Estimating and Simulating a SIRD Model
of COVID-19 for Many Countries, States, and Cities

Jesús Fernández-Villaverde   Charles I. Jones[*]

UPenn and NBER   Stanford GSB and NBER

May 2, 2020 — Version 1.0

Abstract

We use data on deaths in New York City, various U.S. states, and various coun-
tries around the world to estimate a standard epidemiological model of COVID-19.
We allow for a time-varying contact rate in order to capture behavioral and policy-
induced changes associated with social distancing. We simulate the model forward
to consider possible futures for various countries, states, and cities, including the
potential impact of herd immunity on re-opening. Our current baseline mortality
rate (IFR) is assumed to be 0.8% but we recognize there is substantial uncertainty
about this number. Our model fits the death data equally well with alternative mor-
tality rates of 0.3% or 1.0%, so this parameter is unidentified in our data. However,
its value matters enormously for the extent to which various places can relax social
distancing without spurring a resurgence of deaths.

1. Introduction

We use data on deaths in New York City, various U.S. states, and various countries around the world to estimate a standard epidemiological model of COVID-19. Relative to existing frameworks, our contributions are

- We do not use data on cases or tests because of differential selection in testing in different cities, states, and countries. Instead we only use data on deaths.

- We estimate a time-varying R_0 to capture changes in behavior and policy that occur at different times and with different intensities in different locations.

- We show how simulating our model after a location has reached a peak in the number of daily deaths results in very stable results going forward in time. In contrast, simulations of the future *before* a location reaches its peak are extremely noisy and sensitive to daily shocks.

- Our models allow us to back out the percent of people ever-infected versus still susceptible and therefore estimate the extent to which herd immunity effects are large. We find large effects in New York City, noticeable effects in Italy and Spain, and negligible effects in New York state outside of New York City and in places like California and Sweden.

We are not epidemiologists, so these results should be interpreted with caution and care. We study a standard model of COVID-19 using common tools in econometrics and then analyze its main quantitative implications in ways that resemble how economists study other dynamic models. Our exercise can help us understand where a simple SIRD model has difficulties fitting observed patterns in the data and points out avenues for improvement while maintaining the virtues of simplicity and parsimony. We plan to update these results regularly on our web page, where we also report additional findings for many other countries and states.

2. Literature Review

Much of the mathematical study of the spread of infectious diseases start from the classic compartmental models of Kermack and McKendrick (1927) and Kermack and

McKendrick (1932). These models divide the population into several different compartments (e.g., susceptible, infective, recovered, deceased, ...) and specify how agents move across the separate compartments over time.[1] The SIRD epidemic model that we analyze in this paper is one of the simplest of these compartmental models. Hethcote (2000) presents a useful overview of this class of models and some of their theoretical properties, Brauer et al. (2019) is a recent, comprehensive textbook of the field, and Imperial College COVID-19 Response Team (2020) is an example of how these models have been applied to understand the current health crisis by epidemiologists.

The acute economic impact of the current pandemic has generated much interest among economists in exploring compartmental models and how they can be dovetailed into standard economic models and estimated using econometric techniques. (see Stock, 2020, and Avery et al., 2020, for two general surveys of how economists have addressed this topic).

First, economists have argued that many of the parameters controlling the move among compartments are not structural in the sense of Hurwicz (1962), but depend, instead, on individual decisions and policies. For example, the rate of contact that determines the number of new infections is a function of the endogenous labor supply and consumption choices of individuals. Hence, the rate of contact is amenable to be studied with standard decision theory models. See, for instance, Eichenbaum et al. (2020) and Farboodi, Jarosch and Shimer (2020). Also, the recovery and death rates are not just clinical parameters, but can be functions of policy decisions such as expanded emergency hospital capacity or priorities regarding the allocation of scarce medical resources. Similarly, the case fatality ratio, a key figure to assess the severity of the epidemic, is a complex function of clinical factors (e.g., the severity of a virus) and demographic and selection-into-disease mechanism, which are themselves partly the product of endogenous choices (Korolev, 2020b).[2] Our paper builds on these ideas by allowing the infection rates to be influenced by social distancing and by letting many parameters to vary across countries, states, and cities, which can proxies for demographic and policy heterogeneity.

---

[1]Before the contributions of Kermack and McKendrick, William Farr (1807-1883) had observed that epidemics tend to follow a Gaussian curve; however, he never presented a formal mechanism to account for such a pattern.

[2]More precisely, the case fatality ratio is not the average treatment effect on the treated (ATET), a more explicitly "causal" concept.

Second, economists have been concerned with the identification problems of compartmental models. Many of these models are unidentified or weakly identified, with many set of parameters that fit the observed data so far equally well but have considerably different long-run consequences. Atkeson (2020) and Korolev (2020a) document this argument more carefully. Our findings corroborate this result and highlight the need to develop alternative econometric approaches.

Third, some researchers have dropped the use of compartmental models completely. Instead, they have relied on time-series models from the econometric tradition. See, for instance, Linton (2020) and Kucinskas (2020).

Let us close this section by pointing out that economists are pushing the study of compartmental models in a multitude of dimensions. Acemoglu et al. (2020) and Álvarez et al. (2020) characterize the optimal lockdown policy for a planner who wants to control the fatalities of a pandemic while minimizing the output costs of the lockdown. Berger et al. (2020) analyze the role of testing and case-dependent quarantines. Bethune and Korinek (2020) estimate the infection externalities associated with COVID-19 and find them to be large. Bodenstein et al. (2020) combine a compartmental model with a multisector dynamic general equilibrium model to capture key characteristics of the U.S. Input-Output Tables. Garriga, Manuelli and Sanghi (2020), Hornstein (2020), and Karin et al. (2020) study a variety of containment policies. Toda (2020), besides also estimating a SIRD model as we do, explores the optimal mitigation policy that controls the timing and intensity of social distancing. More papers are appearing every day.

