Chapter 9
COVID-19 and Optimal Lockdown Strategies: The Effect of New and More Virulent Strains

Jonathan P. Caulkins, Dieter Grass, Gustav Feichtinger, Richard F. Hartl, Peter M. Kort, Alexia Prskawetz, Andrea Seidl, and Stefan Wrzaczek

Abstract Most nations have responded to the COVID-19 pandemic by locking down parts of their economies starting in early 2020 to reduce the infectious spread. The optimal timing of the beginning and end of the lockdown, together with its intensity, is determined by the tradeoff between economic losses and improved health outcomes. These choices can be modelled within the framework of an optimal control model that recognises the nonlinear dynamics of epidemic spread and the increased risks when infection rates surge beyond the healthcare system's capacity. Past work has shown that within such a framework very different strategies may be optimal

J. P. Caulkins (✉)
Heinz College, Carnegie Mellon University, 4800 Forbes Avenue, Pittsburgh, PA 15213, USA
e-mail: caulkins@andrew.cmu.edu

D. Grass · S. Wrzaczek
International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, 2361 Laxenburg, Austria
e-mail: dieter.grass@tuwien.ac.at

S. Wrzaczek
e-mail: wrzaczek@iiasa.ac.at

G. Feichtinger · A. Prskawetz
Institute of Statistics and Mathematical Methods in Economics, TU Wien, Wiedner Hauptstr. 8-10, 1040 Vienna, Austria
e-mail: gustav@eos.tuwien.ac.at

A. Prskawetz
e-mail: afp@econ.tuwien.ac.at

R. F. Hartl · A. Seidl
Department of Business Decisions and Analytics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria
e-mail: richard.hartl@univie.ac.at

A. Seidl
e-mail: andrea.seidl@univie.ac.at

P. M. Kort
Tilburg School of Economics and Management, Tilburg University, Warandelaan 2, 5037 AB Tilburg, The Netherlands
e-mail: Kort@uvt.nl

163

ranging from short to long and even multiple lockdowns, and small changes in the valuation on preventing a premature death may lead to quite different strategies becoming optimal. There even exist parameter constellations for which two or more very different strategies can be optimal. Here we revisit those crucial questions with revised parameters reflecting the greater infectivity of variants such as the "UK variant" of the SARS-CoV-2 virus and describe how the new variant may affect levels of mortality and other outcomes.

9.1 Introduction

To reduce social interactions and thereby also contagious transmission of the SARS-CoV-2 virus, most countries have implemented one or several lockdowns of non-essential parts of the economy. While lockdowns have succeeded to varying degrees in reducing new infections, the effects on the economy (Fernández-Villaverde and Jones 2020) can be substantial. The lockdowns can themselves harm health, either directly (e.g., when non-essential healthcare is deferred) or indirectly (unemployment and poverty can reduce life expectancy). The question therefore arises as to what is the ideal duration and intensity of lockdowns. If lockdowns are relaxed too soon, the epidemic may bounce back. If these measures are too severe or prolonged, needless economic hardship may result.

9.1.1 The Challenge of New Virus Variants

Such questions have become even more pressing with the discovery of new, mutated strains of the SARS-CoV-2 virus, notably one detected first in the UK and thus referred to as "UK variant". This strain appears to be much more virulent, in the sense of spreading more rapidly. In particular, the U.S. Centers for Disease Control (CDC) had previously advised that epidemiological models use as a base case assumption that the basic reproduction number (denoted by R_0) of the SARS-CoV-2 virus was 2.5, but the new variant is thought to be about 60% more contagious, suggesting a new R_0 of 4.0.

The new variant is not more lethal, so far as is understood at at the time of this writing; i.e., its infection fatality rate is not higher. So the consequences of COVID-19 spreading through the majority of the population are roughly the same, apart from greater mortality when cases are bunched up in time, swamping the healthcare system, but the severity of lockdown necessary to prevent such spread is greater and so more costly. In particular, lockdowns and other interventions that reduced social interaction by 60% would have been sufficient to stall the spread of the original virus, since $2.5 \times (1 - 60\%) = 1.0$, but with the new virus, those same interventions would have each infection leading to $4.0 \times (1 - 60\%) = 1.6$ new infections.

Since the course of the typical infection plays out over roughly two weeks, that would leave the number of new infections growing at a compound rate of about $1.6^2 - 1 = 150\%$ per month. If it will take another six to nine months to achieve herd immunity through mass vaccinations, that spread would be fast enough to infect essentially everyone in a country that currently has an average rate of infections. Hence, policies that were adequate or even optimal in the past may no longer be so today.

This paper explores how this greater virulence may or may not alter conclusions about what constitutes the ideal timing and duration of a lockdown. It also adjusts the time horizon until an effective vaccine has been widely deployed to $T = 1.5$ years, better reflecting the actual trajectory of vaccine development that has been observed.

9.1.2   Review of Past Findings

We begin here with a brief review of findings obtained from these lockdown models using the older, lower basic reproduction number of $R_0 = 2.5$. In Caulkins et al. (2020) we analysed an epidemiological model of the pandemic overlain with a simple optimal control model that considers the optimal starting and ending times of a lockdown that withdraws part of the population from the labour force. The objective function balances economic costs (lost output) and health costs (COVID-19 related deaths) while considering the limited capacity of intensive care units within the health care system. The number of deaths is modelled as being proportional to the number of infections plus an extra penalty for infections that happen when hospitals are overwhelmed.

Even that rather simple model produces some complex behaviour. For instance, sometimes starting a lockdown later might make it better to have a shorter, not a longer lockdown. Most interestingly, we found the formal mathematical version of the notion of "tipping points" that were popularised by Malcolm Gladwell's famous book of that name. In particular, for certain parameter values two very different strategies (e.g., long versus short lockdown) can be optimal when starting at the same initial condition, and slight deviations away from those starting points may make either type of solution optimal. In optimal control models, such tipping points with two or more alternative optimal solution trajectories have been called Skiba, Sethi-Skiba, DNS, and DNSS points (Grass et al. 2008).

Characterising such points is important for two reasons. First, it may help explain why different countries have pursued such different lockdown strategies despite having similar interests in balancing economic and health considerations. Secondly, these tipping points highlight the need to gather better information about the key parameters that cause these different lockdown strategies to be optimal.

Caulkins et al. (2021) extended Caulkins et al. (2020) by allowing for multiple lockdowns and also considering lockdowns of varying intensity, rather than treating a lockdown as an all or nothing binary choice. In addition, the economic modelling is richer in two respects. First, employment is represented by a state variable, and the

policy maker's choices, or "control variable", adjusts that level of economic activity. Underemployment is costly of course (foregone economic activity), but so too is changes in that level; rapidly alternating between mild and severe lockdowns is more costly than maintaining an intermediate lockdown because change is disruptive to business. Furthermore, the adjustment costs are asymmetric to recognise that shutting down businesses may be easier than re-starting them.

Second, since the public's cooperation can wane when lockdowns are too intensive and long, Caulkins et al. (2021) include "lockdown fatigue" as an additional state variable which may undermine the efficiency of a lockdown.

Within this framework, the optimal lockdown strategies are quite diverse, ranging from long and forceful lockdowns to (a couple of) short and rather soft lockdowns. Again, the specific parameter values determine the optimal strategy. Similar to Caulkins et al. (2020), there are parameter constellations for which two very different strategies are both optimal. In addition, we also found triple Skiba points at which even three different strategies are optimal. The fact that such complex strategies result from rather stylised models hints at the complexity of designing lockdown strategies in practice. While our models cannot specify/recommend any single optimal lockdown strategy, our framework provides insight as to which are the most important parameters that drive the decision about the start, duration and intensity as well as the multiplicity of lockdowns.

Before investigating how the UK variant might affect these models and selected results we give a brief summary of related papers in the literature.

9.1.3   Review of Other Related Literature

Several papers discuss the balancing of health and economic interests (see Layard et al. 2020; Bloom et al. 2020; Scherbina 2020; Brodeur et al. 2020 for a careful evaluation). However only a minority of these papers have investigated the optimal timing, length and extent of the lockdown itself. These exceptions include e.g. Gonzalez-Eiras and Niepelt (2020) who start from a simple SIR model and investigate the optimal lockdown intensity and duration taking into account the tradeoff between health and economic consequences of the lockdown.

In Alvarez et al. (2020) the fraction of people going into lockdown is assumed to be the control variable. The model is derived with and without testing as a control variable where test availability implies that those who are recovered can be identified and are not subject to lockdown. It is shown that absence of testing will increase the economic costs of the lockdown and shorten its duration since the dynamics of the epidemiology imply that an increasing share of recovered will decrease the efficiency of the lockdown.

