FEDERAL RESERVE BANK OF MINNEAPOLIS

QR

QUARTERLY REVIEW

June 2020

On Using SIR Models to Model Disease Scenarios for COVID-19

Andrew G. Atkeson

FEDERAL RESERVE BANK OF MINNEAPOLIS

Quarterly Review Vol. 41, No.1

ISSN 0271-5287

https://doi.org/10.21034/qr.4111

This publication primarily presents economic research aimed at improving policymaking by the Federal Reserve System and other governmental authorities.

The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

The *Quarterly Review* is published by the Research Division of the Federal Reserve Bank of Minneapolis. This has become an occasional publication; however, it continues to be known as the *Quarterly Review* for citation purposes. Subscriptions are available free of charge. To subscribe to the journal and be automatically notified whenever a new issue is published, please sign up at https://www.minneapolisfed.org/economic-research/

On Using SIR Models to Model Disease Scenarios for COVID-19*

Andrew G. Atkeson†
University of California, Los Angeles
Federal Reserve Bank of Minneapolis

1 Introduction

In the face of the rapidly growing COVID-19 pandemic, public health experts are using a wide variety of models of its progression to generate scenarios that are being used to guide decisions about severe mitigation measures on economies worldwide.[1] While experts in the field of public health have previously debated the economic costs and public health benefits of these disease mitigation measures in the face of an influenza-type epidemic,[2] macroeconomists (with Rowthorn and Toxvaerd 2015 and McKibbin and Roshen 2020 being prominent exceptions) had not previously considered the feedback between disease progression and economic activity during an influenza pandemic under different public health and economic policy scenarios. This has now changed. To make sense of this public health and economic crisis, economists are building a large number of models with epidemiological and macroeconomic blocks.[3]

This paper is intended to introduce economists to a simple SIR model of the progression of COVID-19 to aid understanding of how such a model might be incorporated into more standard macroeconomic models. An SIR model is a Markov model of the spread of an epidemic in which the total population is divided into categories of being susceptible to the disease (S); actively infected with the disease (I); and resistant (R), meaning those that have recovered, died from the disease, or have been vaccinated.[4] The initial distribution of the population across these states and the transition rates at which agents move between these three states determine how an epidemic plays out over time. These transition rates are determined by characteristics of the underlying disease and by the extent of mitigation and social distancing measures. This model allows for quantitative statements regarding the tradeoff between the severity and timing of suppression of the disease through social distancing and the progression of the disease in the population.

*This paper is a substantially revised version of Federal Reserve Bank of Minneapolis Staff Report #595 and NBER working paper #26867. All errors are mine.

†Please send corrections or suggestions to andy@atkeson.net.

Sample applications of the model are provided. Special attention is given to the determinants of the peak fraction of agents actively infected with the disease (the peak prevalence of the disease) and the fraction of the population that is ultimately infected with the disease (the cumulative disease burden). The peak prevalence of the disease is of interest because of concerns about overloading the health care system, while the cumulative disease burden is of interest because of its connection to the overall loss of life and labor input due to the disease.[5]

This paper extends Atkeson 2020b in discussing the analytical solutions developed by Harko, Lobo, and Mak 2014 and presented in Toda 2020 for the peak prevalence of the disease and the cumulative disease burden when the transmission rate of the disease is held constant. Analytical formulas allow systematic study of the extent to which temporary mitigation measures could reduce subsequent peak prevalence of the disease and the cumulative disease burden.[6] I present conditions under which temporary mitigation measures simply postpone the peak prevalence and cumulative burden of the disease or make it possible to reduce peak prevalence and the cumulative burden of the disease.[7] The central finding here is that it is possible to reduce the disease's peak prevalence and cumulative burden through a policy of temporary disease mitigation that leads a substantial fraction of the population to become resistant to the disease before the mitigation policy is lifted permanently.[8]

This potential for temporary disease mitigation efforts to permanently affect disease progression underlies the finding in the rapidly emerging literature on the optimal control of COVID-19 through mitigation measures that costly mitigation measures such as lockdown should be applied later rather than earlier in the evolution of the epidemic.[9]

This paper extends Atkeson 2020a in discussing an approach to estimate the SIR model for the purpose of forecasting the evolution of the epidemic going forward. The estimation approach discussed here builds on ongoing work with Karen Kopecky and Tao Zha in Atkeson, Kopecky, and Zha 2020. I compare this structural approach for estimation and forecasting with an SIR model with noteworthy reduced-form forecasting approaches.[10]

This estimation approach does not get around the fundamental problem discussed in Lourenco et al. 2020, Atkeson 2020a and Stock 2020 of identifying key parameters of the model from time series data on deaths from the disease if good testing data on the number of active infections and immune agents is not available. So the estimation and forecasting results presented are conditional on outside estimates of the fatality rate and recovery rate of the disease.

In comparing estimation and forecasting methodologies based on the SIR model versus reduced-form empirical models, I focus in particular on a stylized version of the widely cited model developed by researchers at the Institute for Health Metrics and Evaluation (IHME) at the University of Washington.[11] See also Linton 2020[12] for a similar reduced-form forecasting approach. I explore the underlying reasons why these reduced-form forecasting models tend to produce forecasts for cumulative deaths that are more optimistic than those of the structural SIR forecasting models based on different implicit and explicit assumptions made about the future course of the disease's effective reproduction number under these two approaches.[13] In short, the reduced form approach used by the team at IHME assumes that the effective reproduction number will continue to decline going forward while prominent forecasts using structural models assume that this number will remain the same or even rise as social distancing policies are relaxed. I show that a simple SIR model implies that if the

transmission rate of the disease as measured by its effective reproduction number were to rise, then a second wave of the epidemic would quickly appear.

This paper introduces this basic epidemiological model and its quantitative implications as follows.

In section 2, I present the equations of a basic SIR model. I derive the standard conditions under which an epidemic grows or dies out in the population. I use these conditions to frame the options for mitigating a highly contagious disease like COVID-19.

In section 3, I review the analytical solution of the model with constant parameters. In section 3.1.1, I use this analytical solution together with numerical illustrations to review how the disease's peak prevalence and cumulative burden depend on model parameters. These calculations illustrate how permanent mitigation of transmission of the disease allows one to reduce the peak prevalence of the disease ("flatten the curve") and the cumulative burden of the disease over the long term. Following Toda 2020, in section 3.1.2, I then use these analytical results to consider the impact of temporary mitigation measures followed by a period of umitigated or partially mitigated transmission of the disease at a constant rate. Finally, in section 3.1.3, I illustrate the benefits of delayed temporary mitigation in a specific numerical example.

In section 4, I discuss the sources of data used to choose the parameters of the model. In section 5, I discuss reduced-form and structural approaches to estimating and forecasting the progress of the epidemic. I use a structural approach to construct several scenarios for the epidemic going forward based on a back-of-the-envelope estimation of the model. In section 6, I conclude.

2 A Basic SIR Model

The model is as follows.

The population is set to N. At each moment of time, the population is divided into three categories (states) that sum to the total population. These states are susceptible S, infected I, and resistant R. Agents who are susceptible are at risk of getting the disease. Agents who are infected are contagious and may pass on the disease to others through some form of interaction with susceptible agents.[14] Agents who are resistant are not at risk of getting the disease because they have immunity built up from a vaccine, have immunity from previous experience with this or similar diseases, or they have died. In this specification of the model, we assume that immunity is permanent so that being resistant R is an absorbing state.[15] We normalize the total population $N = 1$, so all results regarding S, I, and R should be interpreted as fractions of the relevant population.

The initial distribution of the population across these states at time $t = 0$ is given by $S_0 > 0$, $I_0 > 0$, and $R_0 \geq 0$. For a new disease such as COVID-19, we assume that all agents are at risk of getting the disease, so that $R_0 = 0$, S_0 is very close to one, and I_0 is a small number corresponding to the initial cases of the disease transmitted to humans either from some animal source (as in Wuhan) or introduced into a country or region through travel.

These fractions of the population evolve over time as follows:

$$dS_t/dt = -\beta_t S_t I_t,$$

$$dI_t/dt = \beta_t S_t I_t - \gamma I_t,$$

$$dR_t/dt = \gamma I_t.$$

Since all of the parameters in these equations are positive, agents flow only in one direction, from the state S to the state I and then to state R.

Note as well that we treat the evolution of the population as deterministic. This may be appropriate once the disease has infected a large number of individuals, but at the early stage of an epidemic with a small number of infected agents, it is more appropriate to think of the evolution of the number of infected agents as stochastic because of the small number of these agents. We abstract from that issue here, but it is of substantive importance if one contemplates the possibility of completely eliminating the disease (driving I_t from a positive number to zero).

The parameters of the model can be interpreted as follows.

The parameter γ governs the rate at which agents who are infected become resistant (transition from the state I to the state R) and hence stop transmitting the disease. I refer to this parameter as the *recovery rate*. Because there is no cure for a viral disease such as COVID-19, this parameter is considered a fixed parameter determined by the biology of the disease.[16]

The parameter β_t is the rate at which infected agents spread the virus to (or "shed" the virus onto) others. I refer to this parameter as the *transmission rate*. This parameter is a reduced-form parameter that is impacted by the biological disease transmission mechanism, the rate at which agents bump into one another in the course of their daily activities, and the extent to which agents use prophylactics in their meetings. This parameter can thus be impacted by mitigation measures such as social distancing, hand washing, the use of masks, and so on. Because of changes in the biological disease transmission mechanism, this parameter is also subject to natural, or apparently random, fluctuations over time and across space.[17]

Note in the first two equations governing the flow of agents over time from the state S to the state I, we assume that the transmission of the disease from infected to susceptible agents is mitigated through random and uniform matching of agents in the population, as indicated by the term S_t in those two equations. That is, we interpret β_t as capturing the rate at which an infected agent interacts with and sheds virus onto agents of any kind in the population.[18] Under the assumption that infected agents' interaction with other agents is random and uniform, the the rate at which an infected agent meets a susceptible agent and sheds virus onto that agent is given by $\beta_t S_t$. This assumption that the transmission of the disease is mitigated by random and uniform matching is a very stark one maintained here for simplicity, and there is a great deal of research on the question of how the precise nature of interaction between agents of different types, together with mitigation policies targeted at reshaping the specific nature of interactions, may shape the transmission of a disease.[19]

The following notation is useful in developing results regarding the solution to the model.

I define the ratio of parameters β_t/γ to be the *normalized transmission rate*. It is standard to refer to the value of the normalized transmission rate at the start of the epidemic, before any mitigation measures and use of prophylactics are undertaken, as the *basic reproduction number* of the disease. We denote this basic reproduction number by $\mathcal{R}_0$. This parameter corresponds to the parameter cited in many news and academic studies.[20]

I refer to the term $\mathcal{R}_t$ as the *effective reproduction number* of the disease at date t. This effective reproduction number is the ratio of the rate at which infected agents infect

susceptible agents when the the disease has progressed for some time to the recovery rate of infected agents. In the model, we assume that the effective reproduction number of the disease is given by the product of the normalized transmission rate and the fraction of agents who remain susceptible to the disease, $\mathcal{R}_t = (\beta_t/\gamma)S_t/N$. Thus, in the model, the effective reproduction number of the disease can differ from the basic reproduction number for two reasons. First, as discussed above, the normalized transmission rate β_t/γ may vary over time with changes in the transmission rate β_t, through either steps undertaken to mitigate the transmission rate of the disease or through naturally occurring changes in its transmission. Second, the effective reproduction number falls as the fraction of agents remaining susceptible to the disease S_t/N falls.

With this notation, we can restate the equations of the model in terms of the effective reproduction number as

$$dS_t/dt = -\mathcal{R}_t\gamma I_t, \tag{1}$$

$$dI_t/dt = (\mathcal{R}_t - 1)\,\gamma I_t, \tag{2}$$

$$dR_t/dt = \gamma I_t. \tag{3}$$

This reformulation of the model equations in terms of the effective reproduction number is useful for measurement of the effective reproduction number from data on either the growth of active cases or cumulative deaths.

We are interested in the *cumulative burden of the disease*, measured as the fraction of the population ever infected with the disease. We denote this fraction at date t as CB_t. The initial value of this fraction is equal to the fraction initially infected I_0. This fraction grows over time according to

$$dCB_t/dt = -dS/dt = \mathcal{R}_t\gamma I_t. \tag{4}$$

The long-run cumulative burden of the disease is given by

$$CB_\infty = I_0 + S_0 - S_\infty. \tag{5}$$

where S_∞ is the limit of S_t as t goes to infinity.

We denote the fatality rate from the disease by ν and the cumulative number of fatalities by D_t. That is, ν is the fraction of agents who are newly resistant because they died.[21] We set the initial number of deaths to $D_0 = 0$. From equation (3), the death rate per unit time is given by

$$dD_t/dt = \nu\gamma I_t. \tag{6}$$

Cumulative deaths are then counted by integrating this death rate over time.

The following properties of the solution of the model are standard.