3. A SIRD Model with Social Distancing

We follow standard notation in the literature. There is a constant population of N people, each of whom may be in one of five states:

$$S_t + I_t + R_t + D_t + C_t = N$$

The states — in temporal order — are

$$
\begin{aligned}
S_t &= \text{Susceptible} \\
I_t &= \text{Infectious} \\
R_t &= \text{Resolving} \\
D_t &= \text{Dead} \\
C_t &= \text{ReCovered}
\end{aligned}
$$

A susceptible person contracts the disease by coming into "adequate" contact with an infectious person, assumed to occur at rate $\beta_t I_t / N$, where β_t is a time-varying contact rate parameter. It varies over time to capture behavioral changes such as social distancing. In particular, we assume that

$$
\beta_t = \beta_0 e^{-\lambda t} + \beta^*(1 - e^{-\lambda t})
$$

That is, the β_t process is characterized by three key parameters. There is a starting value β_0 and an ending value β^*, and β_t decays exponentially from β_0 to β^* at rate λ. We think of β_0 as capturing characteristics of the disease, fixed characteristics of the region such as density, and basic customs in the region. Then, β^* is a key policy and behavioral parameter: how strong is the social distancing that different locations adopt, either because of policy or simply because of voluntary changes in individual behavior. Finally, λ governs the speed at which this adjustment occurs.[3]

The total number of new infections at a point in time is $\beta_t I_t / N \cdot S_t$. Infectiousness resolves at Poisson rate γ, so the average number of days a person is infectious is $1/\gamma$: e.g. $\gamma = .2 \Rightarrow 5$ days.

After the infectious period is over, a person is in the "Resolving" state, R. A constant fraction, θ, of people exit this state each period, and the case is resolved in one of two ways:

$$
\begin{aligned}
\text{Death:} &\quad \text{fraction } \delta \\
\text{Recovery:} &\quad \text{fraction } 1 - \delta
\end{aligned}
$$

In our preliminary analysis, we found that it is important to have a model that distinguishes between the infectious and the recovering periods. This distinction is key to match the data with biologically plausible parameter values. In particular, it appears that the infectious period lasts on average about 4 to 5 days while cases take a total of

---

[3]We tried alternative specifications for the evolution of β_t, including discrete jumps at the time of the introduction of shelter-in-place orders. We found those alternatives fit the data worse, perhaps reflecting our view that λ also encodes individual decisions that are likely to evolve more smoothly over time.

about 2 to 3 weeks or even longer to resolve (Bar-On et al., 2020). If one assumes people are infectious for this entire period, the model has trouble fitting the data, as we discuss further below.

The laws of motion related to the virus are then given by

$$\Delta S_{t+1} = \underbrace{-\beta S_t I_t/N}_{\text{new infections}}$$

$$\Delta I_{t+1} = \underbrace{\beta S_t I_t/N}_{\text{new infections}} - \underbrace{\gamma I_t}_{\text{resolving infectious}}$$

$$\Delta R_{t+1} = \underbrace{\gamma I_t}_{\text{resolving infectious}} - \underbrace{\theta R_t}_{\text{cases that resolve}}$$

$$\Delta D_{t+1} = \underbrace{\delta \theta R_t}_{\text{die}}$$

$$\Delta C_{t+1} = \underbrace{(1-\delta)\theta R_t}_{\text{reCovered}}$$

We assume the initial stocks of deaths are set equal to zero. The initial stock of infections and resolving cases, $I(0)$ and $R(0)$, are parameters that we will estimate.

3.1 Basic Properties of a Standard SIRD Model

Here we review the basic properties of this model when $\beta_t = \beta$ and the difference equations are replaced by differential equations (Hethcote, 2000). A convention in epidemiological modeling is to recycle notation and let R_0 denote the expected number of infections generated by the first ill person when $s_0 \equiv S_0/N \approx 1$:

$$
\begin{array}{ccccc}
R_0 & = & \beta & \times & 1/\gamma \\
\text{\# of infections} & & \text{\# of lengthy} & & \text{\# of days} \\
\text{from one sick} & & \text{contacts per} & & \text{contacts are} \\
\text{person} & & \text{day} & & \text{infectious}
\end{array}
$$

More generally, if $R_0 s_0 > 1$, the disease spreads, otherwise it declines quickly. One can see from this simple equation why $R_0 > 1$ is so natural: if people are infectious for 5 days and have lengthy contacts with even just two new people per day, for example, then $R_0 = 10$.

As shown by Hethcote (2000), the initial exponential growth rate of infections is $\beta - \gamma = \gamma(R_0 - 1)$. Another useful result from these models concerns the long-run number of people who ever get infected (and therefore the fraction δ of these gives the long-run death rate). As $t \to \infty$, the total fraction of people ever infected, e^*, solves (assuming $s_0 \approx 1$)

$$e^* = -\frac{1}{R_0}\log(1 - e^*)$$

In other words, with a constant β, the long run number of people ever infected is pinned down by R_0; the parameters γ and θ only affect the timing, holding R_0 constant. The long-run death rate is then δe^*, so this too depends only on R_0 (and δ).

This explains why modeling the changing β associated with social distancing is so important. With a constant β, the initial explosion rate of the disease implies a value for β and then all the variables in the differential system are determined at that point. Instead, a changing β permits the initial exponential growth rate of deaths to be different from the long-run properties of the system, which is the point of social distancing.

