta)\, dF(\theta) \leq \overline{\beta}(K)$$

and it follows from Lemma 1 that the optimal policy has the bang-bang form specified in the proposition statement. □

A.2 Proof of Proposition 2

Proof. The social planner solves

$$\max_{\ell(\cdot)} \int_{\underline{\theta}}^{\overline{\theta}} A(\theta)L_\ell(\theta)\, dF(\theta)$$

$$\text{s.t.} \quad \int_{\underline{\theta}}^{\overline{\theta}} \tau(\theta)I_\ell(\theta, t_{\max}(\ell))\, dF(\theta) \leq K,$$

where $t_{\max}(\ell)$ denotes the time at which the maximum number of individuals are hospitalized under the policy $\ell : [\underline{\theta}, \overline{\theta}] \to \{0,1\}$. Let $\mu^* \geq 0$ be the solution value of the Lagrange multiplier. By assumption the constraint is binding, so we have $\mu^* > 0$ and hence ℓ^* must solve

$$\max_{\ell(\cdot)} \int_{\underline{\theta}}^{\overline{\theta}} \left(A(\theta)L_\ell(\theta) - \mu^*\tau(\theta)I_\ell(\theta, t_{\max}(\ell))\right) dF(\theta) + \mu^* K.$$

Since we have both $\frac{dA(\theta)}{d\theta} < 0$ and $\frac{d\tau(\theta)}{d\theta} > 0$, it follows that ℓ increases monotonically in θ.

We thus end up with two independent SIR models. For the type cohorts with $\theta \geq \theta^*_{PD}$ the dynamics of the epidemic are governed by

$$\frac{dS_1(t)}{dt} = -\beta_0 I_1(t)S_1(t), \quad \frac{dI_1(t)}{dt} = \beta_0 I_1(t)S_1(t) - \gamma I_1(t) \quad \text{and} \quad \frac{dR_1(t)}{dt} = \gamma I_1(t)$$

and for the type cohorts with $\theta < \theta^*_{PD}$ the dynamics of the epidemic are governed by

$$\frac{dS_0(t)}{dt} = -\beta_0 I_0(t)S_0(t), \quad \frac{dI_0(t)}{dt} = (\beta_0 + \beta_1)I_0(t)S_0(t) - \gamma I_0(t) \quad \text{and} \quad \frac{dR_0(t)}{dt} = \gamma I_0(t).$$

The path of hospitalized individuals over time is given by

$$T_{\theta^*_{PD}}(t) = \int_{\underline{\theta}}^{\theta^*_{PD}} \tau(\theta)\, dF(\theta) I_0(t) + \int_{\theta^*_{PD}}^{\overline{\theta}} \tau(\theta)\, dF(\theta) I_1(t)$$

and θ^*_{PD} is pinned down by $T_{\theta^*_{PD}}(t) = K$. Under a given bang-bang lockdown policy, by assumption the rate of transmission is lower among higher types under policy-dependent mixing relative to homogeneous mixing. Since higher type cohorts are hospitalized at a higher rate upon becoming infected it immediately follows that $\theta^*_{PD} > \theta^*$. $\square$