1. Model steady-states have $I = 0$. If $I = 0$, then any combination of S and R that sums to N is a steady-state.
2. From equation (2), if $I_t > 0$, then $dI_t/dt \leq 0$ if and only if $\mathcal{R}_t \leq 1$. Thus, the steady-states reached from an initial value of $I > 0$ must have

$$S_t \leq \gamma/\beta_t. \tag{7}$$

3. The growth rate of the log of the number of active cases is given by

$$\frac{d \log I}{dt} = (\mathcal{R}_t - 1)\,\gamma. \tag{8}$$

Note from equation (2) that if the population is distributed across states such that the condition (7) is satisfied for values of the normalized transmission rate equal to the basic reproduction number $\mathcal{R}_0$, then any initial infection rate $I_0 > 0$ would die out monotonically over time. In this case, we say that this population has *herd immunity* to the disease, since initial infections do not spread further. Note that herd immunity can be achieved either if the disease infects enough agents to reduce the fraction of agents remaining susceptible S/N to a low enough level relative to the normalized transmission rate so that condition (7) is satisfied or through the use of a vaccine to directly convert enough agents from the susceptible state S to the resistant state R. In the context of this model, in the absence of a vaccine, the only option to reduce the infection rate I when the population does not have herd immunity is to take mitigation steps to reduce the normalized transmission rate to a level consistent with condition (7), as determined by the current distribution of agents across states.

3 An Analytical Solution of the Model with a Constant Transmission Rate

In this section, I review the analytical solution of the model with constant parameters developed by Harko, Lobo, and Mak 2014 and presented in Toda 2020. This analytical solution is useful in at least two dimensions. First, we can use the solution, together with initial conditions corresponding to the start of this epidemic (S_0 close to one, I_0 positive but close to zero, and $R_0 = 0$), to study how the shape of the epidemic, in terms of the disease's peak prevalence and long-run cumulative burden of the disease, depends on the normalized transmission rate β_t/γ if it is held constant at $\mathcal{R}_0$.

Second, we can use the solution, together with arbitrary initial conditions, to study how the epidemic might evolve following an initial period of mitigation that is lifted if the normalized transmission rate is held constant after that time. For this second application, we will consider initial conditions with the fraction of agents who are resistant $R_0 > 0$, corresponding to the state of the population after an initial period of evolution of the epidemic. The idea here is that after some initial period with temporary mitigation, the epidemic progresses from that point on with the normalized transmission rate equal to a constant. Thus, we see that a temporary period of mitigation can impact long run outcomes only through its impact on the state of the population at the end of the period of mitigation, taken here as the initial condition for the model of the remainder of the epidemic.

Let the initial conditions of the model be given by $S_0 > 0$, $I_0 > 0$, and $R_0 \geq 0$. Let the recovery rate be given by γ, and let the normalized transmission rate be given by $\beta_t/\gamma = \mathcal{R}_0$ and be constant over time.

The solution of the model for S_t, I_t, and R_t is expressed in terms of a decreasing function of time $v(t)$ defined implicitly below. This solution is

$$S_t = S_0 v(t), \tag{9}$$

$$I_t = \frac{1}{\mathcal{R}_0}\log v(t) - S_0 v(t) + 1 - R_0, \tag{10}$$

$$R_t = -\frac{1}{\mathcal{R}_0}\log v(t) + R_0, \tag{11}$$

with $v(t) \in (0,1]$ implicitly defined as the solution to

$$t = \frac{1}{\beta}\int_{v(t)}^{1} \frac{1}{\zeta f(\zeta)} d\zeta, \tag{12}$$

where the function $f(v)$ is defined on a domain $(v^*, 1]$ by

$$f(v) = S_0(1-v) + I_0 + \frac{1}{\mathcal{R}_0}\log v.$$

To understand the definition of $v(t)$ in equation (12), observe that $f(1) = I_0 > 0$ and $f(v)$ is strictly concave since

$$f''(v) = -\frac{1}{\mathcal{R}_0}\frac{1}{v^2} < 0.$$

We also have that $f(v) \to -\infty$ as $v \to 0$, so there is a unique value of $v^* \in (0,1)$ such that $f(v^*) = 0$. Moreover, $f(v) > 0$ for $(v^*, 1]$. This implies that $v(t)$ is a decreasing function of t, that it is equal to 1 when $t = 0$, and that it approaches v^* as t goes to infinity.

We see from this analytical solution of the model that when the normalized transmission rate is constant, the basic reproduction number $\mathcal{R}_0$ plays an important role in shaping the model's predictions for the shape of the progression of the epidemic in terms of the its peak prevalence and cumulative burden. But this parameter does not determine the time scale of the epidemic's progression.[22] The speed of the epidemic is determined by the time derivative of v, which is obtained by differentiating the equation defining $v(t)$ and is given by

$$-\frac{dv}{dt} = \beta v f(v) = \mathcal{R}_0 \gamma v f(v).$$

We see that speed of the epidemic scales directly with the recovery rate γ, holding fixed the basic reproduction number.

We now use these formulas to compute the model's implications for the peak prevalence cumulative burden of the disease. To compute the peak prevalence of the disease, observe that from equation (2) and our discussion of herd immunity above, we have two possibilities. In the first case, with herd immunity, if $\mathcal{R}_0 \leq 1$, then from equation (2), the solution for $I(t)$ is declining over time for all $t \geq 0$, so the peak prevalence of the disease is the initial infection rate I_0. In the alternative case, with $\mathcal{R}_0 > 1$, from equation (2), the peak prevalence of the disease occurs when $\mathcal{R}_0 S_t = 1$ or, using equation (9), when $v = 1/\mathcal{R}_0 S_0$. Thus, from equation (10), the peak prevalence of the disease is given by

$$I_{peak} = -\frac{1}{\mathcal{R}_0}\log\left(\mathcal{R}_0 S_0\right) - \frac{1}{\mathcal{R}_0} + 1 - R_0. \tag{13}$$

To compute the cumulative burden of the disease, we compute the fraction of agents who get infected in the long run. From equations (5) and (9), we have

$$CB_\infty = -\frac{1}{\mathcal{R}_0}\log v^*, \tag{14}$$

where v^* is given by the unique solution in $(0,1)$ of the equation $f(v)=0$ or, equivalently,

$$1 - R_0 - v^* S_0 + \frac{1}{\mathcal{R}_0}\log v^* = 0. \tag{15}$$

3.1 Quantitative Implications of the Model with a Constant Transmission Rate

I now use these formulas to study two types of disease scenarios of interest.

I first use these formulas to construct disease scenarios in which the disease is entirely new to the population, so that initially there are no agents resistant to the disease ($R_0 = 0$), and the fraction initially susceptible $S_0 = 1 - I_0$ is very close to one because the fraction initially infected is very small. In using these formulas to obtain the model's implications starting from these initial conditions, we assume either that there is no mitigation of the disease (so that $\mathcal{R}_t = \mathcal{R}_0$ for all $t \geq 0$) or that permanent steps are taken to lower for all time the normalized transmission rate of the disease.

I then use these formulas to construct disease scenarios starting from initial conditions in which some sizable portion of the population is resistant (so $R_0 \gg 0$) and a non-negligible fraction of the population is initially infected ($I_0 \gg 0$). We can interpret scenarios starting from these initial conditions as scenarios for the disease after an initial period of experience with it under temporary mitigation measures.

I finally illustrate how, using scenarios of this kind, we can understand whether it is possible to use temporary mitigation measures to alter the peak prevalence and cumulative burden of the epidemic. I solve the model under assumptions of no mitigation and temporary mitigation applied for 30 days, but at different points in the progression of the epidemic. I show how temporary mitigation applied early has little impact on long run outcomes, while the same temporary mitigation applied later can have a significant impact on long run outcomes because of its impact on the state of the population when that temporary mitigation comes to an end.

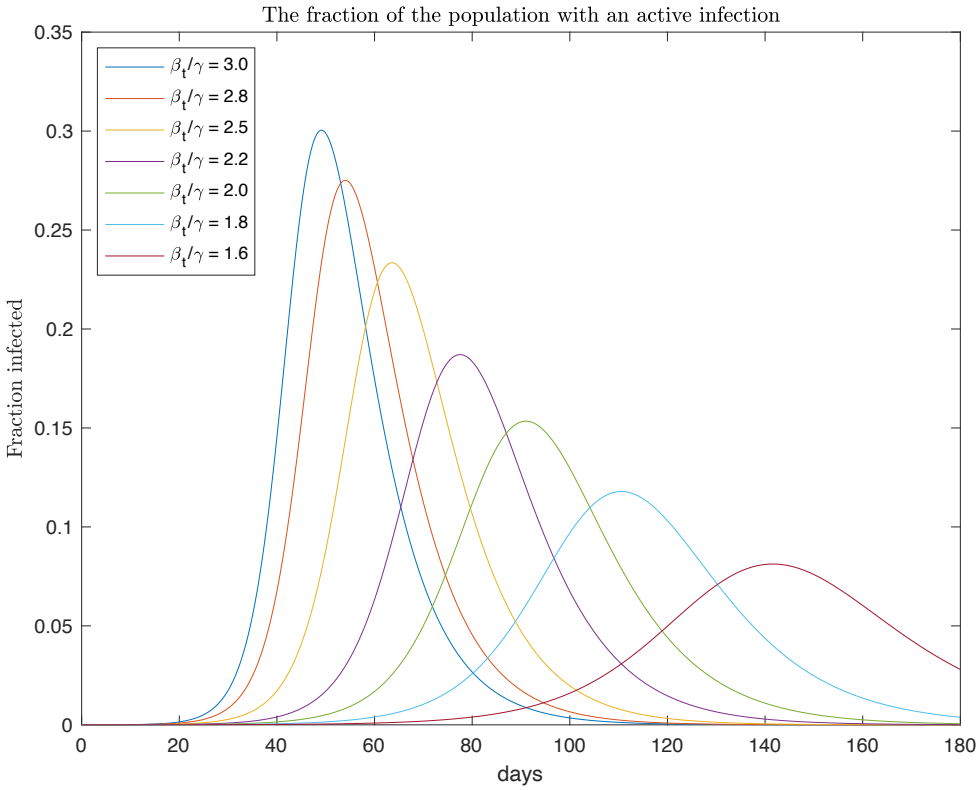
3.1.1 Disease Scenarios Starting with No Resistant Agents in the Population

I start by considering the quantitative implications of the model in the first case, with no agents resistant to the disease to start and an initial importation of cases through travel equal to $I_0 = 1/100000$, or roughly 3,300 initial cases for a population the size of the United States. I set the recovery rate to $\gamma = 1/8$, corresponding to infected agents' being contagious on average for eight days. I illustrate the model's quantitative implications for the peak prevalence and cumulative burden of the disease over the first 180 days of the epidemic, when the normalized transmission rate of the disease ranges from $\beta_t/\gamma = 3.0$ to $\beta_t/\gamma = 1.6$,

corresponding to a doubling time for active cases in the initial phase of the epidemic of 2.8 days and 9.2, respectively. I show results in figures 1 and 2.

Figure 1

Fraction of the population with an active infection over 180 days under different values of β_t/γ held constant over the entire six-month time period.



Note: The initial fraction of the population that is resistant is set to $R_0 = 0$.

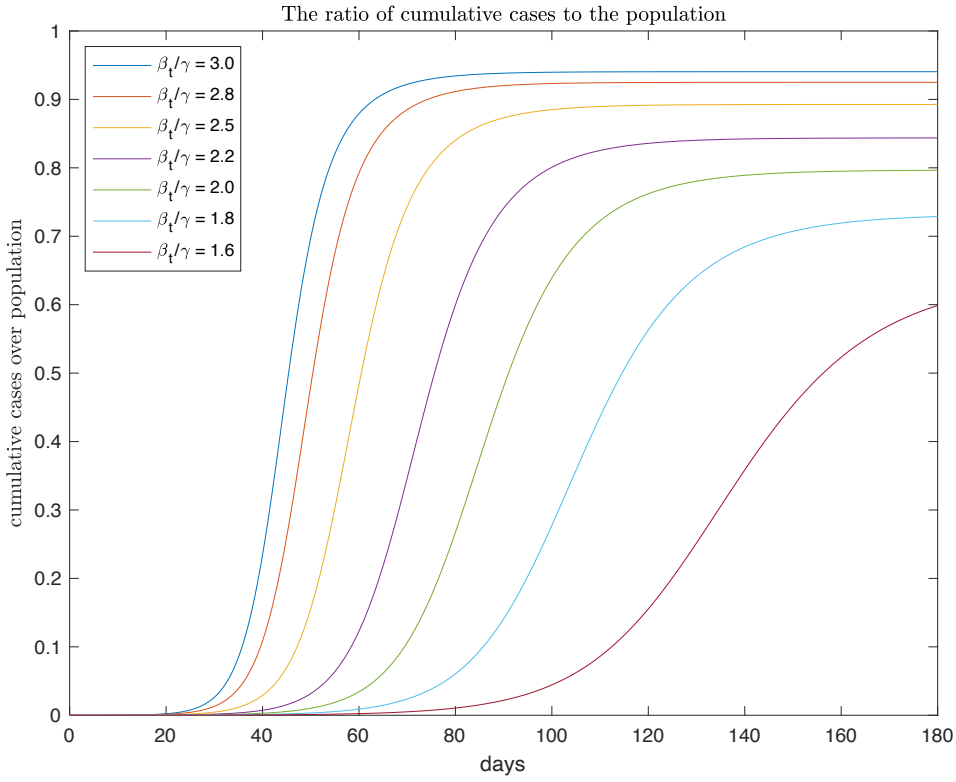
Note in Figure 1 that variation in the basic reproduction number $\mathcal{R}_0 = \beta_t/\gamma$ results in substantial changes in the peak prevalence of the disease, as indicated by equation (13) with initial conditions $R_0 = 0$ and S_0 very close to 1. The timing of the peak in seen in Figure 1 slows down with reductions in the basic reproduction number $\mathcal{R}_0$ because these reductions imply a slower transmission rate per unit of time β_t given our fixed value of the recovery rate γ.