4. Estimation: Countries and States

The following parameters are assumed to be primarily biological and therefore fixed over time and the same in all countries and regions:

- $\gamma = 0.2$: In the continous time version of this model, the average length of time a person is infectious is $1/\gamma$, so 5 days in our baseline, consistent with the evidence in Bar-On, Flamholz, Phillips and Milo (2020). We also consider $\gamma = 0.1$ (10 day duration). The $\gamma = 0.2$ fits slightly better, as shown below, but is not particularly well identified.[4]

- $\theta = 0.1$: In the continous time version of this model, the average length of time it takes for a case to resolve, after the infectious period ends, is $1/\theta$. With $\theta = .1$, this period averages 10 days. Combined with the 5-day infectious period, this implies

---

[4]Note that γ also incorporates choices of individuals and that, therefore, it is not merely pinned down by clinical observations. If an individual experiences symptoms or is suspicious that she might be infectious, hence withdrawing herself from effective contacts with susceptible individuals, we can consider her case has resolved for the purposes of the dynamics of the model, even if she is still under a clinical condition.

that the average case takes a total of 15 days to resolve. The implied exponential distribution includes a long tail that can be thought of as capturing the fact that some cases take even longer to resolve.

- $\delta = 0.8\%$: This parameter is crucial and it would be great to have a good estimate of it. Case fatality rates are not helpful, as we do not have a good measure of how many people are infected. Random testing for antibodies to detect how many people have ever been infected would be quite informative about this parameter. In the previous draft of this paper, our benchmark rate was 0.3% based on the evidence (below) from California and Heinsberg. However, the new evidence from New York suggests a higher death rate of 0.8%. We see this as the current best estimate. However, we will also report results for $\delta = 0.3\%$ as we await better estimates.[5]

  1. On April 23, Governor Andrew Cuomo announced preliminary results suggesting that 21% of New York City residents randomly tested from supermarket and big-box stores had antibodies for COVID-19. According to New York Department of Health (2020), it takes 3-4 weeks for these antibodies to form, so this suggests that around April 1, 21% of NYC residents were "ever infected." As discussed further below, our estimation of NYC requires a death rate of around 0.8% to match this evidence. Moreover, this death rate is consistent with back-of-the-envelope calculations of death rates using the number of deaths and the 21% infection rate.

  2. Sutton, Fuchs, D'Alton and Goffman (2020) in a NEJM study of 215 mothers giving birth in NYC found a remarkably high 15 percent currently-infected rate from March 22 to April 4. This is broadly consistent with the 21% ever-infected rate from the Cuomo report above, though perhaps it is a little higher than one might have expected.

  3. Two other recent studies of Santa Clara County and Los Angeles, two of the hardest hit places in California, suggest that around 2.5% to 5% of the populations there had been infected as of early April (Bendavid et al., 2020). Given the low numbers of deaths in these locations, these infection rates are also

---

[5]Hortasu, Liu and Schwieg (2020) estimate a IFR of 0.3% using data on travelers from Iceland.

consistent with a death rate of around 0.3%. Unfortunately, because the infection rate is so low in California, if the error rate of the test is around 3%, then this evidence could be consistent with almost any death rate, so it is not decisive.

4. A recent report of random testing in Heinsberg, Germany implied a mortality rate of 0.37%, but this study seems to have been heavily criticized.

We explore robustness to different values of these parameters.

The following parameters are assumed to vary across countries/states and are estimated to minimize a sum-of-squared residuals criterion:

- β_0 and β^*: The contact rates, influenced by behavioral changes and social distancing. Because $R_0 \equiv \beta/\gamma$, these parameters are mapped into R_0 and R_0^* in the numbers we report below. Countries, regions, and cities differ in density, culture, and climate, and each of these factors can contribute to differences in R_0. Think about how different is life in New York City versus Denver versus rural Oregon. Changes in behavior over time and changes in policies toward social distancing imply differences in R_0^* as well.

- λ: The rate at which a country or region transitions from R_0 to R_0^*. Some countries adopted social distancing policies sooner than others, and differences in λ can help capture this fact.

- I_0: the initial number of infections. This parameter is crucial in getting the timing right.

- $R(0)$: the initial number of recovering cases. We set $R(0) = \eta I_0$ and estimate η subject to the constraint that $0 \leq \eta \leq 1$, so that the initial number of recovering cases is less than the initial number of infectious cases.

The objective function we use in the estimation is the fit of cumulative deaths (in logs) and daily deaths (in logs). We use the sum of squared residuals for each series, added together with equal weights.

Data. Our data are taken from the GitHub repository of Johns Hopkins University CSSE (2020), which reports cumulative death numbers daily for countries, states, coun-

ties, and provinces throughout the world. We manipulate this data in three important ways before feeding it into the model. First, on April 15, New York City added more than 3,500 deaths to their counts, increasing the total by more than 43%. We therefore apply this same factor of proportionality (1.4325) to the deaths before April 15 to get a consistent time series for New York City. Second, The Economist (2020) reports that similar adjustments need to be made in other countries. In particular, vital statistics records in countries including Spain, Italy, England, France, and Sweden suggest that "excess deaths" relative to an average over past years exceed deaths officially attributed to COVID-19 by a large margin. We therefore increase deaths in all non-New York City locations by 33% for all dates.[6] Finally, there are pronounced "weekend effects" in the raw data: there are days, often on the weekend, in the middle of the pandemic when a country reports zero deaths, only to make up for this with a spike in deaths in subsequent days. We initially ran the model with this raw data, and it works fine. However, applying a 5-day centered moving average to the data produces more stable results, so we make this final adjustment.

Guide to Graphs. We present our results in several graphs for each country and region. These include cumulative deaths through the latest date, daily deaths (data and simulating forward), and cumulative deaths simulating forward. Data are shown as blue circles or blue bars and simulations are solid lines.