An optimal control model on reducing the transmission rate is presented in Abel and Panageas (2020) that also allows for positive vital rates (births and non-COVID deaths). They show that it is not optimal to eradicate the disease but to limit interactions until a cure or vaccination becomes available. In Acemoglu et al. (2020) a het-

erogeneous SIR model is applied that distinguishes between "young", "middle-aged" and "old". It is shown that a stricter lockdown on the old is particularly important. Compared to optimal uniform policies such targeted policies imply a considerable reduction in mortality but may also reduce economic damage since the young and middle-aged groups can be released from lockdown earlier. A similar argument is presented in Gershon et al. (2020) who show that if ICU beds are in short supply, partial quarantine of the most vulnerable group may be enough compared to a shut down of the whole economy. Aspri et al. (2021) consider a SEIRD model, where the population is divided into susceptibles, exposed but asymptomatic, infected, recovered and deceased. Similar to other papers they model the optimal tradeoff between reduction in fatalities and the loss in output. However, different to the literature so far, they assume that containment policies are piece-wise linear functions representing a more realistic policy modelling. Based on the specific value assumed for a statistical life they obtain multiple lockdowns as well as Skiba points. While the previous models apply numerical solution methods, Rachel (2020) presents an analytical model of COVID-19 lockdowns. By differentiating between the individual versus social optimal mitigation strategy it is shown that at the individual level too much social distancing will result in equilibrium relative to the social optimum. This result is explained by the fact that higher social distancing today will reduce infections and flatten out the curve, but raise infection rates later on. In contrast, a social planner considers the cumulative infection risk and not just the infection risk today. In Huberts and Thijssen (2020) a stochastic version of the SIR model is introduced. Based on a continuous-time Markov chain model the optimal timing of intervention and the option to end the intervention are studied. Federico and Ferrari (2021) present a model on the optimal lockdown policy where not only the transmission rate in the standard SIR model is stochastic, but also the time horizon is assumed to be stochastic. Within their framework they show that the optimal policy is first to let the epidemic evolve, followed by a pronounced containment policy and in the last phase to reduce the strength of the containment again. Similar to our model the limited capacity of health system is considered in Piguillem and Shi (2020). Testing is an important strategy to accrue welfare gains in their paper.

The body of this paper now proceeds in two parts. Section 9.2 explores a model in which the policy maker only gets to decide the start and end times of a lockdown, but the intensity of that lockdown is fixed. Section 9.3 then examines the more general situation when the lockdown intensity can be varied continuously over time, and there can even be more than one lockdown.

9.2 The Optimal Start and Length of a Lockdown

This section updates the model presented in Caulkins et al. (2020) to address the higher infectivity of the UK variant of the SARS-CoV-2 virus.

9.2.1 The Model

The epidemiological model we apply is based on an open-population SLIR model Kermack and McKendrick (1927) with a birth rate ν and extra mortality for individuals who are infected (μ_I) above and beyond that for those who are susceptible or recovered (μ):

$$\dot{S}(t) = \nu N(t) - \beta \frac{S(t)\mathcal{I}(I(t), L(t))}{N(t)} - \mu S(t) \tag{9.1a}$$

$$\dot{L}(t) = \beta \frac{S(t)\mathcal{I}(I(t), L(t))}{N(t)} - (\mu + \varphi)L(t) \tag{9.1b}$$

$$\dot{I}(t) = \omega\varphi L(t) - (\alpha + \mu + \mu_I)I(t) \tag{9.1c}$$

$$\dot{R}(t) = (1 - \omega)\varphi L(t) + \alpha I(t) - \mu R(t) \tag{9.1d}$$

$$\beta := R_{\text{eff}}(t, \tau_1, \tau_2)\alpha \tag{9.1e}$$

$$\mathcal{I}(I, L) := I + fL \tag{9.1f}$$

$$N(t) := S(t) + L(t) + I(t) + R(t). \tag{9.1g}$$

The state variables $S(t)$, $L(t)$, $I(t)$ and $R(t)$ denote the number of individuals respectively who are susceptible to infection, have a latent (asymptomatic and presymptomatic) infection, are infected and symptomatic, and are recovered at time t. The term "recovered" is standard in the literature even though it is a bit of a misnomer because it includes not only those who have recovered from COVID-19 symptoms (i.e., passed through the I state), but also those who previously had asymptomatic infections (passed through the L state only). (CDC guidance is that about 40% of those who become infected remain asymptomatic.)

The parameter β is key to the epidemic dynamics. The term it modifies in Eq. (9.1a) counts potential interactions between those who are susceptible to becoming infected (those in the S state) and those are infected (those in the L and I states). The symbol $\mathcal{I}$ denotes the weighted sum of people in the I and L states, weighting by the (lower) relative likelihood of spreading the virus when in the L state. (As of this writing, the CDC recommends assuming this weighting parameter $f = 0.75$.) β is essentially a proportionality constant that converts social interactions into infections. Outside of the lockdown it has one (higher) value; during the lockdown its value is lower, e.g., because either the infected or susceptible person wears a mask, maintains social distance, interacts only virtually if one or both work from home, or the interaction simply does not occur because it has been banned by the lockdown.

Although lockdowns directly affect β, the effective reproduction number $R_{\text{eff}}(t, \tau_1, \tau_2)$ is more readily interpretable, so we describe the lockdown phases in terms of effects on $R_{\text{eff}}(t, \tau_1, \tau_2)$ and adjust β accordingly. (For a formal derivation of the relationship between the basic reproduction number R_0 and β, see Appendix 2 in Caulkins et al. (2020)). The start and end times of the lockdown are denoted by τ_1 and τ_2. They are chosen by the decision maker and—as we only allow one lockdown in this setup—define three periods: before, during and after that one lockdown.

Following CDC guidance, Caulkins et al. (2020) assumed that the basic reproduction number before the lockdown equals $R_0^1 = 2.5$. The lockdown was assumed to reduce that to $R_0^2 = 0.8$ and not bounce back fully, as some behavioural adjustments (such as not shaking hands) could be expected to continue even after people return to work. The extent to which those behavioural changes persisted depends on the duration of the lockdown. In particular, Caulkins et al. (2020) assumed that after the lockdown there exists a gap between the realised and potential value of $R_0^3 = 2.0$, with the potential value being reached only with increasing length of the lockdown.

Here, we assume that the reproduction numbers before and after the lockdown are 1.6 times larger (so $R_0^1 = 4.0$ and $R_0^3 = 3.2$) but continue to assume that during the lockdown $R_0^2 = 0.8$. I.e., we implicitly assume that the lockdown intensity is increased sufficiently to push the reproduction number appreciably below 1.0 despite the new variant's greater infectivity.

We model COVID-19 deaths by focusing on those who require hospitalisation and critical care. Some calculations (described in Caulkins et al. 2020) suggest that about $p = 2.31\%$ of people who develop symptoms will need critical care, and 45% of them will die prematurely as a result of COVID-19 even if they receive that care. The parameter p converts the 2.31% into a daily rate by multiplying by α, the reciprocal of the average duration of symptoms, which we take to be nine days. Likewise, the death rate per person-day spent in the I state by people who need and also receive critical care is $\mu_I = p\xi_1\alpha$.

In addition, there is an extra risk of death for people who need critical care but do not receive it because hospitals are overwhelmed. That term is proportional to $\max(\{0, pI - H_{\max}\})$ where $H_{\max}$ is the number of critical care hospital beds available.[1] In the U.S., there are about 0.176 critical care beds per 1,000 people. Overall deaths are therefore represented by:

$$\xi_1 pI + \xi_2 \max(\{0, pI - H_{\max}\}, \zeta),$$

where ξ_1 is the death rate from COVID-19 of infected people who need and receive critical care, and ξ_2 is the additional, incremental death rate when such individuals do not receive that care. One aim of the decision maker is to minimise these deaths.

It is of course very difficult to determine what value society should place on averting a premature death generally, or in the case of COVID-19 in particular. We represent that quantity by the parameter M, the cost per COVID-19 death, and consider a very wide range of values for that parameter.

The literature has traditionally used values for M ranging at least from 20 times GDP per capita (Alvarez et al. 2020) up to 150 times GDP per capita (Kniesner et al. 2012). Hammitt (2020) argues that lower values may be appropriate for COVID-19 deaths, so we consider a range from 10 to 150 times GDP per capita.