In figure 2, we see that the cumulative burden of the disease (shown here after 180 days) exceeds half of the population in all of the numerical illustrations considered and is close to one if the basic reproduction number for the disease is as high as three. These illustrations suggest that unmitigated transmission of COVID-19 would result in a large cumulative burden of the disease over a relatively short period of time.

What does the model imply for fatalities from the disease? The cumulative fraction of the population that dies in the long run is given by $D_\infty = \nu C B_\infty$. With estimates of the fatality rate ν from COVID-19 in the range of one tenth of one percent at the low end to

Figure 2

The cumulative burden of the disease as measured by cumulative cases as a fraction of the population over 180 days under different values of β_t/γ held constant over the entire six month time period.



Note: The initial fraction of the population that is resistant is set to $R_0 = 0$.

one percent or more, the results in Figure 2 and the corresponding figures for the long-run cumulative disease burden suggest that one would expect a substantial number of deaths in the United States from unmitigated transmission of the disease: roughly 200,000 deaths at the low end and several million deaths at the high end.

3.1.2 Continuation of Disease Scenarios Once There is a Substantial Fraction of Resistant Agents in the Population

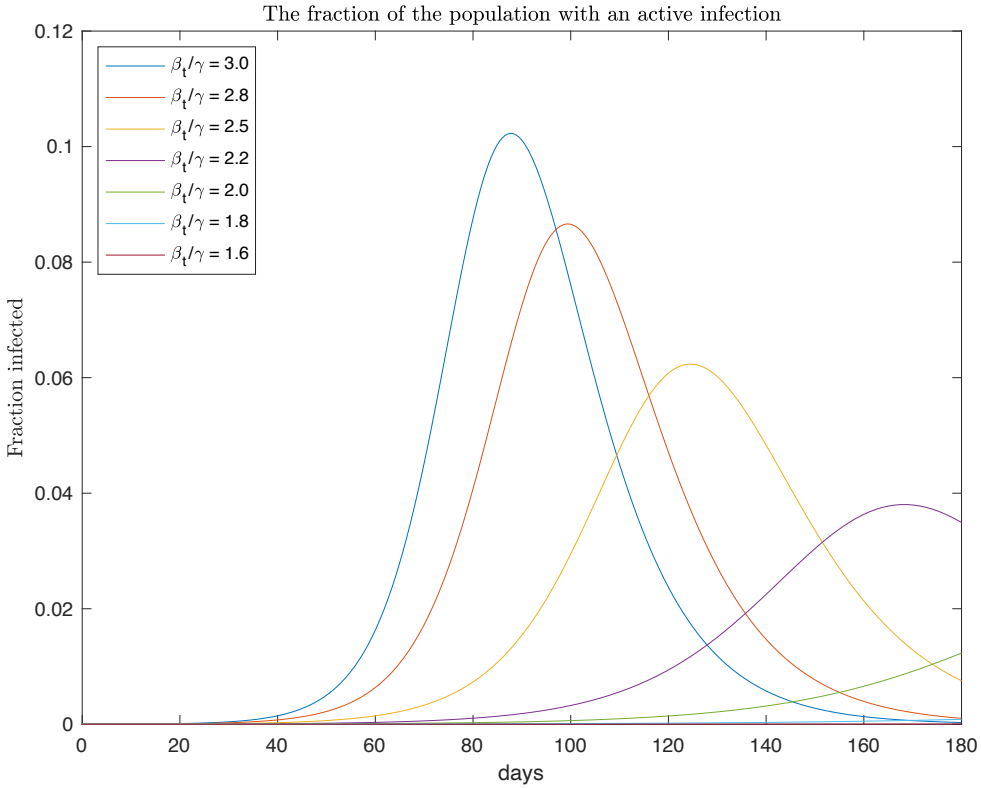
I next consider the quantitative implications of the model in the second case, in which the population has an initial period of experience with the disease which leads to some fraction of the population's being resistant. Here, for illustrative purposes, I set this initial fraction to $R_0 = 1/3$. I set the other parameters of the model equal to their values considered above ($\gamma = 1/8$, $I_0 = 1/100000$, and values of β_t/γ held constant at values ranging from 1.6 to 3.0).

In figures 3 and 4, I show the model's implications for the peak prevalence and cumulative

disease burden over the next 18 months. Here, I consider the peak prevalence of the disease in the period after the initial period of experience with the disease which results in $R_0 = 1/3$ of the agents' being resistant. But I do count these initially resistant agents as part of the cumulative burden of the disease.

Figure 3

Fraction of the population with an active infection over 180 days under different values of β_t/γ held constant over the entire six-month time period.



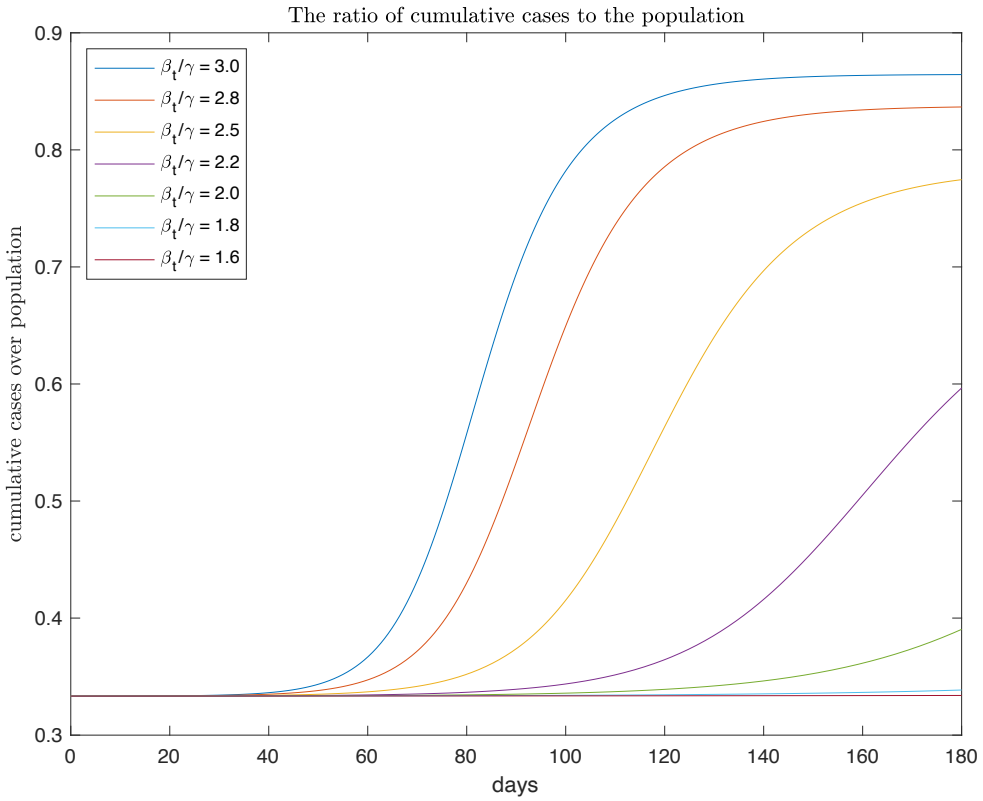
Note: The initial fraction of the population that is resistant is set to $R_0 = 1/3$.

By comparing figures 1 and 3, we see that the continuation of the epidemic after an initial period that led to one-third of the population's being resistant to the disease features much lower peak prevalence of the disease than we found when the epidemic started with no one's being resistant to the disease. This finding suggests that it is possible, through a policy of temporary mitigation of disease transmission, to then relax those mitigation measures and experience a substantially reduced peak prevalence of the disease in a second wave. This is the case only if the initial period of disease transmission subject to mitigation measures allows a substantial portion of the population to become resistant to the disease.

Perhaps more interesting, by comparing figures 2 and 4 showing the cumulative burden of the disease over 180 days in our two scenarios, we see that it may be possible to substantially reduce the cumulative burden of the disease through a policy of temporary mitigation of

Figure 4

The cumulative burden of the disease as measured by cumulative cases as a fraction of the population over 180 days under different values of β_t/γ held constant over the entire six month time period.



Note: The initial fraction of the population that is resistant is set to $R_0 = 1/3$.

disease transmission followed by a relaxation of those mitigation measures once a substantial portion of the population has become exposed to the disease. I present corresponding results for the cumulative disease burden in the long run (CB_∞) computed using equation (14) in Table 1. We see that there is indeed a permanent difference in the cumulative disease burden in the two scenarios.

In addition, we see in figure 4 that this cumulative disease burden grows very slowly if the normalized transmission rate is held to a level such as 1.6 or 1.8, so that, in those cases, the difference in disease burden after 180 days is much more substantial than the corresponding difference in the long-run disease burden shown in Table 1.

3.1.3 Choosing the Timing of Temporary Mitigation

The results from section 3.1.2 and Toda 2020 suggest that temporary mitigation may be most useful in reducing peak disease prevalence and the cumulative burden of the disease

Table 1
Long-run cumulative disease burden $\%$ of total population

β_t/γ	1.6	1.8	2.0	2.2	2.5	2.8	3.0
No initial resistance	0.6420	0.7324	0.7968	0.8437	0.8926	0.9250	0.9405
$1/3$ initially resistant	0.4150	0.5425	0.6363	0.7070	0.7838	0.8375	0.8645

if it is applied later in the progression of the epidemic rather than earlier. When applied later, it allows for the build-up of a substantial number of resistant agents in the population before the temporary mitigation is relaxed. To illustrate this point, I show results for the peak prevalence of the disease and the cumulative burden of the disease from a numerical example comparing the progression of the epidemic with no mitigation, 30 days of mitigation applied from days 20 to 50, and 30 days of equally effective mitigation applied from days 55 to 85.

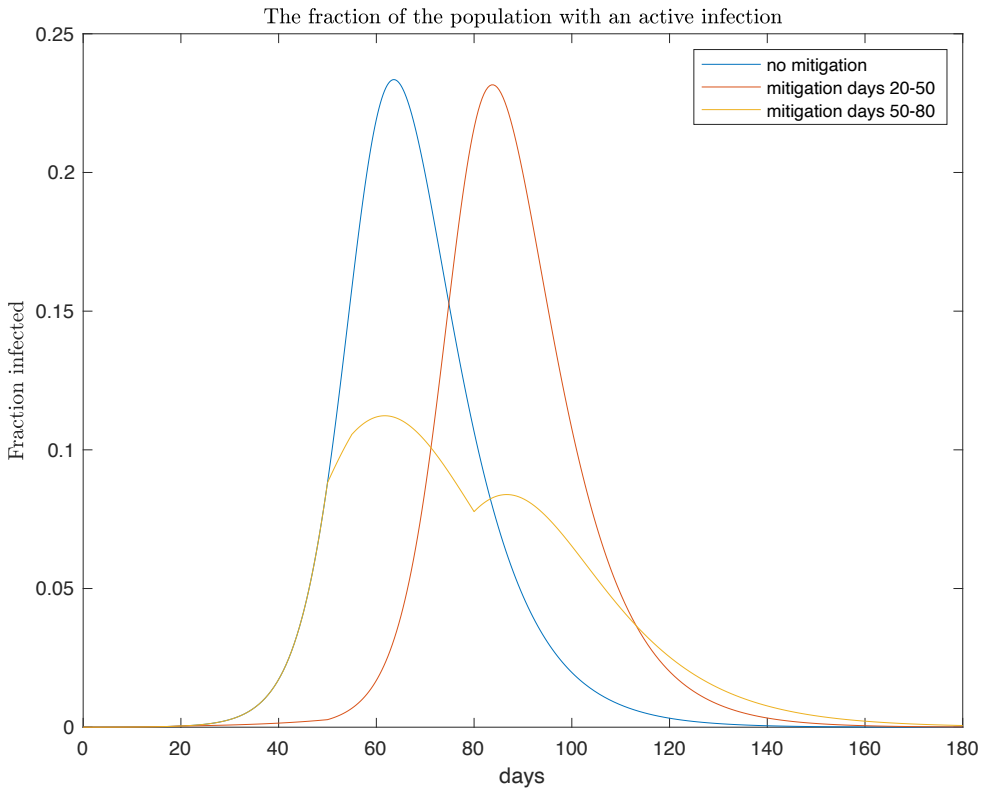
In this numerical example, as before, I set the initial conditions as before to $I_0 = 1/100000$ and the initial fraction resistant to $R_0 = 0$. I set the recovery rate to $\gamma = 1/8$ and the normalized transmission rate of the disease in the absence of mitigation to $\beta_t/\gamma = 2.5$. I assume that mitigation, when applied, reduces the normalized transmission rate of the disease to $\beta_t/\gamma = 1.5$, or to 60% of its unmitigated value. I solve the model under three scenarios. In the first scenario, there is no mitigation. In the second scenario, mitigation is applied for 30 days from $t = 20$ to 30. In the third scenario, mitigation is applied for 30 days from $t = 50$ to 80.