Each graph may have several lines, typically for one of two reasons. In some graphs, we show the simulations from the past seven days. This provides an intuitive assessment of how sensitive the simulations are to one or two recent observations. In other graphs, we show alternatives for baseline, "high," and "low" values of certain parameters.

In these graphs with multiple lines, we use a "rainbow" color scheme. This is easiest to illustrate in the case in which we report simulations from the past seven days. The colors of the rainbow are ordered according to ROYGBIV: red, orange, yellow, green, blue, indigo, violet. We follow this order from oldest to most recent. The only exceptions we make are that "blue" and "indigo" are collapsed into a single color, and the

---

[6]Katz and Sanger-Katz (2020) suggest that the excess deaths in New York City could be even larger than the already-adjusted numbers revealed so far: they report 20,900 excess deaths by April 26, compared to 16,673 in the official counts.

Figure 1: New York City: Cumulative Deaths per Million People ($\delta = 0.8\%/0.3\%/1.0\%$)
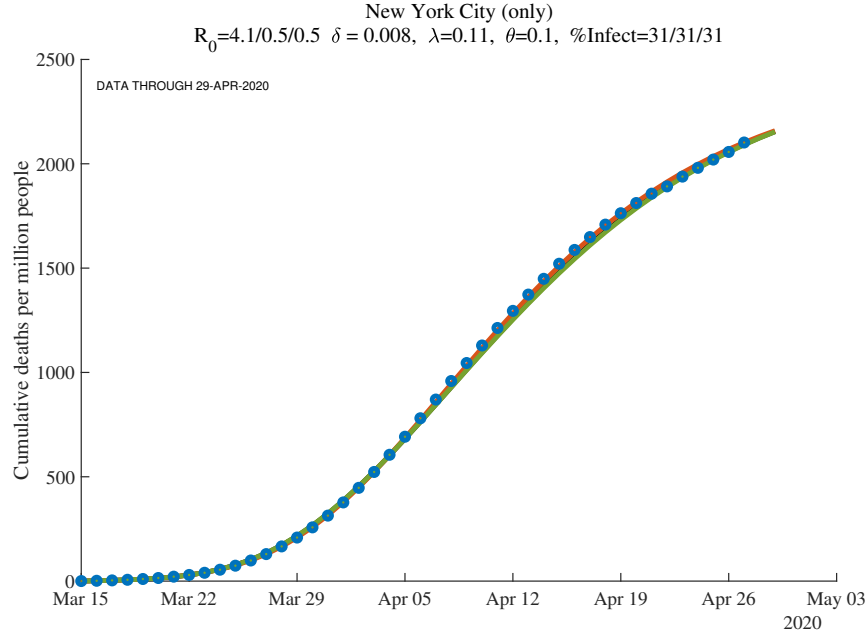
most recent simulation (or baseline value) is always shown in black.

In graphs with three lines, the black line is the baseline value, the red line is a "low" value of the parameter, and the green line is a "high" value, also respecting this rainbow order.

4.1   Baseline

Figures 1, 2, and 3 show results for New York City. The black line is the baseline case, with parameter values and estimates reported in the subtitle to the figure. This figure shows results for three values of δ — .008, .003, and .01.

For the baseline value of $\delta = 0.8\%$, the estimates suggest that New York began with $R_0 = 4.1$, so that each infected person passed the disease to nearly five others at the start. This estimate agrees with other recent findings and its particularly plausible for such a high-density metropolitan area as New York City.[7] Social distancing is estimated to have reduced this value to 0.5 by April 29, with a speed of adjustment λ around 11 percent per day. Either higher values of δ such as 1% or lower values such as 0.3% can

---

[7]For instance, Sanche, Lin, Xu, Romero-Severson, Hengartner and Ke (2020) estimate an even higher median R_0 value of 5.7 during the start of the epidemic in Wuhan.

Figure 2: New York City: Daily Deaths per Million People ($\delta = 0.8\%/0.3\%/1.0\%$)
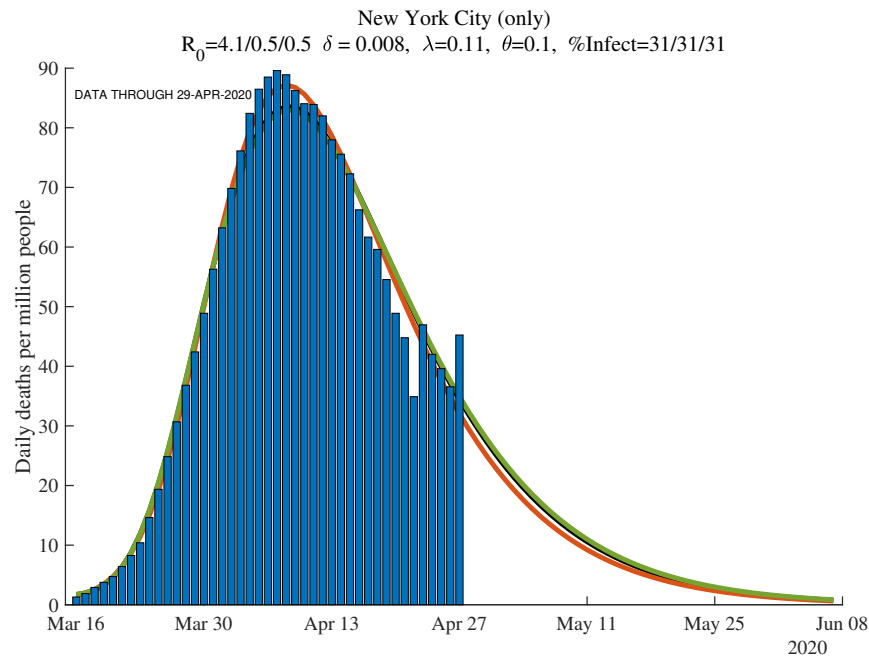
also fit this data — the three lines (black, red, and green) are on top of each other in Figure 1 — so this parameter is not identified.