Economic activity is modelled as being proportional to the number of employed people raised to a power, as in a classic Cobb-Douglas model, with that exponent set

---

[1] In our numerical simulations we have replaced the max function that is not differentiable with a smooth function (see Caulkins et al. 2020, Fig. 1).

to $\sigma = 2/3$ (Acemoglu 2009). Since the time horizon is short, capital is presumed to be fixed and subsumed into the objective function coefficient K for economic activity. Susceptible, latent, and recovered individuals are eligible to work (symptomatic individuals are assumed to be either too sick to work, or are in quarantine). During a lockdown, only a proportion $\gamma(t)$ of those eligible to work are employed. We therefore assume that $\gamma(t) = 1.0$ before the lockdown, $\gamma(t) = 0.25$ during the lockdown, and after lockdown it only partially recovers. The longer the lockdown, the more jobs that are lost semi-permanently because firms go out of business. That recovery is modelled as decaying exponentially in the length of the lockdown with a time constant of 0.001 per day, so that if a lockdown ended after six months, 17% of jobs suspended during the lockdown would not reappear, at least until a vaccine became available.

These economic and mortality costs are summed up from time $t = 0$, when the virus arrives, until time $T = 1.5$ years, when a vaccine has been developed and widely deployed.

The objective function also includes a salvage value that reflects the reduction in economic activity at time T relative to what it was at time 0 (see Caulkins et al. 2020 for further discussion of the salvage value). The summary of the full model and the base case parameter values are given in Appendix 1 and Table 9.4 Appendix 3.

9.2.2 Results

COVID-19 spread very fast in early 2020, so lockdown initiation was often a rushed decision made so quickly that there was no time to build models or optimise them. Hence, we start, in Fig. 9.1, by considering the simpler problem of when to end a lockdown that has already started, answering that question for a wide range of start times. In particular, the left hand panel of Fig. 9.1 shows with the solid blue line how that optimal ending time (measured by the vertical axis, τ_2) varies as a function of the time the lockdown was started (given as the horizontal axis, τ_1). The gap between the blue line and the black line (45-degree line) indicates the duration of the lockdown.

For this model and these parameter values, if the lockdown starts promptly (so on the left side of that panel) the lockdown should be maintained almost until the time when the vaccine has been successfully deployed. That is assumed to happen in 1.5 years; since time is measured in days, that corresponds to 547.5 on the vertical axis. That the blue line starts out at a level of about 500 days indicates maintaining the lockdown until only a month or two before the vaccine has been successfully deployed would minimise total costs, including both health and economic costs. (Ending the lockdown before full deployment does not require an implausible degree of forecasting ability; predicting how long it will take to invent an effective vaccine is hard, but deployment takes approximately six months, so recognising when it is within a couple months of wrapping up is not that hard.)

So the first conclusion is, if a nation starts to lock down early, it should keep that lockdown in place more or less for the duration of the epidemic.
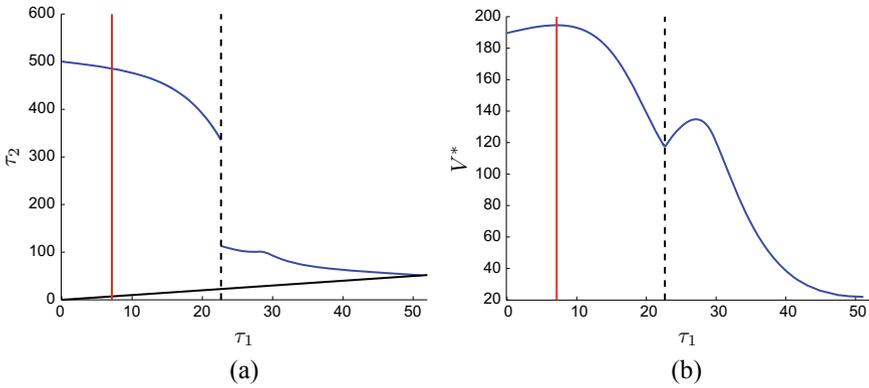
Fig. 9.1 Panel a shows the solutions for a fixed initial lockdown time τ_1 and optimally chosen time τ_2. In panel b the objective value is shown for the optimally chosen time τ_2. For $\tau_1 = 22.6$, which is indicated by the black dashed line, there exists a Skiba solution, i.e. there are two different solution paths which deliver the same objective value. The red vertical line denotes the optimally chosen τ_1. The parameter values are those of Table 9.4 in Appendix 3, with $M = 60,000$ and $R_0^2 = 0.8$

Now suppose the lockdown's initiation was delayed a bit, meaning we slide a little to the right along the horizontal axis of Fig. 9.1. Intuitively one might have expected that getting a late start would imply one should maintain the lockdown longer to compensate, but the opposite is true in this model. The fact that the blue line slopes downward implies that the later one starts the lockdown, the sooner it should end.

The second surprising result is that the blue line does not decline smoothly; it contains a discontinuous jump when the lockdown starts at $\tau_1 = 22.6$ days. As one delays the start of the lockdown from τ_1 from 0 up to 22.5 days, the ideal ending time τ_2 decreases smoothly from about 500 days (roughly a year and four months) down to a little less than a year. Then suddenly, when the lockdown starts just a little later, at $\tau_1 = 22.7$ days, it becomes optimal to end the lockdown fairly soon, at only $\tau_1 = 120$ days, or after about three months.

What has happened at that point is that the epidemic has had a chance to spread so widely in those first 22.7 days that it is just too hard to rein the epidemic in for it to be worthwhile. If a prolonged lockdown were going to spare most of the population from getting infected it would be worth the cost; but if the lockdown hasn't started until $\tau_1 = 22.7$ days, it is just too late for it to be wise to pursue that strategy. One should still lockdown, but only relatively briefly. That can "flatten the curve" a bit and avoid totally swamping the limited capacity of the healthcare system.

In simple words, if the lockdown starts too late, then one should abandon the "long lockdown" strategy that protects most people from infection, and instead employ a much more limited "curve flattening" strategy.

The discontinuity in the blue line shows that if the lockdown starts at just exactly $\tau_1 = 22.6$ days, then either the "long lockdown" or the "curve flattening" can be followed with equal results.

This equivalence is illustrated more directly in the right-hand panel of Fig. 9.1, which shows the so-called value function (V) versus the lockdown initiation time τ_1.

The value function indicates the performance achieved when the optimal strategy is followed. There is a kink in the value function right at $\tau_1 = 22.6$ days. To the left of that kink it is optimal to follow the "long lockdown" strategy, but the plunging value function shows that the "long lockdown" strategy performs less and less well as τ_1 increases. Likewise, to the right of $\tau_1 = 22.6$ days it is optimal to follow the "curve flattening" strategy, but as the lockdown start time decreases, approaching 22.6 from the right, the "curve flattening" strategy does less and less well. And right at τ_1 the two strategies' value functions cross.

The third surprising result pertains to where the blue line in the right- hand panel of Fig. 9.1 peaks. It is not ideal to start the lockdown immediately at $\tau_1 = 0$. Instead, the value function peaks at about $\tau_1 = 8$ days (a point in time indicated in the left hand panel by a vertical red line). The reason is that every day of lockdown is expensive, because people are out of work, but at the very beginning, when there are very, very few infected people, there are also very, very few new infections to be prevented. When the virus is very scarce, targeted approaches, such as testing and contact tracing, may be preferred to shutting down the entire economy.

Figure 9.1 illustrated two different strategies: a short "curve flattening" lockdown and a long lockdown that starts after a short delay. For other parameter values, two other strategies can be optimal: never locking down at all or a long lockdown that begins immediately.

That raises the question of under what conditions it is optimal to pursue each strategy. Figure 9.2 answers that question with respect to two key parameters: (1) The economic value placed on preventing a COVID-19 death, M, and (2) the epidemic's reproduction number during the lockdown, denoted by the parameter R_0^2. That figure, called a bifurcation diagram, shows for each combination of those two key parameters which strategy is optimal.

The base case values for those parameters were $R_0^2 = 0.8$, meaning the lockdown could still drive the reproduction number below the critical threshold of 1.0, and $M = 60,000$, meaning that the cost of a premature death is set at about 150 times GDP per capita. That point falls within the region labelled IIb, but for other values of R_0^2 and/or M different strategies may be optimal.

Not surprisingly, as M increases—meaning moving from left to right in Fig. 9.2—the optimal strategy changes to make greater and greater use of lockdowns. When M is very small, it may be optimal not to lockdown at all. When M is sufficiently large, then a long lockdown is best. For intermediate values of M, the "curve flattening" strategy may be best.