In figures 5 and 6, I show the actively infected fraction of the population and the cumulative burden of the disease over 180 days for these three scenarios. We see in these figures that temporary mitigation applied for 30 days starting on day 20 of the epidemic simply delays the peak prevalence of the disease and the cumulative burden of the disease by 30 days relative to the case with no mitigation at all, while the same temporary mitigation applied from days 50 to 80 substantially reduces the peak prevalence of the disease and has a measurable impact on the cumulative burden of the disease. As indicated by the analytical formulas above, the key difference between the two cases with temporary mitigation is that in the case with temporary mitigation applied early, only a very small fraction of the population is resistant at the end of the temporary mitigation period (day 50) while in the case of temporary mitigation applied later, nearly 60% of the population is resistant at the end of the temporary mitigation period (day 80).

Rachel 2020 provides useful intuition for this result regarding how the impact of temporary disease mitigation measures on the peak prevalence of the cumulative burden of the disease varies with the timing of the mitigation based on an examination of the phase diagram of the SIR model. This phase diagram is presented as a plot with the fraction of agents susceptible S_t on the x-axis and the fraction of the population actively infected I_t on the y-axis, for all points of time t. Since the fractions of the population across states S, I, and R sums to one, we can characterize the current state of the system in terms of the current fractions susceptible S_t and actively infected I_t, with the fraction resistant R_t computed as a residual.

We present the solution of the model with different timing of mitigation shown over time in figures 5 and 6 in this phase diagram in figure 7. The vertical line at $S = 0.4$ marks the point that the population reaches herd immunity, since the basic reproduction number in

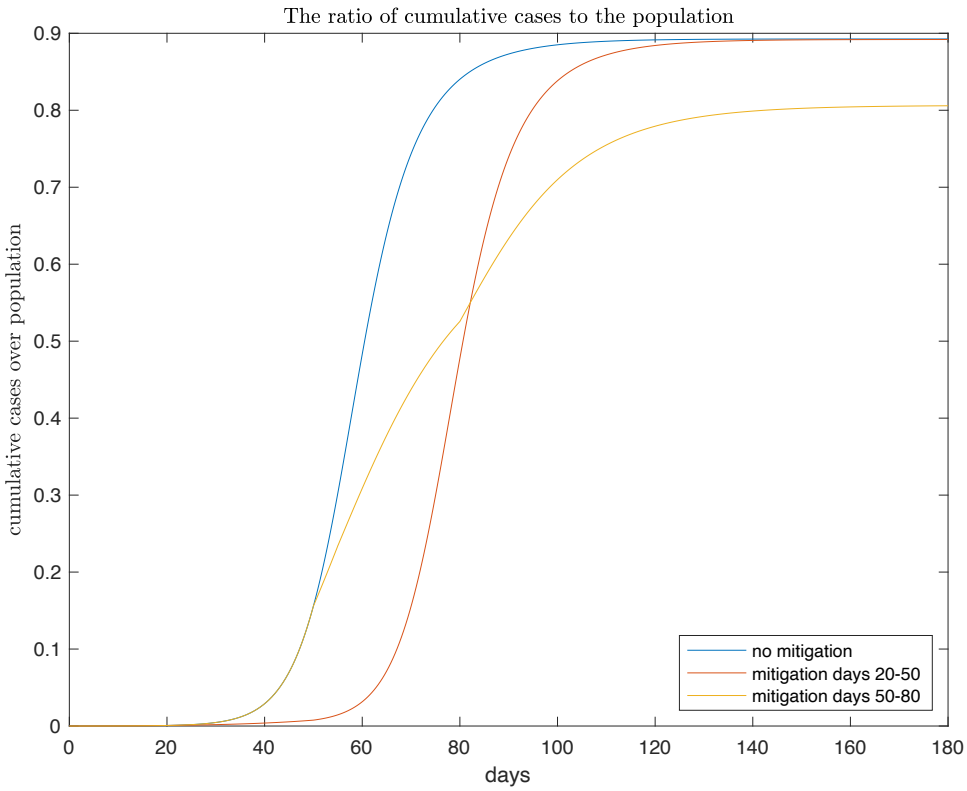
Figure 5
Fraction of the population with an active infection over 180 days under different temporary disease mitigation scenarios.



Note: The normalized transmission rate is set to $\beta/\gamma = 2.5$.

this case is set to $\mathcal{R}_0 = 2.5$. The solution of the model under all three temporary mitigation scenarios considered starts in the lower right hand corner of the diagram, with S_0 close to one and I_0 close to zero. Absent mitigation, I_t rises and S_t falls over time as long as S_t is above the line at 0.4 marking herd immunity, and I_t falls and S_t also falls over time when S_t is below this level. This means that the long-run value to which S_t converges in the lower right hand side of the diagram, equal to one minus the long-run cumulative burden of the disease, is determined by the level of I_t when herd immunity is reached (or when temporary mitigation ends, if that comes later). As we can see in the diagram, early temporary mitigation has a negligible impact on the level of active infections when herd immunity is reached, in comparison to the alternative of no mitigation at all. In contrast, the same temporary mitigation applied later on in the progression of the epidemic does substantially reduce the fraction actively infected when herd immunity is reached, leading to higher long-run values of S_t and thus a lower cumulative burden of the disease.

Figure 6
The cumulative disease burden measured as cumulative cases as a fraction of the population over 180 days under different temporary disease mitigation scenarios.



Note: The normalized transmission rate is set to $\beta/\gamma = 2.5$.

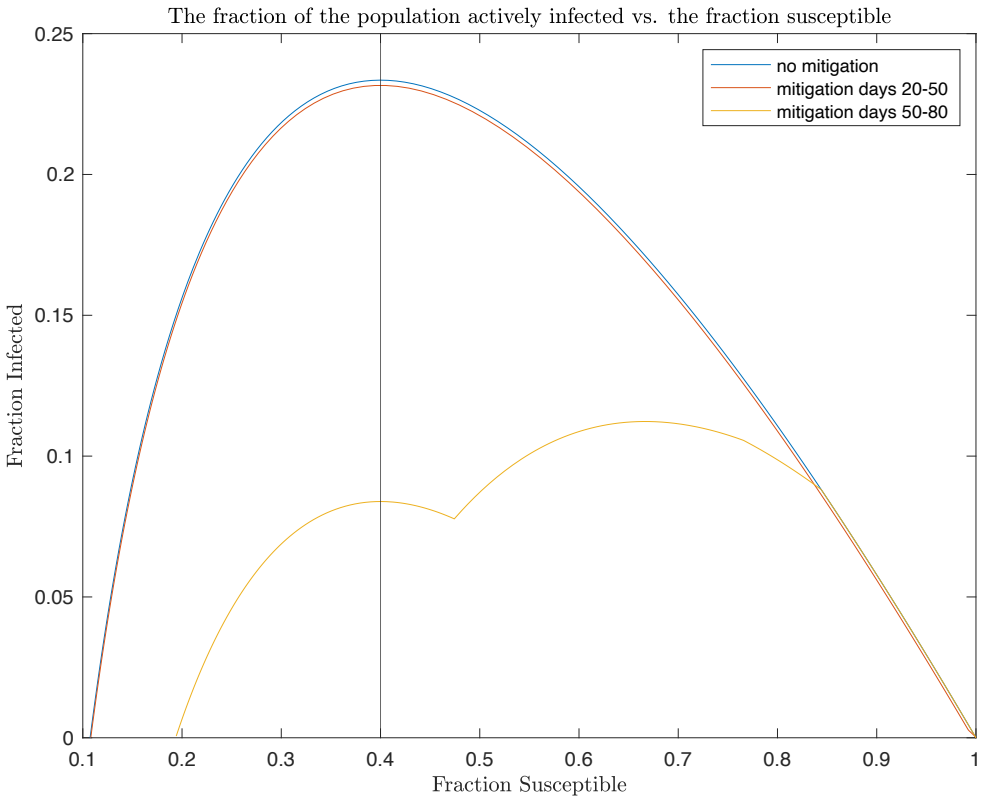
4 Measuring the Parameters of the Model

In this section, I discuss data used to measure the parameters of the model.

The model has the following parameters. There is the initial condition of the population indexed by S_0 and I_0 (with the initial population resistant at $R_0 = 1 - S_0 - I_0$), the recovery rate γ, the basic reproduction number $\mathcal{R}_0$, and the evolution over time of the normalized transmission rate β_t/γ as it changes with mitigation and natural variation. These parameters imply a path for the transmission rate β_t and the effective reproduction number $\mathcal{R}_t$ depending on the solution for the fraction of susceptible agents S_t at each date.

Consider first the measurement of the initial conditions for a new disease such as COVID-19. We start from the assumption that there are no agents resistant to the disease, so $R_0 = 0$. As mentioned above, outside of Wuhan, the initial number of infected agents corresponds to the number of cases of the disease introduced into the area through travel. As a fraction of the total US population, this number is typically set to be quite small (on the order of $I_0 = 1/100000$).[23]

Figure 7
A phase diagram showing the evolution of the fraction of the population actively infected versus the fraction of the population still susceptible under three temporary disease mitigation scenarios.



Note: The vertical line in the figure marks the fraction susceptible at which herd immunity is achieved $S = 0.4 = \gamma/\beta$ with $\beta/\gamma = 2.5$

Now consider estimates of the recovery rate γ. There are a wide range of estimates of this parameter taken from clinical observations of data such as the length of time agents known to be infected shed the virus[24] as well as data from contact tracing determining who got sick when and from contact with whom.[25] Estimates for COVID-19 continue to be updated as new data come in.[26] Values of γ between one-fourth and one-fourteenth are considered in the literature, corresponding to an infectious period of four to 14 days on average. In the numerical illustrations in this paper, I use $\gamma = 1/8$.

How do we measure the basic reproduction number of the disease $\mathcal{R}_0$? This is typically done using data on the growth of the number infected in the early phase of the epidemic using equation (8).[27] For example, estimates of the doubling time of the number of active cases for COVID-19 in the early stages of the epidemic are typically in the range of two to nine days. From equation (8), the doubling time of active cases should be given by $\log(2)/(\mathcal{R}_t - 1)\gamma$. Thus, with $\gamma = 1/8$, a doubling time of two days for active cases corresponds to an estimate of the effective reproduction number of $\mathcal{R}_t = 3.77$ and a doubling time of nine days for

active cases to an estimate of $\mathcal{R}_t = 1.6$. Note that at the start of an epidemic, the effective reproduction number of the disease $\mathcal{R}_t$ should be close to the basic reproduction number of the disease $\mathcal{R}_0$ if no mitigation steps have been taken and if the entire population is susceptible to the disease.

As equation (8) makes clear, however, to use data on the growth of infections in the initial stage of the epidemic to estimate the basic reproduction number of the disease $\mathcal{R}_0$, we must also have data on the recovery rate γ. Thus, uncertainty in our measurement of the recovery rate γ translates, in the early phase of an epidemic, into uncertainty in our estimate of the basic reproductive number $\mathcal{R}_0$.

Given the large uncertainty over estimates of γ and the wide range of case growth rates observed across different regions, there is a great deal of uncertainty over the basic reproduction number of COVID-19. Based on the observation that COVID-19 has grown just about everywhere, it is certainly greater than 1. Many estimates of the basic reproduction number of COVID-19 put it over 2 or 3 or even more. In the numerical illustrations above, I use basic reproduction numbers $\mathcal{R}_0$ ranging from 1.6 to 3, corresponding to doubling time for active infections at the start of the epidemic with recovery rate $\gamma = 1/8$ between 9.2 and 2.8 days.

An additional difficulty with measurement of the basic reproduction number of the disease from early data on the growth rate of active infections arises if diagnostic testing does not keep pace with the initial growth of the disease.[28] In this case, one may wish to use data on deaths due to the disease to measure the growth of the disease, under the presumption that the growth rate of deaths due the disease is measured more accurately than the growth rate of infections.[29]

One can measure the effective reproduction number from the growth rate of deaths as follows.[30] From equation (2), the evolution of the true number of infected agents is given by

$$I_t = I_0 \exp(\gamma \int_0^t (\mathcal{R}_s - 1) ds).$$

From equation (6), we then have that

$$\frac{dD}{dt} = \nu\gamma I_0 \exp(\gamma \int_0^t (\mathcal{R}_s - 1) ds),$$

where this time derivative of deaths is approximated in the data by the daily number of new deaths. If we differentiate this expression again, we get

$$\frac{d^2D}{dt^2} = \nu\gamma^2 I_0 \exp(\gamma \int_0^t (\mathcal{R}_s - 1) ds)(\mathcal{R}_t - 1).$$

Thus, we can get an estimate of the effective reproduction number of the disease at time t from

$$\mathcal{R}_t = 1 + \frac{1}{\gamma} \frac{\frac{d^2D}{dt^2}}{\frac{dD}{dt}}. \tag{16}$$

Note that the term

$$\frac{\frac{d^2D}{dt^2}}{\frac{dD}{dt}}$$

corresponds in a discrete time model to the time derivative of the logarithm of the daily number of new deaths.