Next, notice that with these estimates, 31 percent of New York City is estimated to have ever been infected by April 29. With $\delta = 0.8\%$, our model implies this number for April 1 was 20%. This compares very well to the recent preliminary announcement by Governor Cuomo that as of April 20, about 21% of New York City residents tested positive for antibodies of COVID-19. Because antibodies only appear 3 to 4 weeks after infection, these antibody tests really tell us what the ever-infected rate was 3 to 4 weeks earlier (New York Department of Health, 2020).

Figure 2 shows the daily number of deaths (per million population) in New York, both in the data and for these same three parameter values. Here it is apparent once again that all three death rates can fit the data equally well.

Figure 3 then shows the cumulative deaths per million, running the simulation forward in time. One might have thought that the different death rates would imply very different futures. However, we re-estimate all parameters when we impose the different death rates, and they all end up producing similar futures in this case. In partiuclar, the simulations imply a death rate by the middle of May of around 2500 people per million.

Figure 3: New York City: Cumulative Deaths per Million (Future, $\delta = 0.8\%/0.3\%/1.0\%$)
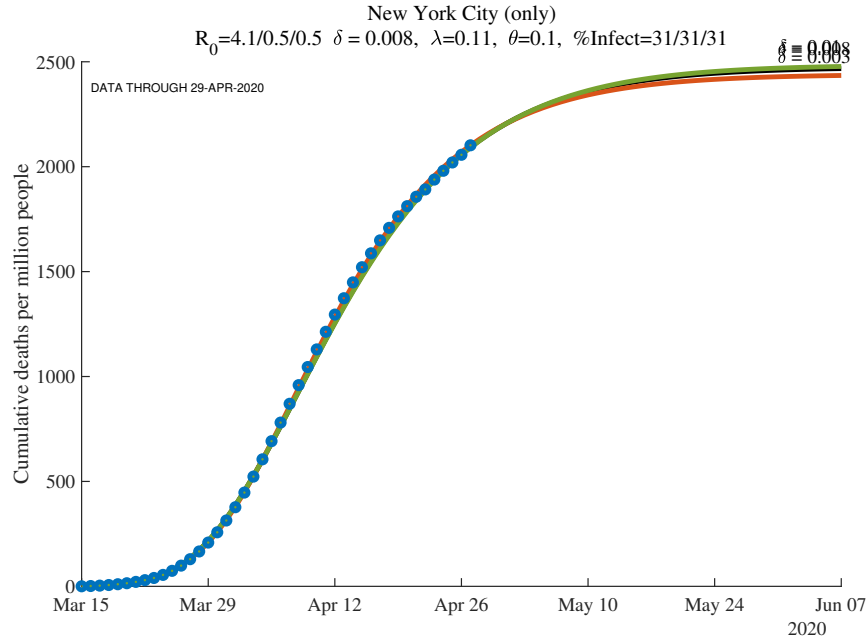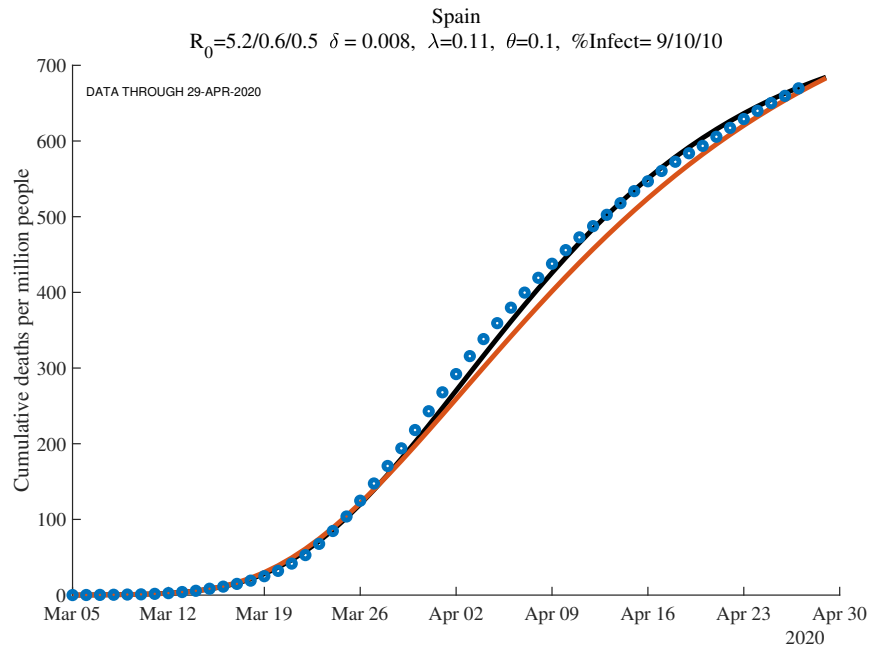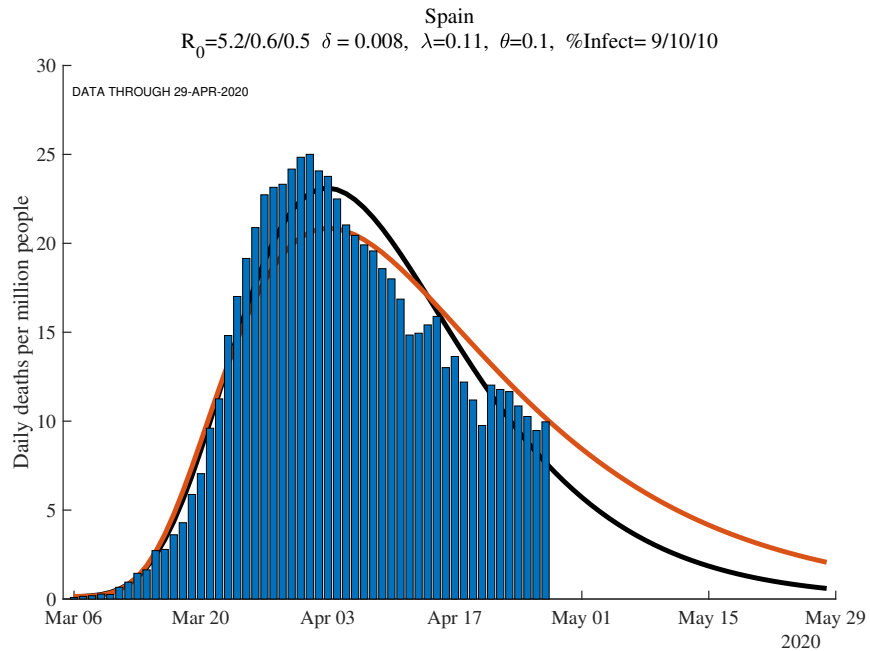