The verticality of the Skiba curve separating the regions where no lockdown vs. a short lockdown are optimal indicates that increasing the epidemic's reproductive number during the lockdown R_0^2 has little effect on the relative merits of not locking down versus using a short lockdown. That makes sense precisely because in neither of those strategies was the lockdown prolonged in any event. However, the Skiba curve separating regions where a short lockdown (Region I) and a long lockdown (Regions IIa and IIb) is preferred slopes up and to the right indicating that the larger R_0^2 is, the larger M must be in order to justify a long lockdown. That also makes sense. If the new variant's higher virulence sufficiently undermines the effectiveness of locking
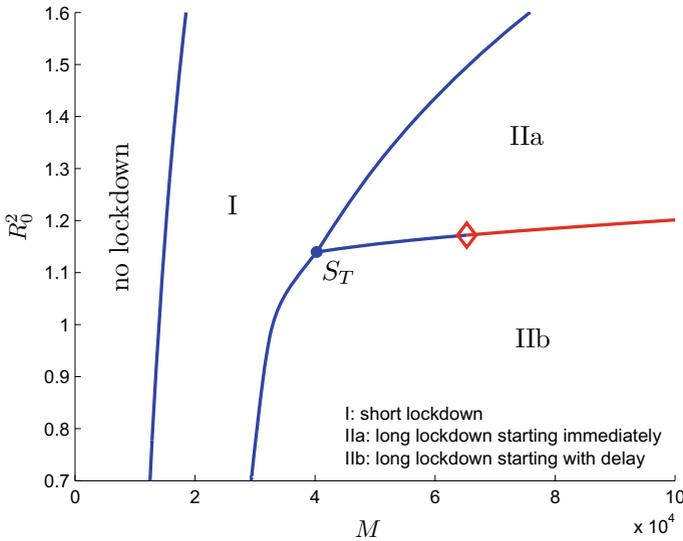
Fig. 9.2 Bifurcation diagram in the $R_0^2 - M$ space. The blue lines denote Skiba curves, separating (discontinuously) regions with different optimality regimes. At the red curve the regimes change continuously. The point S_T corresponds to a triple Skiba point, where three optimal solutions exist. At the red diamond the discontinuous Skiba solution changes into a continuous transition curve

down, then the value per life saved has to be greater to justify the imposition of a long lockdown.

One of the interesting features of this model is that two Skiba curves intersect, namely the curve separating Region I from Regions IIa or IIb and the curve separating Regions IIa and IIb. That intersection, which is denoted by the point S_T, is a triple Skiba point. If the parameters have exactly those values, then any of three distinct strategies can be optimal. It is akin to Snow Dome Mountain in Canada's Jasper National Park, where a drop of water could equally well flow west through the Columbia River system to the Pacific Ocean, east to Hudson's Bay and the Atlantic Ocean, or north via the Athabasca and McKenzie Rivers into the Arctic Ocean. Except that instead of being indifferent between flowing to different oceans, at this point a social planner is indifferent between starting a long lockdown immediately, starting a long lockdown after a short delay, and employing only a short lockdown.

9.3   The Optimal Lockdown Intensity

The previous section updated results from a model based on Caulkins et al. (2020) that sought to determine the optimal start and length of a lockdown when the intensity of that lockdown was given exogenously. We next present an extension of the model

as given in Caulkins et al. (2021) that allows the intensity of the lockdown to vary continuously over time.

As in Caulkins et al. (2020), we define $\gamma(t)$ to be the share of potential workers who are employed at time t. However, now we model $\gamma(t)$ as a state variable that can be altered continuously via a control $u(t)$:

$$\dot{\gamma}(t) = u(t), \quad \gamma(0) = 1,$$

We set $\gamma(0) = 1.0$ because the planning horizon begins when COVID-19 first arrives, and so before there is any lockdown. We include a state constraint $\gamma(t) \leq 1$ for $0 \leq t \leq T$, since employment cannot exceed 100%. This formulation allows for multiple lockdowns, takes into account that employment takes time to adjust, and it recognises that changing employment levels induces adjustment costs, which we allow to be asymmetric, with it being harder to restart the economy than it is to shut it down.

Since public approval of lockdowns may wane the longer a lockdown lasts, we introduce a further state variable that models this "lockdown fatigue" $z(t)$:

$$\dot{z}(t) = \kappa_1 (1 - \gamma(t)) - \kappa_2 z(t),$$

where κ_1 governs the rate of accumulation of fatigue and κ_2 measures its rate of decay. Note that if the worst imaginable lockdown ($\gamma(t) = 0$) lasted forever then $z(t)$ would grow to its maximum possible value of $z_{\max} = \kappa_1/\kappa_2$.

We use an epidemiological model based on an open-population SIR[2] model with a birth rate ν and extra mortality for individuals who are infected (μ_I) above and beyond that for those who are susceptible or recovered (μ). In addition, we allow a backflow of recovered individuals back into the susceptible state at a rate ϕ. How long immunity will last with SARS-CoV-2 virus is not known at the time of this writing, but immunity to other corona viruses often lasts 3–5 years, so we set ϕ to 0.001 per day in our base case, which corresponds to a mean duration of immunity of $1000/365 = 2.74$ years.

---

[2] Since the qualitative dynamics in Caulkins et al. (2020) did not change if we excluded the latent state, we opted for a more parsimonious model in our extensions Caulkins et al. (2021) and the parameters have been adapted accordingly.

The state dynamics in our extended model can then be written as

$$\dot{S}(t) = \nu N(t) - \beta(\gamma(t), z(t)) \frac{S(t)I(t)}{N(t)} - \mu S(t) + \phi R(t)$$

$$\dot{I}(t) = \beta(\gamma(t), z(t))) \frac{S(t)I(t)}{N(t)} - (\alpha + \mu + \mu_I)I(t)$$

$$\dot{R}(t) = \alpha I(t) - \mu R(t) - \phi R(t)$$

$$\dot{\gamma}(t) = u(t), \quad \gamma(0) = 1$$

$$\dot{z}(t) = \kappa_1(1 - \gamma(t)) - \kappa_2 z(t), \quad z(0) = 0$$

$$\gamma(t) \leq 1, \quad 0 \leq t \leq T$$

where $N(t) = S(t) + I(t) + R(t)$ is the total population. As before, the factor $\beta(\gamma, z)$ captures the number of interactions and the likelihood that an interaction produces an infection. It is assumed to depend on both the intensity of the lockdown γ and the level of lockdown fatigue z in the following way:

$$\beta(\gamma, z) := \beta_1 + \beta_2 \left(\gamma^\theta + f \frac{\kappa_2}{\kappa_1} z(1 - \gamma^\theta) \right)$$

This expression can be interpreted as follows. In the absence of lockdown fatigue, we might model $\beta(\gamma, 0)$ as some minimum level of infection risk β_1 that is produced just by essential activities plus an increment β_2 that is proportional to γ raised to an exponent $\theta > 1$. Having θ greater than 1 is consistent with locking down first the parts of the economy that generate the most infections per unit of economic activity (perhaps concerts and live sporting events) and shutting down last industries with high economic output per unit of social interaction (perhaps highly automate manufacturing and mining).

The term $(\kappa_2 / \kappa_1)z$ is the lockdown fatigue expressed as a percentage of its maximum possible value. So if $f = 1$ and z reached its maximum value, then all of the potential benefits of locking down would be negated. Lockdown fatigue will not actually reach that maximum because the planning horizon is relatively short. Also, we choose a relatively small value of $f = 0.05$, so this lockdown fatigue has only a modest effect. Nonetheless, including this term at least acknowledges this human dimension of the public's response to lockdowns.

The objective function includes health costs (due to deaths from COVID- 19), economic loss (due to locking down), and the adjustment costs of changing the employment level γ. We assume these adjustment costs to be quadratic in the control u and allow for them to be asymmetric with different constants for shutting down businesses c_l and reopening them c_r, with an extra penalty for reopening after an extended shut down so that

$$V_u(u, \gamma) := \begin{cases} c_l u^2 & u \leq 0 \\ c_r(z + 1)u^2 & u > 0. \end{cases}$$

The resulting optimal control model and the base case parameter values are summarised in Appedix 2 and Table 9.4 Appendix 3.

9.3.1 Results

9.3.1.1 The Effect of Increased Infectivity

Our main interest is in how a mutated strain that is more contagious alters what strategies are optimal. That is perhaps best captured in Fig. 9.3, which has two panels. The one on the left corresponds to the old reproduction number of $R_0 = 2.5$; the one on the right corresponds to the new, higher number of $R_0 = 4$. Both are similar to the right panel of Fig. 9.1 in that they show how the value function depends on the parameter M describing the cost per premature death.

This value function can be thought of as the "score" that a social planner "earns" when he or she follows the optimal strategy. Naturally in both panels the value function slopes down. The greater the penalty the social planner "pays" for each premature death, the lower the score. On the left side of each panel the value function slopes down steeply because there isn't much locking down so there are a lot of deaths; thus, a given increment in the cost per death gets "paid" many times. On the right side of each panel, the optimal strategy involves an extended lockdown, so there are fewer deaths and the same increment in the cost per death reduces the social planner's score by less.