5 Forecasting an Epidemic

In this section, I review two alternative methods for forecasting the progression of the COVID-19 epidemic. I use this discussion to highlight the identification issues that arise when one does not have adequate testing data to measure the number of actively infected or recovered individuals[31] and to outline a procedure based on work in Atkeson, Kopecky, and Zha 2020 for using the SIR model to construct forecasts of the progression of the epidemic going forward. I also relate this approach to the work of Fernandez-Villaverde and Jones 2020.[32]

The first method I review is reduced form. I base my discussion of this approach on a stylized version of the widely cited model developed by researchers at the Institute for Health Metrics and Evaluation (IHME) at the University of Washington. I discuss how one can use the SIR model to directly interpret these reduced-form forecasts in terms of forecasts of the effective reproduction number of the disease going forward. I use this discussion to illustrate why the reduced-form forecasts of this kind tend to produce a more optimistic forecast for cases and deaths going forward than the structural approaches considered next.

The second method I review makes full use of the structure of the SIR model.[33] The model is first fit to match data to date on deaths from COVID-19 in the same manner as in the reduced-form forecasting exercises. But to generate forecasts, the full structure of the SIR model is used to construct an estimate of the current distribution of the population across states S_t, I_t and R_t and of the path of the normalized transmission rate from the start of the epidemic to the current date. Scenarios for the future path of the epidemic are then computed by solving the model going forward, based on assumptions about the impact of disease mitigation efforts on the future path of the normalized transmission rate. I discuss how the main difference in forecasts generated by these two methods is the assumptions made implicitly or explicitly about the path going forward of the effective reproduction number of the disease.

5.1 Reduced-Form Approaches

The reduced-form approaches followed by the team at IMHE and Linton 2020 are based on the empirical observation that the number of daily new cases and deaths in those places impacted early on by the epidemic (such as Wuhan and regions of Italy and Spain) that then imposed mitigation measures has followed a hump-shaped pattern.[34] The forecasts from these models are predicated on the assumption that other regions experiencing the epidemic and imposing social distancing measures later on should follow a similar hump-shaped pattern of daily new cases and/or deaths.

The central idea behind these reduced-form approaches can be explained using a simple version of the IHME model's approach.[35] This model assumes that the path of cumulative

daily deaths beginning from some initial calendar date t_0 is given by

$$D_t = \frac{p}{2}\left(1 + \frac{2}{\sqrt{\pi}}\int_0^{a(t-t_0-b)} \exp(-\tau^2)d\tau\right),$$

where the parameter p denotes the limiting number of deaths D_∞, a is a growth parameter, and b is an inflection point. This specification produces an implication for daily deaths (approximated by the derivative of cumulative deaths) given by

$$\frac{dD_t}{dt} = \frac{pa}{\sqrt{\pi}} \exp(-a^2(t-t_0-b)^2),$$

which peaks at time $t = b + t_0$ and then falls thereafter. Note that the implied growth rate of the logarithm of daily deaths is given by

$$\frac{\frac{d^2D}{dt^2}}{\frac{dD}{dt}} = 2a^2(b + t_0 - t).$$

The model is fit to the available data on deaths between the initial date t_0 and the current date, then the assumed parametric form is used to project daily deaths beyond the current date. This projection is based on the claim that this parametric form fits the data for locations that are further along in disease progression. One can incorporate measures of the extent and timing of mitigation in the estimation as covariates for the parameters p and b, based on experience across locations.

Some argue that this approach produces optimistic forecasts, at least in comparison with forecasts based on structural models. To illustrate the basis for this argument, I relate this empirical specification for cumulative deaths to our SIR model as follows.

With a constant fatality rate ν, our SIR model implies that observed deaths are related to active infections by equation (6), and equation (16) implies that the path of the effective reproduction number over time implied by the parameters of this empirical specification for cumulative deaths is given by

$$\mathcal{R}_t = 1 - \frac{2a^2}{\gamma}(t - t_0 - b).$$

That is, for our SIR model to replicate this pattern of cumulative deaths, the normalized transmission rate β_t/γ would have to vary over time so as to produce the path given above for the evolution of the effective reproduction number $\mathcal{R}_t = \beta_t/\gamma S_t$ over time.

Note then that this empirical specification implies that the effective reproduction number falls linearly over time, not only in the period of estimation but also going forward beyond the current period.[36] Thus, in the context of a structural model, unless one assumes that the current value of S_t/N is substantially below one, this is equivalent to assuming that the impact of mitigation measures on disease transmission will continue to lower the transmission rate over time rather than keep it stable or even allow it to rise. It is not clear that this is a natural assumption regarding the impact of lockdowns or other disease mitigation measures — that their impact on disease transmission would grow over time. One can see then that this

forecasting procedure produces “optimistic" forecasts relative to a fully structural approach based on the assumption that the normalized transmission rate will remain at some constant value or even return to some increasing path.

5.2 A Structural Approach

I now consider model estimation and forecasting based on the full structure of the SIR model.

Conceptually, to use the full structure of our SIR model to generate forecasts for the progression of the epidemic, we must identify the current state of the population in terms of its distribution across states, S_t, I_t and R_t, and the path of the normalized transmission rate going forward β_{t+s}/γ for dates $t+s$ with $s > 0$. Once the state of the system is identified and the normalized transmission rate forecast, then the model equations imply a path going forward for the effective reproduction ratio $\mathcal{R}_{t+s}$ and the evolution of the state S_{t+s}, I_{t+s} and R_{t+s}. To add a forecast of deaths from the epidemic, one can augment the model with equation (6), which links daily deaths to the measure of active infections.

To use this approach, we must wrestle with the data problems inherent in measuring the true number of infected and resistant agents with incomplete testing of the population as discussed in Atkeson 2020a, as well as the difficulties with forecasting the path of the normalized transmission rate. I discuss those issues here. I do so in the context of a specific model estimation exercise presented next.

Assume that we have complete data on cumulative deaths D_t from dates $t = t_0 > 0$ to T, where t_0 is the calendar date on which cumulative deaths D_{t_0} are equal to some threshold (like 5 or 15 or 25) and T represents the current period. Assume that we also have data on the first and second derivatives of cumulative deaths during this time period. These derivatives are continuous time versions of data on daily deaths and the change in daily deaths. Assume that we do not have data on active infections or resistant agents.

Consider the following thought experiment based on these assumptions. Imagine that we fix parameters ν and γ governing the fatality rate and recovery rate of the disease. What state of the population at date t_0 and course of the time-varying normalized transmission rate β_t/γ from dates t_0 to T would allow the model to match exactly the data on deaths from t_0 to T?

We have the following equations to work with in estimating these additional parameters of the model. Note that we assume that these equations apply for all $t > 0$ but we have data only on the level and derivatives of cumulative deaths from $t \geq t_0$. From (6), we have

$$I_t = \frac{1}{\nu\gamma} dD_t/dt.$$

Using (3) and (6) together and the assumption that $R_0 = D_0 = 0$, we have that

$$R_t = \frac{1}{\nu} D_t.$$

Using that the states must sum to one, we have

$$S_t = 1 - \frac{1}{\nu} D_t - \frac{1}{\nu\gamma} \frac{dD_t}{dt}.$$

These equations then give the full path of the state of the population from dates t_0 to T. Note that the intuition behind these equations is straightforward. To determine the number of resistant agents at time t, we simply use the observation that the level of cumulative deaths, together with an estimate of the fatality rate of the disease, tells us how many agents are resistant to the disease. To determine the number of actively infected agents at time t, we use the observation that the level of daily deaths (the derivative of cumulative deaths), together with an estimate of the fatality and recovery rates of the disease, tells us the number of active infections. The number of susceptible agents at time t is then one minus these two quantities.

To recover the implied path of the normalized transmission rate, note that equation (16), together with the death data, gives the following estimate for the path of the product of the effective reproduction between dates t_0 and T. This estimate, together with the estimate of S_t above, implies that the path of the normalized transmission rate needed to exactly match the deaths data for $t \in [t_0, T]$ is given by

$$\frac{\beta_t}{\gamma} = \frac{1 + \frac{1}{\gamma}\frac{\frac{d^2 D_t}{dt^2}}{\frac{dD_t}{dt}}}{1 - \frac{1}{\nu}D_t - \frac{1}{\nu\gamma}dD_t/dt}.$$

This estimation procedure makes clear the identification problem in pinning down the parameters ν and γ from deaths data alone. In particular, if we start with a prior for ν and γ, we should not be able to update this prior based on deaths data alone, unless the estimates above result in inadmissable values of implied I_t, R_t, or S_t (above 1 or negative) or some implausible path for the normalized transmission rate β_t/γ (such as a negative value).[37]

Putting aside this identification problem, if we are given outside estimates of the parameters ν and γ, how might we then use the structure of the SIR model to construct a forecast different from that offered by the reduced-form approach? In joint work with Karen Kopecky and Tao Zhao in Atkeson, Kopecky, and Zha 2020, we outline the following four-step procedure for doing so.

In step 1, estimate an empirical specification for cumulative deaths between dates t_0 and the present that is twice differentiable. The empirical specification used in the previous subsection is one possible specification, but of course, there are many others.

In step 2, use the best available information to choose parameters for the fatality and recovery rates ν and γ, or a range for these parameters to be considered.

In step 3, we use the equations above to estimate the paths of S_t, I_t, R_t and the normalized transmission rate β_t/γ for which the model fits exactly the empirical specification for deaths in step 1.

These steps then give a "best estimate" of the current state of the population and the path that the normalized transmission rate has followed to date.

In step 4, construct a model-based forecast by specifying a path forward for the normalized transmission rate β_t/γ based on estimates of the impact of future mitigation on disease transmission going forward and by solving the SIR model from this current state with this assumed path of the normalized transmission rate.

With this procedure, one can estimate the structural model in as flexible a manner as desired with regard to empirical specifications of the evolution of cumulative deaths to date. One can then present a model-based forecast directly related to estimates of the progress

towards herd immunity and the impact of future disease mitigation efforts.[38]

Consider the following back-of-the-envelope application of this procedure to data on deaths for the United States as a whole from early April to early May 2020.[39]

Begin with an empirical specification for cumulative deaths in the United States between the dates t_0 corresponding to April 8, 2020 and T corresponding to May 8, 2020, given by

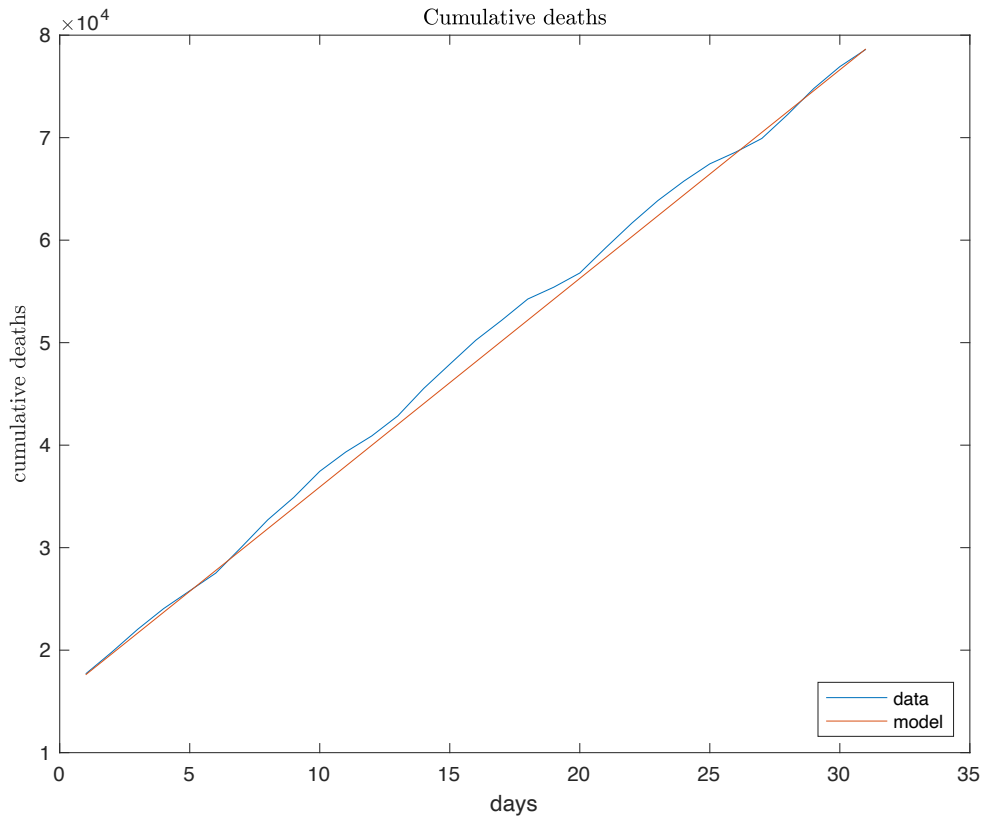
$$D_t = D_{t_0} + a(t - t_0);$$

that is, model cumulative deaths over this time period as growing linearly through time. Note that with this specification, daily deaths are constant