With a population of more than 8 million, this implies approximately 20,000 deaths.

The subtitle lines for these three figures also report the "%Infected" at different dates. These are the percent of people who are estimated to have *ever* been infected with the virus. For New York City, the numbers as of April 29 are 31 percent, and then in 30 days are estimated to equal 31 percent, which is also the value at the end of our simulation (the third number). We return in Section 5 to the implications of these high infection rates for herd immunity and re-opening.

The next set of graphs show results for Spain together with robustness to different values of γ, the rate at which cases resolve. Figures 4 and 5 show that $\gamma = 0.2$ fits the data somewhat better than $\gamma = 0.1$. Spain is estimated to have reduced R_0 from an initial value of 5.2 to 0.6 on April 29. Figure 6 suggest that the cumulative number of deaths per million in Spain may level off at around 700.

Next, we show how different values of the recovery-time parameter θ affect our results. Figure 7 shows the daily death numbers for Italy. The fit is good across a range of values for θ, including our benchmark value of $\theta = 0.1$ but also values of 0.05 (in red) and 0.2 (in green).

As we mentioned in the introduction, our online dashboard contains detailed and

Figure 4: Spain: Cumulative Deaths per Million People ($\gamma = .2/.1$)

Spain
R_0=5.2/0.6/0.5 $\delta = 0.008$, λ=0.11, θ=0.1, %Infect= 9/10/10

DATA THROUGH 29-APR-2020

Figure 5: Spain: Daily Deaths per Million People ($\gamma = .2/.1$)

Spain
R_0=5.2/0.6/0.5 $\delta = 0.008$, λ=0.11, θ=0.1, %Infect= 9/10/10

DATA THROUGH 29-APR-2020

extended results for more than 100 countries, states, and regions. We plan to update it frequently with the latest data:

Figure 6: Spain: Cumulative Deaths per Million (Future, $\gamma = .2/.1$)
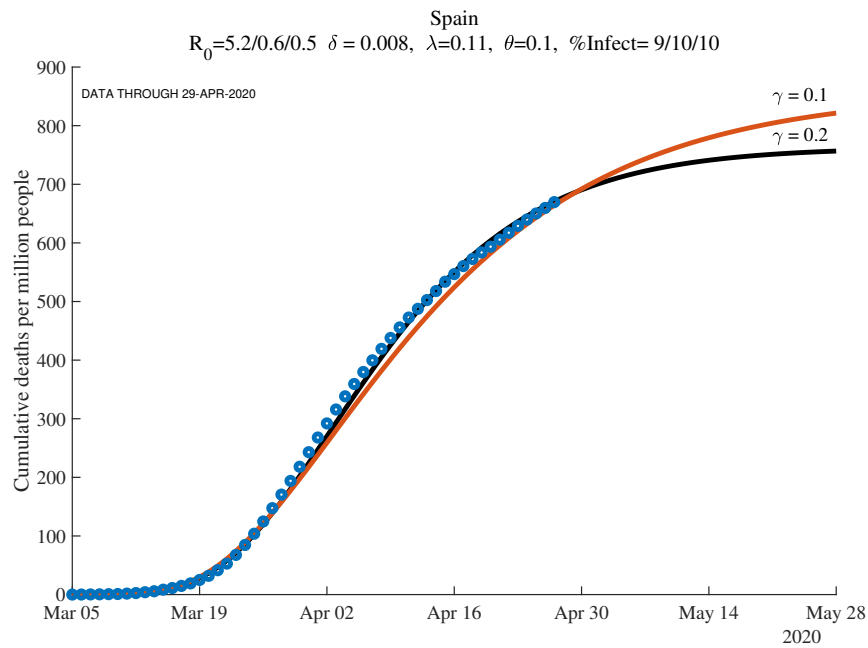
Figure 7: Italy: Daily Deaths per Million People ($\theta = .1/.05/.2$)