There are, though, two noteworthy differences between the value functions across the two panels. First, the kink in the curve, indicating the point at which an extended
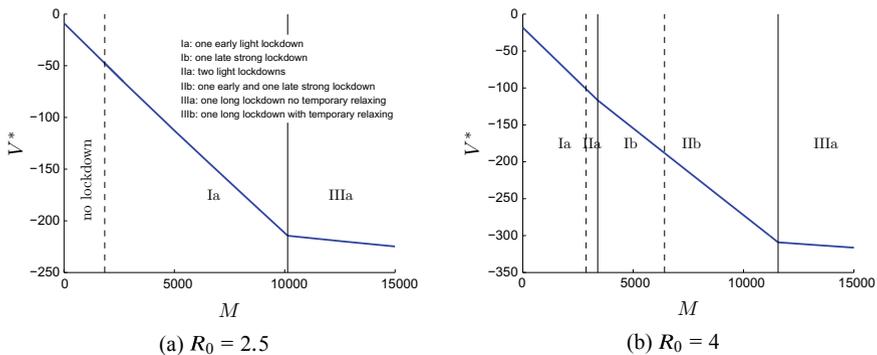

Fig. 9.3 Dependence of the value function on the social cost of a death M for the base case parameters in Table 9.4 Appendix 3 and lockdown fatigue $f = 0$. There are four main regimes (no lockdown, I, II, III) which differ by the duration, intensity, and number of lockdowns of the optimal solutions. For the value of M highlighted by solid vertical black lines two different solution paths are optimal. The dashed vertical lines denote continuous transitions from one regime to the other

lockdown becomes preferred, occurs at a larger value of M in the right-hand panel. That is because when the reproduction number is larger, it takes a more determined lockdown to pull off the extended lockdown strategy, making it more costly and less appealing unless the penalty per premature death is larger. The difference is not enormous though, with valuations equivalent to about ($M = 11, 560$) 32 times GDP per capita in the right panel and ($M = 10, 140$) 28 times GDP per capita in the left panel.

The second difference is that—at least with all other parameters at their base case values—increasing R_0 increased the number of different types of strategies that can be optimal. With $R_0 = 4$ there are five distinguishable types of lockdown strategies that can be optimal, not just two.

Here is how to interpret the labels of the four regions Ia, Ib, IIa, and IIb. The Roman numeral I or II refers to whether there are one or two lockdowns. The 'b' versus 'a' roughly indicates whether there is a substantial lockdown later in the planning horizon to prevent a rebound epidemic. (A rebound may be possible after an appreciable number of previously infected individuals have lost their immunity and returned to the susceptible state S via the backflow.)

Figure 9.4 shows example control trajectories for all five regions. The vertical axis is γ, the proportion of workers who are allowed to work, so any dip below 1.0 indicates a lockdown. If the social planner places a very low value on preventing COVID-19 deaths (e.g., $M = 1500$ in panel a), then there is only a small, short early lockdown which does little except to take a bit of the edge off the initial spike in infections. Such a small effort does not prevent many people from getting infected, but it shifts a few infections to later, when hospitals are less overwhelmed. When M is a little larger (specifically $M = 3200$ in panel c), then there is also a similarly small lockdown later, to take a bit of the edge off of the rebound epidemic. But in neither of those cases is there much locking down or much reduction in infections.

When M is still larger ($M = 5000$ in panel b) the later lockdown gets considerably larger—large enough to essentially prevent the rebound epidemic. Curiously, at this point the initial lockdown disappears, but it wasn't very big to begin with, so this qualitative change is not actually a very big difference substantively. When M increases further ($M = 11, 000$) the initial lockdown reappears, albeit as a very small blip.

Then rather abruptly when M crosses the Skiba curve separating type I and II strategies from type III strategies it becomes optimal to use a very large and sustained lockdown to reduce infections and deaths dramatically. Panel e shows the particular optimal lockdown trajectory when $M = 13, 000$, which is equivalent to valuing a premature death at 35 times GDP per capita. That sustained lockdown averts most of the infections and deaths, but at the considerable cost of almost 50% unemployment for about a year and a half.

Thus, when the lockdown intensity is allowed to vary continuously, many nuances emerge, but the overall character still boils down to an almost binary choice. If M is high enough, then use a sustained and forceful lockdown to largely preempt the epidemic despite massive levels of economic dislocation. Otherwise, lockdowns are too blunt and expensive to employ as the primary response to the epidemic. Thus, the model prescribes an almost all-or-nothing approach to economic lockdowns.
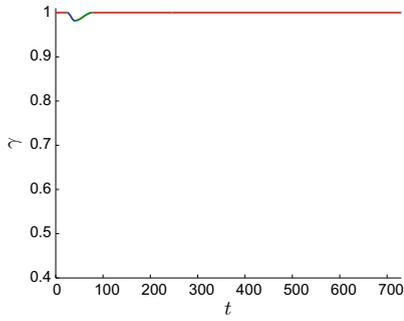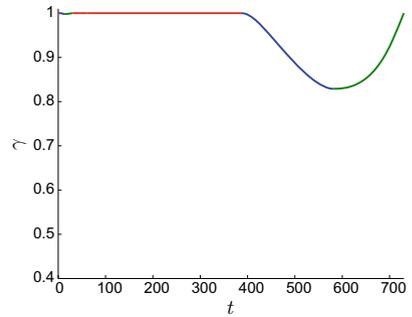
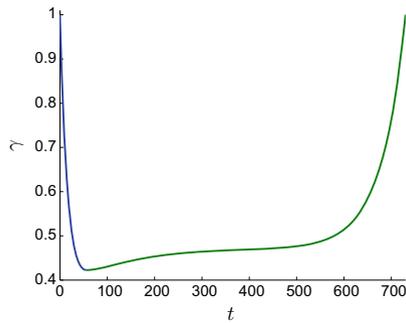(a) Regime Ia ($M = 1500$)
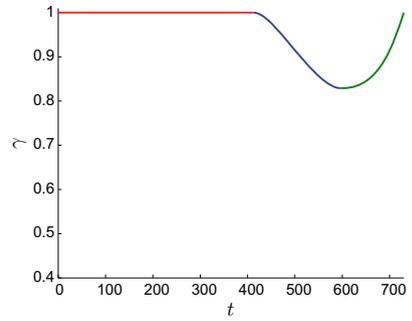
(b) Regime Ib ($M = 5000$)
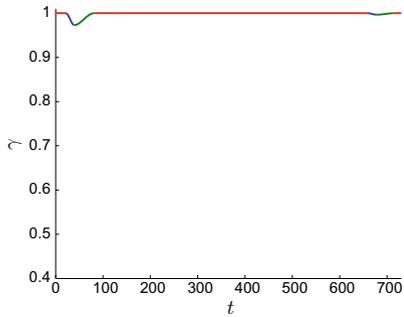
(c) Regime IIa ($M = 3200$)

(d) Regime IIb ($M = 11000$)

(e) Regime IIIa ($M = 13000$)

Fig. 9.4 Showing the time evolution of the optimal lockdown for the different regimes in Fig. 9.3b

For certain combinations of parameter values (e.g., Fig. 9.4 panels b and d corresponding to $M = 5,000$ and $M = 11,000$) it can be optimal to act fairly decisively against the rebound epidemic even if all one does in response to the first epidemic is a bit of curve flattening. It may seem odd to lock down more aggressively in response to the second, smaller epidemic, but the reason is eminently practical. When the reproduction number is high enough, it is very hard to prevent the epidemic from exploding if everyone is susceptible. But there is already an appreciable degree of herd immunity when the second, rebound epidemic threatens, so a less severe lockdown can be sufficient to preempt it.

9.3.1.2 Interpreting the Types of Lockdown Strategies that Can be Optimal

Table 9.1 summarises the nature and performance of each of the strategies in the right-hand panel of Fig. 9.3. Its columns merit some discussion. The lockdowns' start and end times are self-explanatory except to note that with strategies IIa and IIb, there are two separate lockdowns, so there are two separate start and end times. The intensity of the lockdown measures the amount of unemployment that the lockdown creates on a scale where 365 corresponds to no one in the population working for an entire year.

Table 9.1 shows that even when lockdown intensity and duration are allowed to vary continuously, there are basically only three sizes that emerge as optimal: very small (less than 1.04), modest (around 30–35, or the equivalent of the economy giving up one month of economic output), and large (around 360, or the equivalent of the economy giving up a full year of economic output).

The levels of deaths also fall into basically three levels. High (around 2.9% of the population) goes with small lockdowns. Medium-high deaths (around 2.4%) goes with modest lockdowns. Small deaths (around 0.2%) goes with large lockdowns. It would be nice to have a small number of deaths despite only imposing a small lockdown, but that just isn't possible.

In sum, there are basically three strategies: (1) Do very little locking down and suffer deaths both from the initial epidemic and also the rebound epidemic as people lose immunity, (2) Only do a bit of curve flattening during the first epidemic but use a modest sized lockdown later on to prevent the rebound epidemic and so have a medium-high number of deaths, or (3) Lockdown forcefully more or less throughout the entire planning horizon in order to avert most of the deaths altogether.