$$\frac{dD_t}{dt} = a,$$

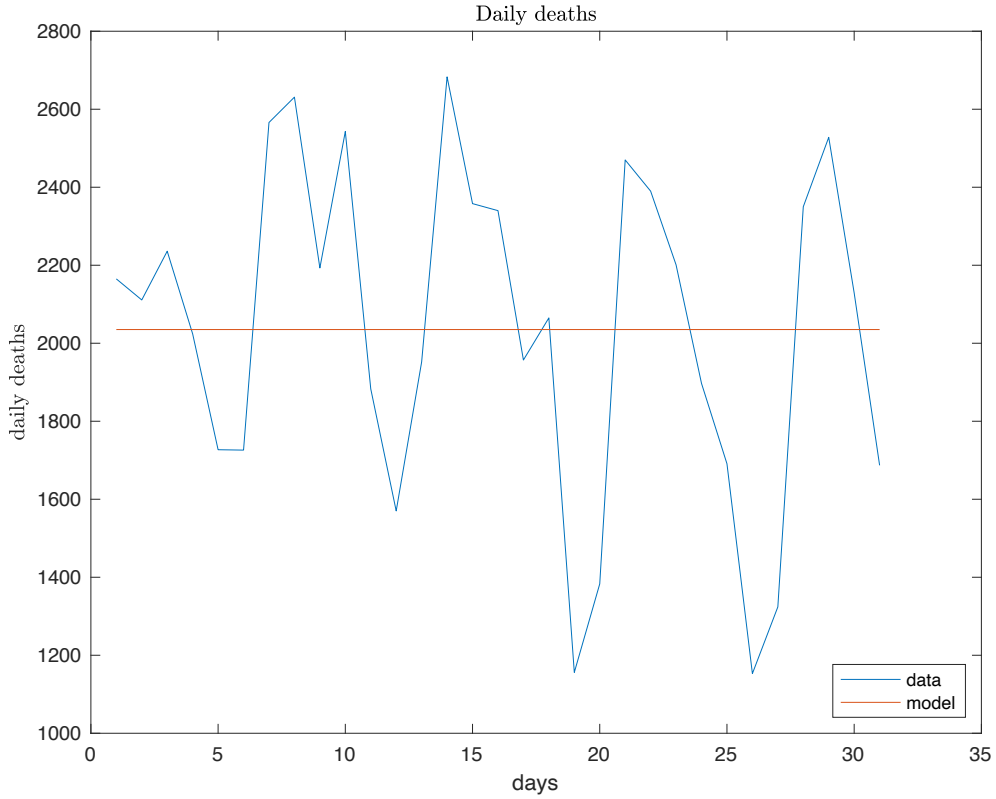
and the second derivative of cumulative deaths is zero. In figures 8 and 9, I show cumulative deaths and daily deaths in the data and this empirical specification of the model.

Figure 8
Cumulative deaths, April 8 to May 8, 2020, data and model



Note: Data on deaths at various dates used in the exercise are taken from `https://www.worldometers.info/coronavirus/country/us/`.

Figure 9
Daily deaths, April 8 to May 8, 2020, data and model



Note: Data on deaths at various dates used in the exercise are taken from
`https://www.worldometers.info/coronavirus/country/us/`.

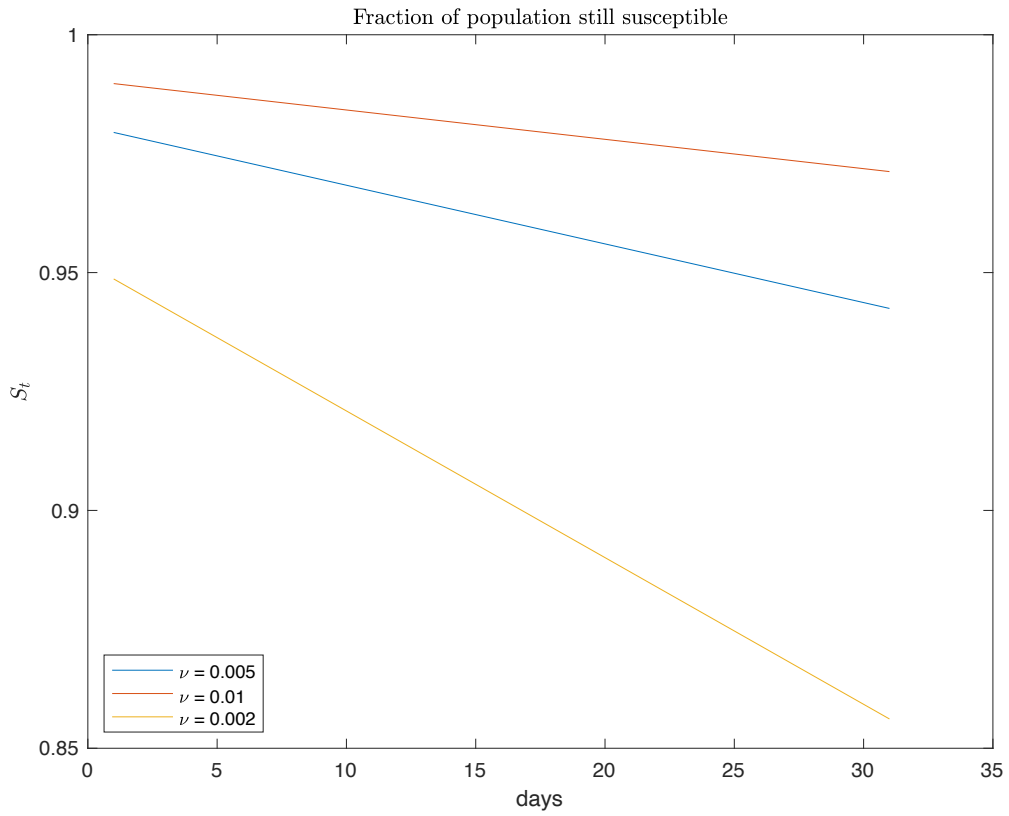
To explore the implications of these data for the evolution of the state of the population and for the normalized transmission rate, I use a recovery rate of $\gamma = 1/8$ and three values for the fatality rate: $\nu = 0.005$ as a baseline value, and $\nu = 0.01$ and $\nu = 0.002$ as alternative values.

Consider first the implications of these parameter choices for the fraction of agents currently infected. Since daily deaths are constant over time in our empirical specification, we have that the fraction of agents actively infected is also constant over time, at just under 1% of the population in our baseline fatality rate, and just under 0.5% and 2.5% of the population with our two alternative fatality rates.

In figure 10, I show the corresponding implications of the model for the fraction of agents still susceptible . We see here that different assumptions about the fatality rate lead to substantially different implications for the infection rate and the fraction of agents still susceptible.

Consider finally the implications of this specification of the model for the evolution over time of the effective reproduction number $\mathcal{R}_t$ and the normalized transmission rate

Figure 10
Model-implied fraction of agents remaining susceptible to infection, April 8 to May 8, 2020
for three different fatality rates

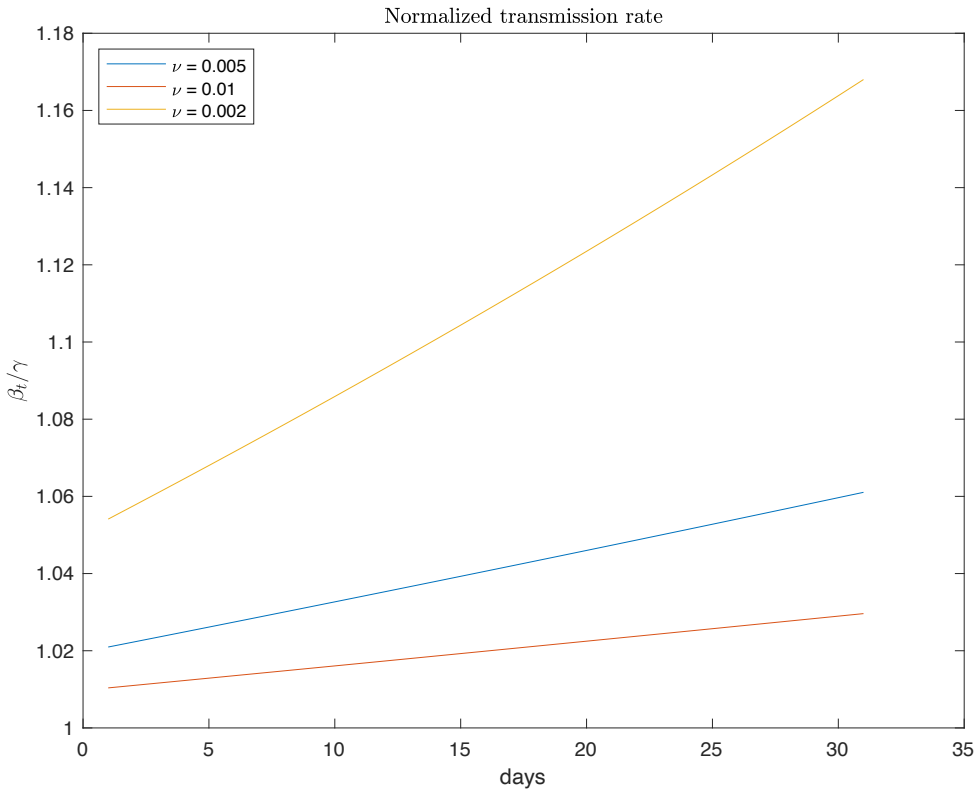


Note: The fatality rates considered are $\nu = 0.01, 0.005, 0.002$

β_t/γ. From equation (16), we have immediately that our linear empirical specification for cumulative deaths over the past month implies an effective reproduction number equal to one for that time period because the second derivative of cumulative deaths is equal to zero in this specification. Thus, in this case, our model's implications for the evolution of the normalized transmission rate is simply given by $1/S_t$, where S_t is shown in figure 10. Therefore, this specification of our model implies that the normalized transmission rate has been rising over the course of the past month. I plot this estimate of the evolution of the normalized transmission rate in Figure 11.

I have now used the model and data on deaths over the past month to compute the state of the population in terms of S_t, I_t, and R_t, the evolution over time of the effective reproduction number $\mathcal{R}_t$ and the normalized transmission rate β_t/γ. To make forecasts of the evolution of the disease, one must make a forecast of the evolution of the normalized transmission rate going forward.

Figure 11
Estimated normalized transmission rate April 8 to May 7, 2020 for three different fatality rates



Note: The fatality rates considered are $\nu = 0.01, 0.005, 0.002$

I consider two alternative scenarios.

First, one might conjecture that decentralized disease avoidance behavior would adjust endogenously to rising or falling daily deaths and/or changes in the fraction of the population currently infected leading these quantities to remain constant in equilibrium.[40] Alternatively, this outcome might be targeted by policy. For example, in Germany, the stated policy is to tune mitigation steps to keep the effective reproduction number at or below one to avoid overloading the health system.[41] In either case, the effective reproduction number going forward would remain at one, the normalized transmission rate going forward would rise over time with $1/S_t$, and the current pattern of linear growth of cumulative deaths and the current constant fractions of the population actively infected would continue.