http://web.stanford.edu/people/chadj/Covid/Dashboard.html.
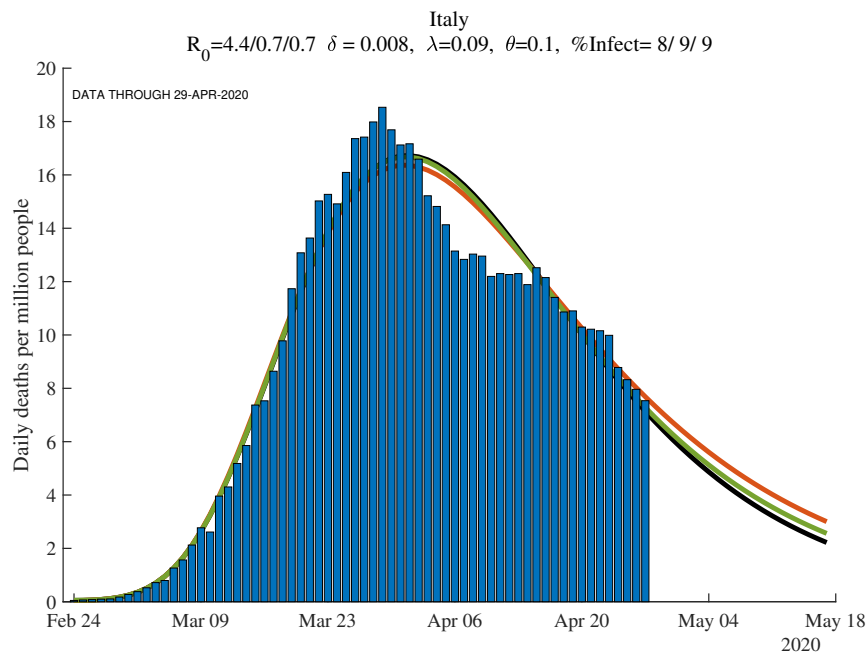
Figure 8: Italy: Cumulative Deaths per Million (Future, $\theta = .1/.05/.2$)
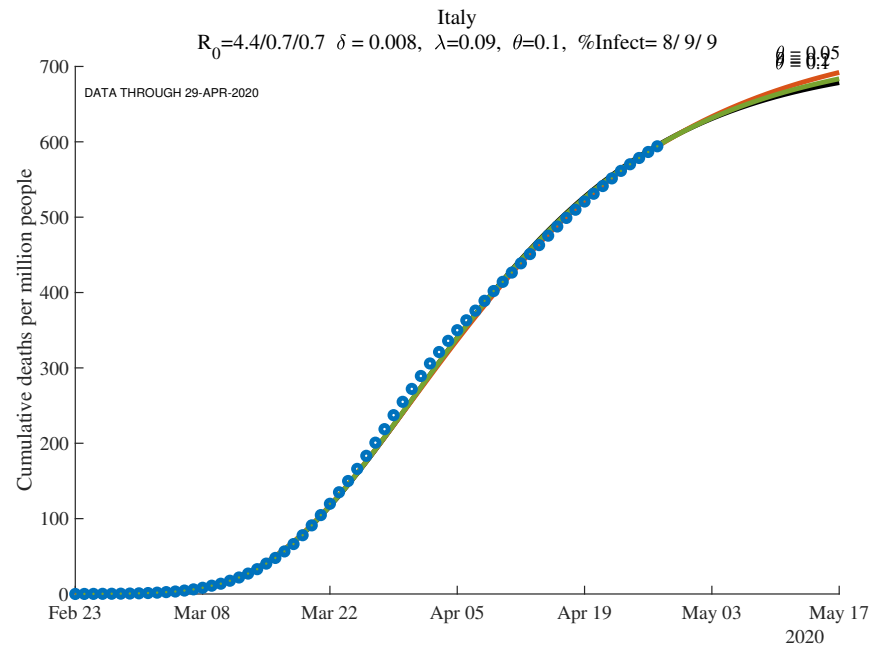
4.2   Seven Days of Simulations

Our next set of graphs illustrate how the simulations change as we get more data. Two key findings will emerge from these graphs. First, once countries or regions reach the peak and deaths start to decline, the forecasts converge well. Second, however, before that happens, the forecasts are very noisy. This makes sense: we are trying to forecast 30 to 60 days into the future based on 3 to 4 weeks of data using a very nonlinear model.

To begin, we illustrate the first point by showing the simulations for the seven most recent days in Spain and Italy in Figures 9 and 10. Both countries appear to have past their peak deaths and the simulations into the future are relatively similar for the past seven days. As a reminder, the ordering of the lines follows the colors of the rainbow, with the oldest forecast in red.

New York City.   The next two figures, Figures 11 and 12, show results for New York City now broadly defined to include the surrounding counties of Nassau, Rockland, Suffolk, and Westchester (which we call "New York City (plus)" in the graphs). Like Italy and Spain, New York City is now past its peak, so the forecasts have converged nicely.

California.   The preceding set of graphs show the forecasts converging for Spain, Italy, and New York. The next set of graphs show the enormous uncertainty that occurs before a region has peaked, in this case in California.

In Figure 13, the lines change noticeably as each new day provides new data. Notice that the latest parameter values suggest the R_0 fell in California from 2.8 to 0.9 as a result of social distancing and that only 1 percent of the population of California had ever been infected as of April 29. This number is broadly consistent with the recent (but not conclusive) studies of Santa Clara County and Los Angeles, two of the hardest hit places in California, that suggested around 2.5% to 5% of the populations there had been infected (Bendavid et al., 2020).

Figure 14 shows the wild fluctuations in the forecasts of daily deaths in California as new data come in. Notice also that there is no systematic pattern. When deaths come in low, the curves can bend sharply, guessing that perhaps the peak has arrived. And when deaths come in high, the curve can unbend and suggest more exponential growth to come.