Figure 9.2 provides the corresponding information when $R_0 = 2.5$. It shows that when the virus is less contagious the large lockdown does not need to be quite as large (size of 257 or about 8.5 months of lost output, not a full year) in order to hold the number of deaths down to low levels. Perhaps surprisingly, the minimalist strategies (Ia) are less minimalist when $R_0 = 2.5$; when $R_0 = 4.0$ the epidemic is just so powerful that it is not even worth doing as much curve flattening as it is when $R_0 = 2.5$.

Table 9.1 $R_0 = 4$: Data characterising the optimal solutions for the different regimes of Fig. 9.4. The size of the lockdown is defined as $\int_{\tau_s}^{\tau_e}(1 - \gamma(t))dt$, where τ_s is the starting and τ_e the exit time of the lockdown

	Start time	End time	Duration	Size of lockdown	Deaths (%)	M
Lockdown: Ia	24.8	77.4	52.6	0.5	2.9	1500
Lockdown: Ib	412.8	730.0	317.2	32.8	2.4	5000
Lockdown: IIa						
1	21.4	81.3	59.9	0.8	–	–
2	659.5	713.8	54.2	0.1	–	–
Total	–	–	114.1	0.9	2.9	3200
Lockdown: IIb						
1	3.2	30.9	27.7	0.0	–	–
2	388.1	730.0	341.9	35.8	–	–
Total	–	–	369.6	35.8	2.3	11,000
Lockdown: IIIa	0.0	730.0	730.0	360.8	0.2	13,000

Table 9.2 $R_0 = 2.5$. The optimal solutions for $R_0 = 2.5$ evaluated at the same M values as for $R_0 = 4$

	Start time	End time	Duration	Size of lockdown	Deaths (%)	M
Lockdown: Ia	–	–	–	0	2.1	1500
Lockdown: Ib	35.2	145.8	110.6	4.8	2.0	5000
Lockdown: IIa	44.3	122.2	77.9	1.3	2.1	3200
Lockdown: IIb	0.0	730.0	730.0	257.6	0.2	11,000
Lockdown: IIIa	0.0	730.0	730.0	258.8	0.2	13,000

9.3.1.3   The Effects of Lockdown Fatigue

One feature of the current model is its recognition of lockdown fatigue. Recall that fatigue means that the infection-preventing benefits of an economic lockdown may be eroded over time by the public becoming less compliant, e.g., because the economic suffering produces pushback. The results above used parameter values that meant the power of that fatigue was fairly modest. In this subsection we explore how greater tendencies to fatigue can influence what strategy is optimal.

The tool again is a bifurcation diagram with the horizontal axis denoting M, the value the social planner places on preventing a premature death. (See Fig. 9.5.) Now, though, the vertical axis measures the strength of the fatigue effect, running from 0 (no effect) up to 1.0. The units of this fatigue effect are difficult to interpret, but roughly speaking, over the time horizons contemplated here, if $f = 1.0$ then when employing the sustained lockdown strategies, the lockdowns lose about half of their effectiveness by the time they are relaxed.

Figure 9.5 shows the results. When that fatigue parameter is small (lower parts of Fig. 9.5), the march across the various strategies with increasing M is the same as that depicted in Fig. 9.3. With large values (top of Fig. 9.5), there are two differences. First, Region Ib disappears but Region IIb remains, meaning if it is ever optimal to use a moderately strong lockdown to forestall a rebound epidemic, then one also does at least something in response to the first epidemic. Second, Region IIIa gives way to Region IIIb in which some degree of lockdown is maintained for an extended time, but it is relaxed somewhat between the first and rebound epidemics in order to let levels of fatigue dissipate somewhat.

The still more important lesson though pertains to the curve separating regions where some major lockdown is optimal (whether that is of type IIIa or IIIb) and regions where only small or moderate sized lockdowns are optimal (Regions Ia, Ib, IIa, or IIb). That boundary slopes upward and to the right, meaning that the greater the tendency of the public to fatigue, the higher the cost per premature death (M) has to be in order for a very strong and sustained lockdown to be optimal. That makes sense. If fatigue will undermine part of the effectiveness of a large lockdown, then the valuation of the lockdown's benefits has to be greater in order to justify its considerable costs.

This suggests that those advocating for very long lockdowns might want to think about whether there are ways of making that lockdown more palatable in order to minimise fatigue. For example, some Canadian provinces tempered their policies limiting social interaction to people within a household bubble so that people living alone were permitted to meet with up to two other people, to avoid the mental health harms of total isolation.
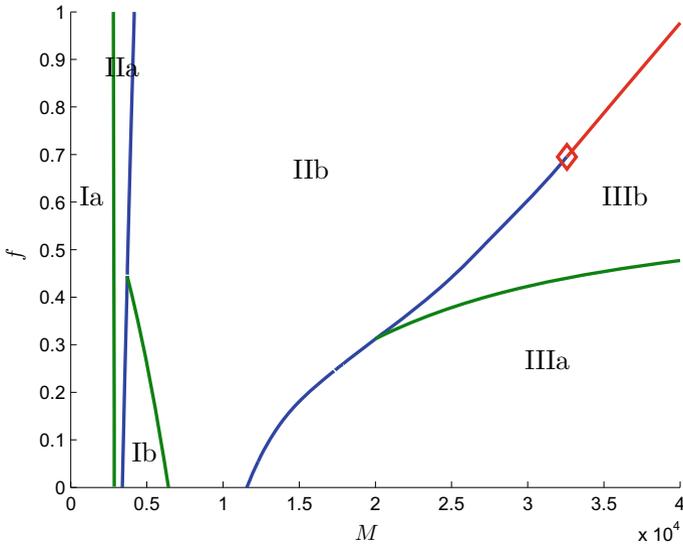
Fig. 9.5 This figure shows the different regions in the $M - f$ space. The green lines denote continuous transitions from region region I to II ($R_0 = 4$). The blue curve is a Skiba curve, where the transition from region II to III is discontinuous and at the Skiba curve two optimal solutions exist. The Skiba curve switch to a continuous transition curve (red) at the red diamond. In Regime IIIb the lockdown is relaxed in between and then tightened again, whereas in Regime IIIa the lockdown is steadily increased and then steadily relaxed

9.3.1.4   Illustrating Skiba Trajectories

One key finding here is that for certain sets of parameter values, two—or sometimes even three—very different strategies can produce exactly the same net value for the social planner. We close by illustrating this phenomenon in greater detail.

Returning to Fig. 9.3b, with the higher level of infectivity believed to pertain for the UK variant of the virus, as the valuation placed on preventing a premature death (M) increases, one crosses two Skiba thresholds, one at $M = 3395$ separating Regions IIa and Ib and another at $M = 11,560$ separating Regions IIb and IIIa. These thresholds are denoted in Fig. 9.3b by solid vertical lines. They can also be seen in Fig. 9.5 by moving left to right at the bottom level ($f = 0$).

Figure 9.6 shows the two alternate strategies, in terms of γ, the proportion of employees who are allowed to work. The left panel shows the two equally good strategies when $M = 3395$; the right-hand panel shows the two strategies that are equally good when $M = 11,560$. We have already discussed their nature. On the left side one is choosing between two very small lockdowns and one moderately large lockdown later. On the right side one is choosing between a pair of lockdowns (very small early and moderately large later) and one very deep and sustained lockdown.

Table 9.3 Data on costs for Skiba point at $M = 11,560$

	Uncontrolled	Flattening	long and sustained lockdown
Health costs	335.67	271.42	25.33
Economy	18.27	36.77	270.13
Adjustment costs	0	0.91	13.64
Total costs	353.94	309.10	309.10
Deaths (%)	2.9	2.3	0.2

The observation to stress for present purposes is just how different the trajectories are in each pairing. When one crosses a Skiba threshold, what is optimal can change quite radically. Likewise, when one is standing exactly at that Skiba threshold, one has two equally good options, but those options are radically different.

That means that when two people advocate very different lockdown strategies in response to COVID-19, one cannot presume that they have very different understandings of the science or very different value systems. They might actually share very similar or indeed even identical worldviews, but still favour radically different policies.

Table 9.3 illustrates how this can be so. Its first column summarises the outcomes (costs) when there is no control. Health costs are enormous because more or less everyone gets infected and 2.9% of the population dies; the numbers are on a scale such that 365 is one year's GDP, so the health cost of 335.7 is almost as bad as losing an entire year's economic output. There are also some economic losses from losing the productivity of those who die prematurely, producing a total cost of 353.9.