Under first scenario, the model forecasts many months of roughly 2000 daily deaths, with an associated portion of the population actively infected at just under 1% of the population in our baseline fatality rate, and just under 0.5% and 2.5% of the population with our two alternative fatality rates. This linear growth of cumulative deaths and constant infection rate would end only when the population reaches herd immunity, so that even with no mitigation

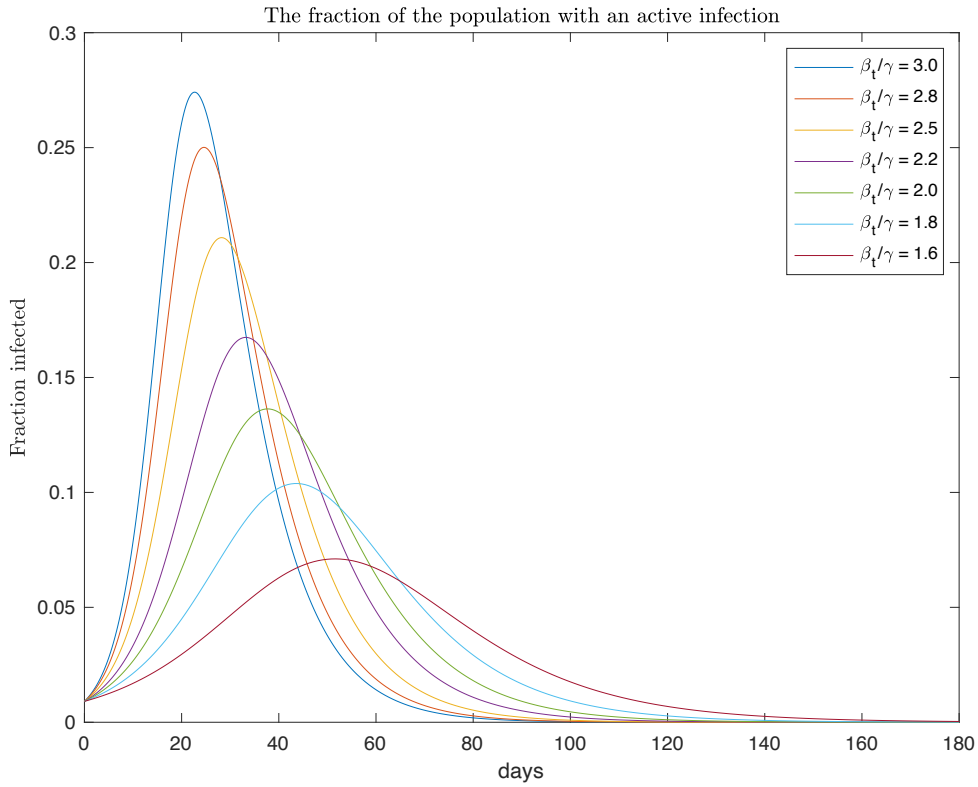
efforts at all, the effective reproduction number would begin to fall below one. With a normalized transmission rate of $\beta_t/\gamma = 2.5$, this would occur in 470 days from May 8 with our baseline fatality rate, and in 957 or 178 days from May 8 with our two alternative fatality rates.

Alternatively, one might conjecture a second scenario going forward in which a relaxation of policies regarding social distancing leads to an increase in the normalized transmission rate to a level above that shown in figure 11 for the month from April 8 to May 8, 2020. For example, one might conjecture that, going forward, the normalized transmission rate might rise to a value in the range of 1.6 to 3, as considered in figures 1 to 4.

I illustrate the forecast progression of the epidemic under such a scenario in figures 12 and 13. Specifically, given our baseline estimate of the fatality rate, our back-of-the-envelope estimate of the fraction of the population susceptible, infected, and resistant on May 8, 2020 is $S = 0.9425$, $I = 0.0448$ and $R = 0.0477$, respectively. I then use the SIR model to simulate the path of the epidemic going forward with a normalized transmission rate held constant at a value in the range of 1.6 to 3.

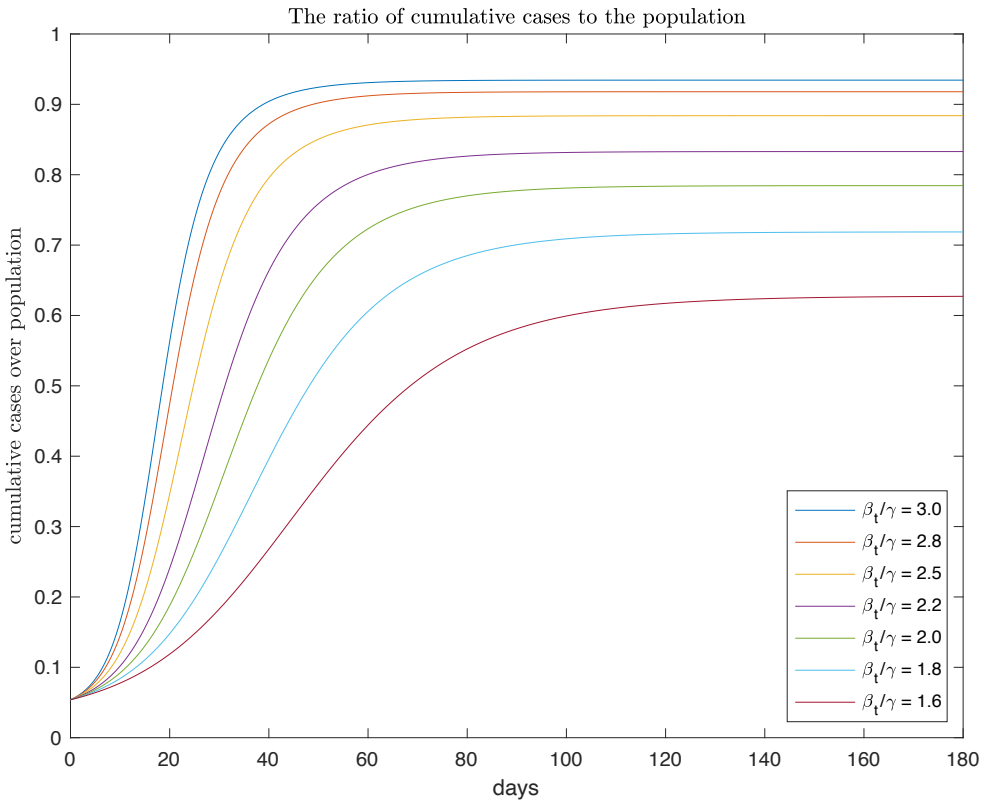
Figure 12

Fraction of the population with an active infection over 180 days going forward from May 8, 2020 under different values of β_t/γ held constant over the entire time period.



Note: The distribution of the population across states on May 8, 2020 is set to $S = 0.9425$, $I = 0.0448$ and $R = 0.0477$.

Figure 13
The cumulative burden of the disease as measured by cumulative cases as a fraction of the population over 180 days going forward from May 8, 2020 under different values of β_t/γ held constant over the entire time period.



Note: The distribution of the population across states on May 8, 2020 is set to $S = 0.9425$, $I = 0.0448$ and $R = 0.0477$

In figures 12 and 13 we see that if, as conjectured in this second scenario, a relaxation of social distancing measures were to lead the normalized transmission rate to rise to a level in this range of 1.6 to 3, then a large second wave of the epidemic would appear relatively quickly (in twenty to thirty days), with a peak prevalence and cumulative burden of the disease much higher than has been experienced to date. These forecasts under this second scenario for the normalized transmission rate of the disease going forward are driven by the simple logic of the SIR model that diseases with high normalized transmission rates (high basic reproduction numbers) and low numbers of agents initially resistant spread quickly through the population and, in the end, infect a very large fraction of that population.

6 Conclusion

Several lessons can be drawn from this review of the properties of a simple SIR model of the progression of COVID-19.

First, as shown in figures 1 and 2, a researcher who applies a simple SIR model to construct scenarios for an epidemic that is new to a population will be driven to conclude that, absent significant intervention, the peak prevalence and cumulative burden of the disease will be very high if, in the early stages of the epidemic, the data show a a rapid recovery rate for infected individuals and a rapid growth rate of active infections and daily deaths. This is because, in the SIR model, there is no inherent force to slow the rapid growth of the epidemic other than the acquisition of herd immunity by the population as a whole.

Second, as shown in figures 5 and 6, an SIR model implies that temporary mitigation measures (either mandated or undertaken in a decentralized fashion) applied early on in the progression of the epidemic do little to change the peak prevalence and cumulative burden of the disease. Such early temporary mitigation measures simply postpone the progression of the epidemic by the length of time that such temporary measures are imposed.

In contrast, as also shown in figures 5 and 6, well-timed application of temporary mitigation measures can have an impact on the long-run peak prevalence and cumulative burden of the disease. The logic of this result, as shown in figures 3, 4, and 7, depends on the impact that well-timed temporary mitigation measures might have on the fraction of agents resistant (or equivalently, actively infected) when the population reaches herd immunity. If such measures result in the population having few actively infected agents when it reaches herd immunity after these measures are lifted, then the cumulative burden of the disease is reduced as the epidemic dies out more quickly.

Third, as shown in section 5, estimated reduced-form and structural SIR models of an epidemic should result in similar implications for the progression of the epidemic to date in terms of the evolution of the fractions of the population susceptible, actively infected, and resistant and for the evolution of the effective reproduction number of the disease to date.

Where these two estimation and forecasting approaches differ is in the forecast generated regarding the evolution of the epidemic going forward. Reduced-form models, in projecting forward based on estimated empirical specifications for data on deaths or cases, make implicit assumptions about the evolution of the effective reproduction number of the disease going forward. One prominent version of this model, developed by the team at IMHE, implicitly assumes that the effective reproduction number of the disease will continue to fall over time. Structural SIR models do not impose such an implicit assumption about the evolution of the transmission of the disease going forward. Instead, explicit scenarios for the evolution of the normalized transmission rate of the disease going forward must be constructed.

I have considered two such scenarios using a back-of-the-envelope estimation of an SIR model for COVID-19 for the United States here. In one scenario, I conjecture that the effective reproduction number of the disease will continue to hover around one as it has done from early April to early May. Under this scenario, we should expect active infections and daily deaths to remain roughly constant and cumulative infections and deaths to grow linearly for many months to come. Under a second scenario, I conjecture that relaxation of social distancing policies will lead to an increase in the transmission rate of the disease back to levels more in line with those seen in the early phase of this epidemic. Under this second

scenario, we should expect to see a potentially dramatic second wave of the epidemic that then ends in a few months.

To finish up, I briefly discuss scenarios for the evolution of COVID-19 in the coming years if immunity is temporary, rather than permanent as is assumed in the standard SIR model. To account for temporary immunity, one can introduce a new parameter that governs the rate at which agents transition from the state R (resistant) back to the state S (susceptible). This modification of the model changes the dynamics of the disease substantially, allowing for multiple waves of infection in the population.[42] This modification also changes the calculus of optimal policy. See, for example, Rowthorn and Toxvaerd 2015.

The scenarios for COVID-19 based on temporary immunity outlined in Kissler et al. 2020 have the disease recurring on a seasonal basis every year, or every two years, for the next several years. The simulations shown in this article suggest that we may be dealing with COVID-19 for a long time to come.

Notes

1. See, for example, https://www.nature.com/articles/d41586-020-01003-6. See Cobey 2020 for a useful summary of issues connected to applying these models to COVID-19.

2. See this story in the *New York Times* regarding the decision in 2006 to advocate social distancing as a disease mitigation measure for influenza pandemics: https://www.nytimes.com/2020/04/22/us/politics/social-distancing-coronavirus.html. See also the dissent from that decision in Inglesby et al. 2006 based on the uncertainties regarding the effectiveness of this measure and its economic costs https://assets.documentcloud.org/documents/6841076/2006-11-Disease-Mitigation-Measures-in-the.pdf.

3. See, for example, Alvarez, Argente, and Lippi 2020, Berger, Herkenhoff, and Mongey 2020, Baskozos, Galanis, and Di Guilmi 2020, Chudik, Pesaran, and Rebucci 2020, Eichenbaum, Rebelo, and Trabant 2020, and Glover et al. 2020. See Eksin, Paarporn, and Weitz 2019 for work by epidemiologists to build and sequentially estimate a model of behavioral feedbacks in an SIR model.

4. The model was first developed in Kermack and McKendrick 1927. Here, I consider a model in which agents cannot get the disease again once they have transitioned into the R state. It is not yet clear whether this assumption is correct for COVID-19.

5. The scenarios considered here should not be considered definitive forecasts. They are intended only to allow the reader to see how a model of the progression of the epidemic might be applied to economic analysis of COVID-19 and to allow readers trained in economics to begin conversations with public health experts in this area.

6. See also Rachel 2020 for analytical formulas for the impact of disease mitigation on peak prevalence of the disease and the cumulative burden of the disease.

7. Barro 2020 uses data from the 1918-19 Spanish Flu pandemic to argue that temporary mitigation measures reduced peak disease prevalence but did not impact total mortality over the long run.

8. This finding differs from that in Atkeson 2020b, because in that previous paper, I did not consider a scenario with temporary disease mitigation that allowed for anything more than a negligible fraction of agents in the population to become resistant to the disease.

9. See, for example, Alvarez, Argente, and Lippi 2020, Rachel 2020, and Toda 2020. See Abakuks 1972 for an early application of optimal control to the study of epidemics. See also Farboodi, Jarosch, and Shimer 2020 and Kruse and Strack 2020.

10. I also relate this estimation and forecasting method to the structural estimation and forecasting model of Fernandez-Villaverde and Jones 2020.

11. See the working paper discussion of the IMHE forecasting model here: http://www.healthdata.org/sites/default/files/files/Projects/COVID/RA_COVID-forecasting-USA-EEA_

042120.pdf. See also a similar model developed at the University of Texas-Austin at https://covid-19.tacc.utexas.edu/media/filer_public/87/63/87635a46-b060-4b5b-a3a5-1b31ab8e0bc6/ut_covid-19_mortality_forecasting_model_latest.pdf.

12. See updated forecasts at http://covid.econ.cam.ac.uk/linton-uk-covid-cases-predicted-peak.

13. See https://reichlab.io/covid19-forecast-hub/ for a comparison of forecasts from a variety of epidemiological models.

14. Note that the precise nature of the interaction required to spread the disease depends its the natural transmission mechanism. Some diseases such as HIV or Ebola require direct exchange of body fluids; others are transmitted through a fecal-oral mechanism. COVID-19, like influenza, is transmitted through droplets or aerosol expelled by infected agents and taken in by susceptible agents either by getting them on their hands and touching their face or through breathing them in. A regularly updated discussion of the transmission mechanism for COVID-19 is available here: https://www.who.int/news-room/q-a-detail/q-a-coronaviruses.

15. It is unclear whether exposure to COVID-19 leads to long-lasting immunity to the disease. Rowthorn and Toxvaerd 2015 is an early study of the optimal mitigation of diseases with temporary or no immunity. Kissler et al. 2020 provide a model of the recurrence of COVID-19 in coming years under different scenarios for the length of time for which those who have had the disease remain immune.

16. Many diseases have an initial period during which an agent is infected but not contagious. This is referred to as a *latency period* and is modeled by adding a state E between S and I, as in Atkeson 2020b. We abstract from that here. Note that the latency period is distinct from the *incubation period*, which is the time between exposure to the virus and the appearance of symptoms. It appears that COVID-19 can be transmitted before symptoms appear.

17. Seasonal influenza is an example of a disease whose transmission rate regularly fluctuates with the weather. The Spanish Flu of 1918-19 came and went in three big waves in the spring and fall of 1918 and the spring of 1919. It is not fully understood what drove the changes over time in the transmission rate of that disease. The available data on COVID-19 indicate that its transmission rate widely varies across different geographies. See https://www.cidrap.umn.edu/sites/default/files/public/downloads/cidrap-covid19-viewpoint-part1_0.pdf for a careful discussion of natural fluctuation in transmission of similar diseases. See https://www.nytimes.com/2020/05/03/world/asia/coronavirus-spread-where-why.html for a discussion of geographic variation in transmission of COVID-19.

18. Note that in this simple model, we do not differentiate between the roles of infected agents with mild and severe cases in spreading the disease. Clearly, the severity of the disease may impact the spread, particularly since the very sick are likely to stay at home or go to the hospital. Research into the question of the extent to which those who are infected but asymptomatic spread the disease is ongoing. See, for example this research in *Science*: https://science.sciencemag.org/content/early/2020/03/24/science.abb3221.

19. To mark the anniversary of the 1918 Spanish Flu epidemic, in 2018, the BBC recruited a large number of volunteers to track their movements and whom they met and in what locations. These data were stratified on many dimensions and were then used to simulate the spread of a flu pandemic in the UK population. See https://www.sciencedirect.com/science/article/pii/S1755436518300306 for an epidemiological model based on that detailed data on social interactions, and see this article for an updated version of the social interaction matrix used in that model https://www.medrxiv.org/content/10.1101/2020.02.16.20023754v2.

20. The use of the notation $\mathcal{R}_0$ to denote the basic reproduction number and the letter R to denote the fraction of agents who are resistant is an unfortunate choice, but it is standard. See, for example, https://mathworld.wolfram.com/Kermack-McKendrickModel.html.

21. For the purposes of this paper, I assume that this death rate is constant and thus independent of the stress placed on the health care system at points of peak infection. That assumption is clearly

incorrect. Evidence from Wuhan, Italy, and New York City suggests that the fatality rate from COVID-19 is much higher in periods of peak infection.

22. Note that $\mathcal{R}_0$ is a ratio of two rates per unit time and hence is not denominated in units of time — a slow moving disease such as HIV and a fast moving disease such as COVID-19 could both have the same normalized transmission rate and hence the same peak prevalence of the disease and cumulative disease burden.

23. See https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html for an estimate by the CDC of the number of cases of COVID-19 that are travel related in the United States. There is considerable uncertainty about this number. See https://www.nytimes.com/2020/04/04/us/coronavirus-china-travel-restrictions.html for data on the large volume of travel from China to the United States in the early part of 2020 and https://www.nytimes.com/2020/05/07/us/new-york-city-coronavirus-outbreak.html for information about how travel from New York to other parts of the country may have seeded outbreaks across the United States.

24. See, for example, https://www.nature.com/articles/s41591-020-0869-5.

25. See, for example, https://wwwnc.cdc.gov/eid/article/26/7/20-0282_article.

26. See, for example, this pre-print in *Nature*: https://www.nature.com/articles/s41586-020-2196-x.

27. See Wallinga and Teunis 2004 and Chowell, Nishiura, and Bettencourt 2007 for discussions of how to estimate the effective reproduction number from case data.

28. In the United States, during April, the number of diagnostic tests being conducted every day is growing, at best, at a linear rather than exponential rate. Time series data on tests performed are available here: https://covidtracking.com/data/us-daily.

29. See https://www.economist.com/graphic-detail/2020/04/16/tracking-covid-19-excess-deaths-across-countries for a discussion of the extent to which data on deaths due to COVID-19 are accurately measured. Also see https://www.cdc.gov/nchs/nvss/vsrr/covid19/tech_notes.htm for provisional estimates of excess mortality in the United States.

30. Thanks to James Stock for pointing out this calculation.

31. This discussion substantially extends Atkeson 2020a.

32. I abstract from their additional state to keep the model in the SIR framework. See updates of their results at https://web.stanford.edu/~chadj/Covid/Dashboard.html.

33. There are a large number of more complex forecasting models based on a structural approach. See, for example the description of the Columbia University Mailman School of Public Health Model at https://www.medrxiv.org/content/10.1101/2020.03.21.20040303v2, or the model used by COVIDActNow described at https://data.covidactnow.org/Covid_Act_Now_Model_References_and_Assumptions.pdf.

34. See https://www.endcoronavirus.org/countries for plots of the pattern of active case counts across many countries.

35. The specific procedures implemented in the IMHE model are described here: https://www.medrxiv.org/content/medrxiv/suppl/2020/04/25/2020.04.21.20074732.DC1/2020.04.21.20074732-2.pdf.

36. Clearly, at some point, this implied effective reproduction number becomes negative, which is inadmissable. This observation implies that this functional form for cumulative deaths cannot be reproduced by our SIR model with any set of parameters and time varying transmission rate β_t.

37. In their updated paper Fernandez-Villaverde and Jones 2020 use a procedure for estimating their model closely related to the one described here. In earlier versions, they imposed the assumption that the transmission rate of the disease declines over time because of mitigation from an initial high level to a lower level during the estimation period at a rate determined by a parameter λ according to

$$\beta_t = \exp(-\lambda t)\beta_0 + (1 - \exp(-\lambda t))\beta^*.$$

The estimation procedure outlined above does not add this restriction to the estimating model.

38. See Gupta et al. 2020 and Chudik, Pesaran, and Rebucci 2020 for useful studies on which forecasts of future transmission of the disease might be based.

39. Data on the level of deaths at various dates used in the exercise are taken from https://www.worldometers.info/coronavirus/country/us/.

40. See, for example, Eksin, Paarporn, and Weitz 2019 and https://johnhcochrane.blogspot.com/2020/05/an-sir-model-with-behavior.html.

41. See https://www.wsj.com/articles/germanys-r0-coronavirus-experiment-11588115565 and an explanation of this policy by Chancellor Merkel: https://youtu.be/22SQVZ4CeXA.

42. See, for example, this discussion of the patterns of infections for COVID-19 that might be based on analogies to influenza pandemics: https://www.cidrap.umn.edu/sites/default/files/public/downloads/cidrap-covid19-viewpoint-part1_0.pdf.

References

Abakuks, Andris. 1972. "An optimal isolation policy for an epidemic." *Journal of Applied Probability* 10, no. 2 (June): 247–262.

Alvarez, Fernando, David Argente, and Francesco Lippi. 2020. "A Simple Planning Problem for COVID-19 Lockdown." Working Paper 26981, NBER, April.

Atkeson, Andrew. 2020a. "How Deadly is COVID-19? Understanding the Difficulties with Estimation of its Fatality Rate." Staff Report 598, Federal Reserve Bank of Minneapolis, March.

———. 2020b. "What Will be the Economic Impact of COVID-19 in the U.S.? Rough Estimates of Disease Scenarios." Staff Report 595, Federal Reserve Bank of Minneapolis.

Atkeson, Andrew, Karen Kopecky, and Tao Zha. 2020. "Estimating and Forecasting Disease Scenarios for COVID-19 with an SIR Model." June.

Barro, Robert. 2020. "Non-Pharmaceutical Interventions and Mortality in U.S. Cities during the Great Influenza Pandemic, 1918-1919." Working Paper 27049, NBER, April.

Baskozos, Giorgos, Giorgos Galanis, and Corrado Di Guilmi. 2020. "Social Distancing and Contagion in a discrete choice model of COVID-19." Working Paper. Australian National University, April.

Berger, David, Kyle Herkenhoff, and Simon Mongey. 2020. "An SEIR Infectious Disease Model with Testing and Conditional Quarantine." Staff Report 597, Federal Reserve Bank of Minneapolis, April.

Chowell, Gerado, Hiroshi Nishiura, and Luis M.A. Bettencourt. 2007. "Comparative estimation of the reproduction number for pandemic influenza from daily case notification data." *Journal of the Royal Society Interface* 4, no. 2 (February): 155–166.

Chudik, Alexander, M. Hashem Pesaran, and Alessandro Rebucci. 2020. "Voluntary and Mandatory Social Distancing: Evidence on COVID-19 Exposure Rates from Chinese Provinces and Selected Countries." Working Paper 27039, NBER, April.

Cobey, Sarah. 2020. "Modeling infectious disease dynamics." *Science,* no. 10.1126/science.abb5659 (May).

Eichenbaum, Martin S., Sergio Rebelo, and Mathias Trabant. 2020. "The Macroeconomics of Epidemics." Working Paper 26882, NBER, April.

Eksin, Ceyhun, Keith Paarporn, and Joshua S. Weitz. 2019. "Systematic biases in disease forecasting – The role of behavior change." *Epidemics* 27 (June): 96–105.

Farboodi, Maryam, Gregor Jarosch, and Robert Shimer. 2020. "Internal and External Effects of Social Distancing in a Pandemic." Working Paper 27059, NBER, April.

Fernandez-Villaverde, Jesus, and Charles I. Jones. 2020. "Estimating and Simulating a SIRD Model of COVID-19 for Many Countries, States, and Cities." Working Paper 27128, NBER, April.

Glover, Andrew, Jonathan Heathcote, Dirk Krueger, and Jose-Victor Rios-Rull. 2020. "Health versus Wealth: On the Distributional Effects of Controlling a Pandemic." Staff Report 600, Federal Reserve Bank of Minneapolis, April.

Gupta, Sumedha, Thuy D Nguyen, Felipe Lozano Rojas, Shyam Raman, Byungkyu Lee, Ana Bento, Kosali I Simon, and Coady Wing. 2020. "Tracking Public and Private Response to the COVID-19 Epidemic: Evidence from State and Local Government Actions." Working Paper 27027, NBER, April.

Harko, Tiberiu, Francisco 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, no. 1 (June): 184–194.

Inglesby, Thomas V., Jennifer B. Nuzzo, Tara O'Toole, and D.A. Henderson. 2006. "Disease Mitigation Measures in the Control of Pandemic Influenza." *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science* 4 (4): 366–375.

Kermack, William Ogilvy, and A.G. McKendrick. 1927. "A contribution to the mathematical theory of epidemics." *Proceedings of the Royal Society London A* 115, no. 772 (August): 700–721.

Kissler, Stephen M., Christine Tedijanto, Edward Goldstein, Yonatan H. Grad, and Marc Lipsitch. 2020. "Projecting the transmission dynamics of SARS-COV-2 through the postpandemic period." *Science* 10.1126/science.abb5793 (April).

Kruse, Thomas, and Philipp Strack. 2020. "Optimal Control of an Epidemic through Social Distancing." Working Paper 3581295, SSRN, April.

Linton, Oliver. 2020. "When will the COVID-19 Pandemic Peak?" Technical report, Cambridge-INET Working Paper Series, April.

Lourenco, Jose, Robert Paton, Mahan Ghafari, Moritz Kraemer, Craig Thompson, Peter Simmonds, Paul Klenerman, and Sunetra Gupta. 2020. "Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-COV-2 epidemic." *Cold Spring Harbor Laboratory Press* (March).

McKibbin, Warwick, and Fernando Roshen. 2020. "The Global Macroeconomic Impacts of COVID-19: Seven Scenarios." Report. `https://www.brookings.edu/research/the-global-macroeconomic-impacts-of-covid-19-seven-scenarios/`. Brookings Institution, March.

Rachel, Lucasz. 2020. "An Analytical Model of COVID-19 Lockdowns." May.

Rowthorn, Robert, and Flavio Toxvaerd. 2015. "The Optimal Control of Infectious Diseases via Prevention and Treatment." Technical report, Cambridge-INET Working Paper Series.

Stock, James H. 2020. "Data Gaps and the Policy Response to the Novel Coronavirus." Working Paper 26902, NBER, March.

Toda, Alexis Akira. 2020. "Susceptible-Infected-Recovered (SIR) Dynamics of COVID-19 and Economic Impact." Technical report. Cornell University, March.

Wallinga, Jacco, and Peter Teunis. 2004. "Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures." *American Journal of Epidemiology* 160, no. 6 (September): 509–516.