The second column shows that modest deployment of lockdowns only reduces health costs by 20%, to 271.4, whereas a severe and sustained lockdown reduces them by 92.5%, to 25.3. However, the severe and sustained lockdown multiplies costs of lost labor fifteenfold, to 270.1, and creates an additional cost equivalent to 13.6 days of output from forcing businesses to adjust to changing lockdown policies. Summing across all three types of costs produces the same total of 309.1 for both types of lockdown strategies.

Thus the two lockdown strategies produce the same aggregate performance (309.1), but with very different compositions. The moderate lockdown strategy creates smaller economic costs but only reduces health costs by 20%. The severe and sustained lockdown eliminates most of the healthcare costs but creates very large economic dislocation.

What is quite sobering is that either optimal policy only reduces total social cost by 13%, from 353.9 to 309.14. The COVID-19 pandemic is truly horrible; at least within this model, even responding to it optimally alleviates only a modest share of the suffering. Lockdowns can convert health harms to economic harms, but they cannot do much to reduce the total amount of harm.
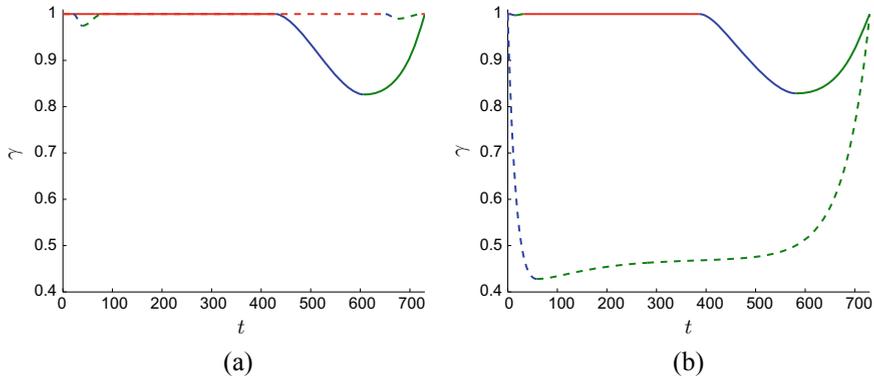
Fig. 9.6 Optimal time paths for the Skiba solutions for $M = 3395, 11560$ in Fig. 9.3b

9.4 Discussion

This paper investigated implications of the SARS-CoV-2 virus being more contagious than has previously been understood, e.g., because of a mutation or variant strain. In particular, it investigates implications for economic lockdown strategies within an optimal control model that balances health and economic considerations. A number of results were confirmed that had been obtained earlier with parameters reflecting the earlier understanding of the epidemic's reproduction number. In particular, we continue to find that:

- Very different lockdown policies can be optimal—ranging from very little to long and sustained lockdowns—depending on the value of parameters that are difficult to pin down, notably including the valuation placed on averting a premature death.
- For certain parameter constellations, the nature of the optimal policy can change radically even with quite small changes in these parameters.
- There are even situations in which two very different policies can both yield exactly the same aggregate performance, with one strategy's better performance at reducing deaths being exactly offset by its worse performance in other respects.
- As we have discussed previously in Caulkins et al. (2020), these results suggest a degree of humility is in order when advocating for one policy over another. Another person who favours a very different policy might actually share a very similar scientific understanding of the disease dynamics and even hold similar values, and yet still reasonably reach quite different conclusions.

There are, though, differences here. One is that a greater variety of strategies emerged as candidates. Some concerned how to address a potential rebound epidemic among people who were previously infected but then flowed back from the recovered to the susceptible state as their immunity wore off. For some parameter values, if the virus is sufficiently contagious and the time until a vaccine arrives long enough, it may be prohibitively difficult to substantially avoid the initial wave of infection, but

nonetheless be desirable to use a moderately aggressive lockdown to avert a rebound epidemic, for two reasons. First, it is easier to deal with the rebound epidemic because there will still be some degree of herd immunity at that time, in contrast to the situation when the virus first arrives. Second, the time until a vaccine's arrival is obviously shorter when addressing a rebound as opposed to the initial epidemic, so any lockdowns do not need to be sustained as long.

Indeed, whereas in the past the multiplicity of strategies basically fell into two camps, either a fairly modest lockdown that served only to flatten the curve and a much more intensive and sustained lockdown that largely protected the population from infection, now there is a third category of strategies. It might be thought of as flattening the first wave and eliminating the second.

We also investigated more thoroughly than before the potential effects of lockdown fatigue. The primary results are perhaps as expected. The greater the tendency for fatigue to undermine the effectiveness of a lockdown, the higher one must value the benefits created by a lockdown in order for a large and sustained lockdown to be optimal. That suggests that those wishing to impose long and deep lockdowns might want to think about ways of reducing resistance to those measures.

In sum, when lockdown intensity is allowed to vary continuously, many nuances emerge, but the overall character still boils down to an almost binary choice, one made perhaps even more stark if the virus becomes more contagious. If the value placed on preventing a premature death is high enough, then a social planner should use a sustained and forceful lockdown to preempt the epidemic despite incurring massive economic dislocation. Otherwise, lockdowns are too blunt and expensive to employ as the primary response to the epidemic. Thus, the model prescribes an almost all-or-nothing approach to economic lockdowns.

That finding does not mean that modulated approaches to what might be termed social lockdowns do not have a role. It may be entirely possible for a government to ramp up or down when and where it requires masks and social distancing outside the workplace, or to do the same with travel restrictions and quarantines. Our model is looking only at economic lockdowns.

Here is one way to think about this conclusion. We credited policy makers with a degree of common sense that could have made intermediate levels of economic lockdowns appealing. In particular, we assumed that the benefits in terms of reduced infection were a concave function of the amount of the economy that is shut down. In plain language, we presumed policy makers would shut down first the economic activities that had the greatest ratio of infection risk to economic value (e.g., in-person concerts and other crowd gatherings) and shut down last those that produce a lot of economic value per unit of infection risk (e.g., mining and highly automated manufacturing). If the virus' behaviour were linear, that might be expected to favour lockdowns of intermediate intensity. However, the contagious spread of a virus is highly nonlinear, involving very powerful positive feedback loops that produce exponential growth. To speak informally, if the virus gets its nose into the tent and locking down is the only policy response, then the virus will rip through the population if the lockdown is anything other than very strong.

The good news is that policy makers do have other tools besides lockdowns. For example, rapid, intense testing and contact tracing might be able to hold down infections when there are relatively few people getting infected. But if the virus spreads beyond the ability of such targeted measures, and the only remaining tool is broad-based economic lockdowns, the analysis here suggests being decisive; waffling efforts may produce the worst of both worlds, with substantial economic losses and still high rates of infection.

We close with a caveat. Despite its apparent complexity, this model explored here is of course vastly simplified compared to the real world, and there is much that remains unknown and uncertain about optimal economic response to pandemic threats. We hope we have usefully provoked thinking and advanced understanding, but hope even more fervently that society will invest heavily in much more such analysis, so that we can all be better prepared the next time the world confronts a novel pandemic.

Appendix 1

The decision variables are τ_1 and τ_2, the times when the lockdown begins and ends, and the full model can be written as:

$$V(X_0, \tau_1, \tau_2) := \int_0^T \big(V_l(W(t), \tau_1, \tau_2) - V_h(I(t))\big)\mathrm{dt}$$
$$- (T + \Gamma)KW(0)^\sigma \gamma(0, \tau_1, \tau_2)^\sigma$$
$$+ \Gamma KW(T)^\sigma \gamma(T, \tau_1, \tau_2)^\sigma$$

$$V^*(X_0) := \max_{\tau_1, \tau_2} V(X_0, \tau_1, \tau_2), \quad X := (S, L, I, R), \quad W := S + L + R.$$

$$\text{s.t.} \quad \dot{X}(t) = \begin{cases} \text{SLIR}_1(X(t), \tau_1, \tau_2) & 0 \le t < \tau_1 \\ \text{SLIR}_2(X(t), \tau_1, \tau_2), & \tau_1 \le t \le \tau_2 \\ \text{SLIR}_3(X(t), \tau_1, \tau_2) & \tau_2 < t \le T \end{cases}$$

$$X(0) = X_0 \ge 0$$

$$\gamma(t, \tau_1, \tau_2) := \begin{cases} \gamma_1 & 0 \le t < \tau_1 \\ \gamma_2 & \tau_1 \le t \le \tau_2 \\ \gamma_3(\tau_1, \tau_2) := \gamma_2 + (\gamma_1 - \gamma_2)e^{\kappa_2(\tau_1 - \tau_2)} & \tau_2 < t \le T \end{cases}$$

$$R_{\text{eff}}(t, \tau_1, \tau_2) := \begin{cases} R_0^1 & 0 \le t < \tau_1 \\ R_0^2 & \tau_1 \le t \le \tau_2 \\ R_0^3(\tau_1, \tau_2) := \bar{R}_0^3 + (R_0^1 - \bar{R}_0^3)e^{\kappa_1(\tau_1 - \tau_2)} & \tau_2 < t \le T \end{cases}$$

with $R_0^2 \le \bar{R}_0^3 \le R_0^1$.

We specify the health care term and the economic (labor) term in the objective as

$$V_h(I) := M\left(\xi_1 pI + \xi_2 \max(\{0, pI - H_{\max}\}, \zeta)\right)$$
$$V_l(W, \tau_1, \tau_2) := K\gamma(t, \tau_1, \tau_2)^\sigma W(t)^\sigma.$$

The derivation of the necessary optimality conditions can be found in the Appendix 1 in Caulkins et al. (2020). The Matlab toolbox `OCMat` is used for the numerical calculations (see http://orcos.tuwien.ac.at/research/ocmat_software).

Appendix 2

$$V(X_0, u(\cdot)) := \int_0^T \left(V_l(W(t), \gamma(t)) - V_h(I(t)) - V_u(u(t), z(t))\right)\mathrm{dt}$$
$$- (T + \Gamma)KW(0)^\sigma \gamma(0)^\sigma$$
$$+ \Gamma KW(T)^\sigma \gamma(T)^\sigma$$

$$V^*(X_0) := \max_{u(\cdot)} V(X_0, u(\cdot))$$

$$X(t) := (S(t), I(t), R(t), \gamma(t), z(t)), \quad W(t) := S(t) + R(t)$$

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

s.t.
$$\dot{S}(t) = \nu N(t) - \beta(\gamma(t), z(t))\frac{S(t)I(t)}{N(t)} - \mu S(t) + \phi R(t)$$

$$\dot{I}(t) = \beta(\gamma(t), z(t))\frac{S(t)I(t)}{N(t)} - (\alpha + \mu + \mu_I)I(t)$$

$$\dot{R}(t) = \alpha I(t) - \mu R(t) - \phi R(t)$$

$$\dot{\gamma}(t) = u(t), \quad \gamma(0) = 1$$

$$\dot{z}(t) = \kappa_1(1 - \gamma(t)) - \kappa_2 z(t), \quad z(0) = 0$$

$$\gamma(t) \le 1, \quad 0 \le t \le T$$

$$\beta(\gamma, z) := \beta_1 + \beta_2\left(\gamma^\theta + f\frac{\kappa_2}{\kappa_1}z(1 - \gamma^\theta)\right)$$

$$V_l(W, \gamma) := K\gamma^\sigma W^\sigma$$

$$V_h(I) := M\left(\xi_1 pI + \xi_2 \max_s(\{0, pI - H_{\max}\}, \zeta)\right)$$

$$V_u(u, z) := \begin{cases} c_l u^2 & u \le 0 \\ c_r(z + 1)u^2 & u > 0 \end{cases}$$

Appendix 3

Table 9.4 Base case parameter values and initial state variable values. The $*$ denotes a free parameter and is specified in the figures

Variable	Model Sect. 9.2	Model Sect. 9.3	Description
α	1/9	1/15	Reciprocal of average duration of the infection
$R_0^{1,2,3}$	4, $*$, 3.2	–	Level of infection risk
$\gamma_0^{1,2,3}$	1, 0.25, 0.75	–	Level of lockdown intensity
β_1	–	0	Minimum level of infection risk
β_2	–	0.2	Increment in the level of infection risk
$H_{\max}$	1.76×10^{-4}	2×10^{-4}	Capacity of intensive care units
p	2.311×10^{-2}	2.25×10^{-2}	Probability that infected person needs critical care
M	$*$	$*$	Social cost of a premature death due to COVID-19
K	1	1	Coefficient on economic activity
Γ	365	365	Reflects time required to return to full employment
f	0.75	–	Likelihood of spreading the virus when in the L state
f	–	$*$	Impact of lockdown fatigue on infection risk
κ_1	2×10^{-3}	–	Rate of decay for lockdown intensity
κ_1	–	0.15	Rate of accumulation of fatigue
κ_2	10^{-3}	–	Rate of decay for reproduction number
κ_2	–	0.2	Rate of exponential decay of fatigue
ω	0.6	–	Proportion of infections that become symptomatic
φ	1/7.2	–	Reciprocal of average duration of the latency
σ	2/3	2/3	Labor elasticity in Cobb-Douglas production function
ν	0.01/365	0	Birth rate
μ	0.01/365	0	Death rate (not caused by COVID-19)
μ_I	13/10800	0	COVID-19 death rate
ζ	5000	5000	Parameter in the approximation of the max-function
ξ_1	0.05	0.03	Death rate of infected individual in critical care
ξ_2	0.55/9	0.55/15	Incremental death rate if IC capacity is exceeded
ϕ	–	0.001	Rate by which recovered get susceptible again
c_l	–	1000	Parameter in business shutting down costs
c_r	–	5000	Parameter in business reopening costs
θ	2	2	Exponent in the proportionality function $\beta(t)$
$S(0)$	0.999	0.999	Initial susceptible population
$L(0)$	0.001	–	Initial latent population
$I(0)$	0	0.001	Initial infected population
$R(0)$	0	0	Initial recovered population
$\gamma(0)$	–	1	Initial employment level
$z(0)$	–	0	Initial lockdown fatigue

References

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

D. Acemoglu, *Introduction to Modern Economic Growth* (Princeton University Press, Princeton, 2009)

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

F.E. Alvarez, D. Argente, F. Lippi, A simple planning problem for COVID-19 lockdown. Am. Econ. Rev. Insights (forthcoming) (2020)

A. Aspri, E. Beretta, A. Gandolfi, E. Wasmer, Mortality containment vs. economics opening: optimal policies in a SEIARD model. J. Math. Econ. (2021)

D.E. Bloom, M. Kuhn, K. Prettner, Modern infectious disease: Macroeconomic impacts and policy responses. J. Econ. Liter. (2020)

A. Brodeur, D.M. Gray, A. Islam, S. Bhuiyan, A literature review of the economics of COVID-19. IZA Discussion Paper No. 13411 (2020)

J.P. Caulkins, D. Grass, G. Feichtinger, R. Hartl, P.M. Kort, A. Prskawetz, A. Seidl, S. Wrzaczek, How long should the COVID-19 lockdown continue? PloS One **15**(12), e0243413 (2020)

J.P. Caulkins, D. Grass, G. Feichtinger, R.F. Hartl, P.M. Kort, A. Prskawetz, A. Seidl, S. Wrzaczek, The optimal lockdown intensity for COVID-19. J Math. Econ. (2021)

S. Federico, G. Ferrari, Taming the spread of an epidemic by lockdown policies. J. Math. Econ. (2021)

J. Fernández-Villaverde, C.I. Jones, Macroeconomic outcomes and COVID-19: A progress report. Nber working paper 28004 (2020)

D. Gershon, A. Lipton, H. Levine, Managing COVID-19 pandemic without destructing the economy. arXiv preprint arXiv:2004.10324 (2020). https://ui.adsabs.harvard.edu/abs/2020arXiv200410324G/abstract

M. Gonzalez-Eiras, D. Niepelt, On the optimal 'lockdown' during an epidemic. CEPR Discussion Paper 14612 (2020)

D. Grass, J.P. Caulkins, G. Feichtinger, G. Tragler, D.A. Behrens, *Optimal Control of Nonlinear Processes: With Applications in Drugs, Corruption, and Terror* (Springer, Berlin, 2008)

J.K. Hammitt, Valuing mortality risk in the time of COVID-19. SSRN Electron. J. (2020). https://doi.org/10.2139/ssrn.3615314

N. Huberts, J. Thijssen. Optimal timing of interventions during an epidemic. Available at SSRN 3607048, University of York (2020)

W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics. Proc. Roy. Soc. London **115**(772), 700–721 (1927)

T.J. Kniesner, W.K. Viscusi, C. Woock, J.P. Ziliak, The value of a statistical life: evidence from panel data. Rev. Econ. Stat. **94**(1), 74–87 (2012). https://doi.org/10.1162/REST_a_00229

R. Layard, A. Clark, J.-E. De Neve, C. Krekel, D. Fancourt, N. Hey, G. O'Donnell, When to release the lockdown? A wellbeing framework for analysing costs and benefits. IZA Institute of Labor Economics IZA DP No. 13186 (2020)

F. Piguillem, L. Shi, Optimal COVID-19 quarantine and testing policies. Einaudi Institute for Economics and Finance EIEF Working Papers Series 20/04 (2020)

L. Rachel, *An Analytical Model of Covid-19 Lockdowns* (Technical report, Center for Macroeconomics, 2020)

A. Scherbina, *Determining the Optimal Duration of the COVID-19 Supression Policy: A Cost-Benefit Analysis* (Technical report, American Enterprise Institute, 2020)