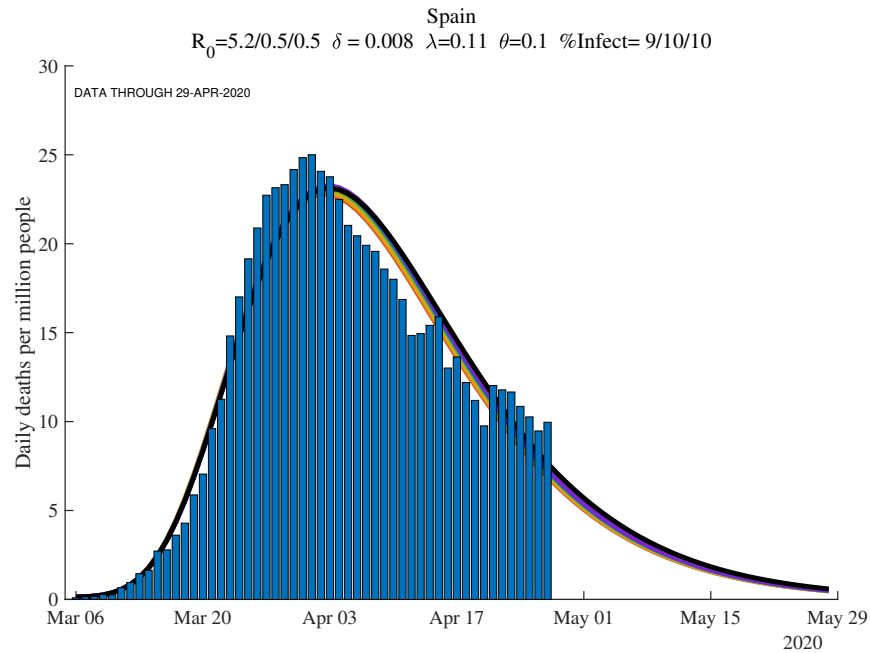Figure 9: Spain (7 days): Daily Deaths per Million People

Spain
R_0=5.2/0.5/0.5 $\delta = 0.008$ λ=0.11 θ=0.1 %Infect= 9/10/10

DATA THROUGH 29-APR-2020

Figure 10: Italy (7 days): Daily Deaths per Million People

Italy
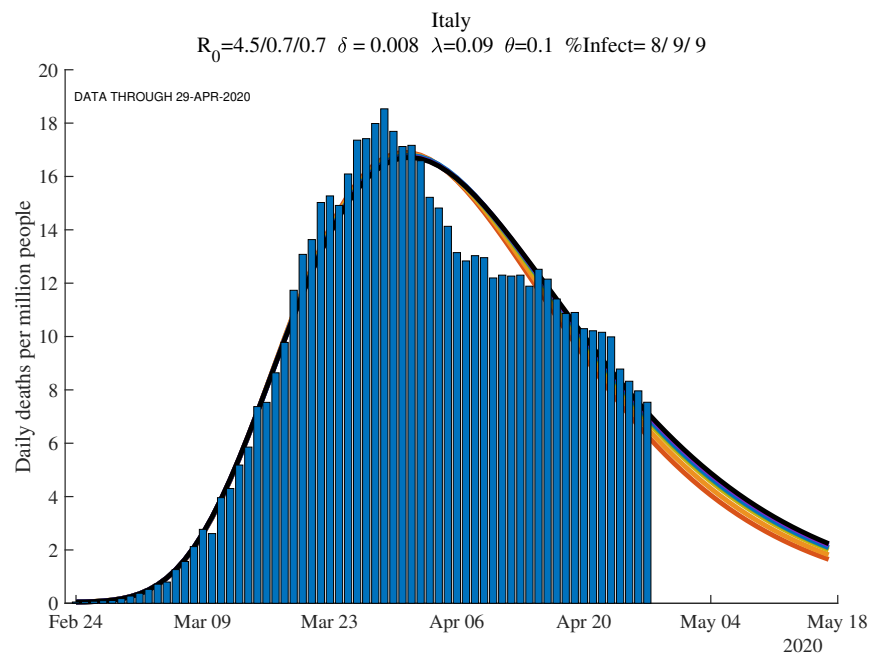R_0=4.5/0.7/0.7 $\delta = 0.008$ λ=0.09 θ=0.1 %Infect= 8/ 9/ 9

DATA THROUGH 29-APR-2020

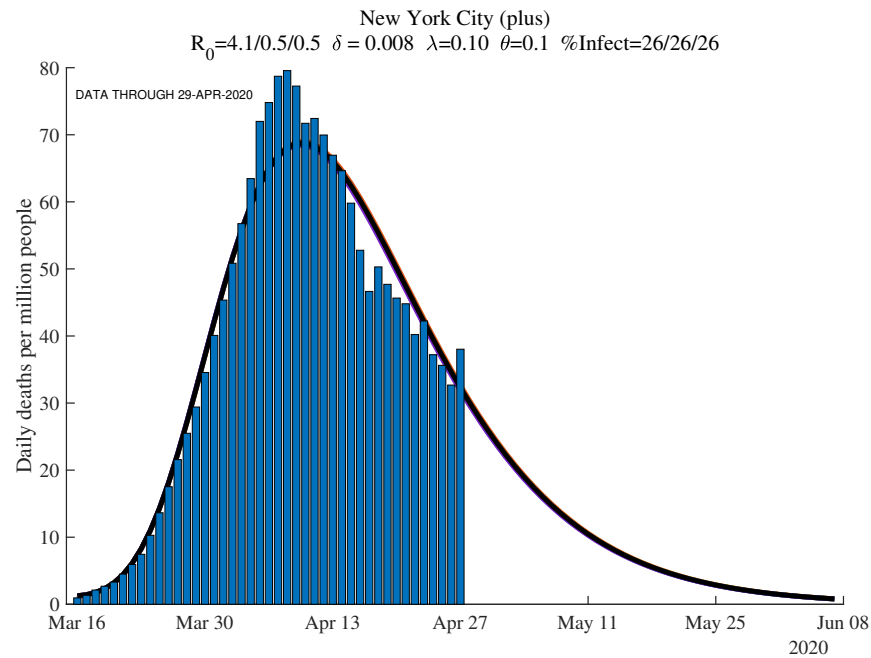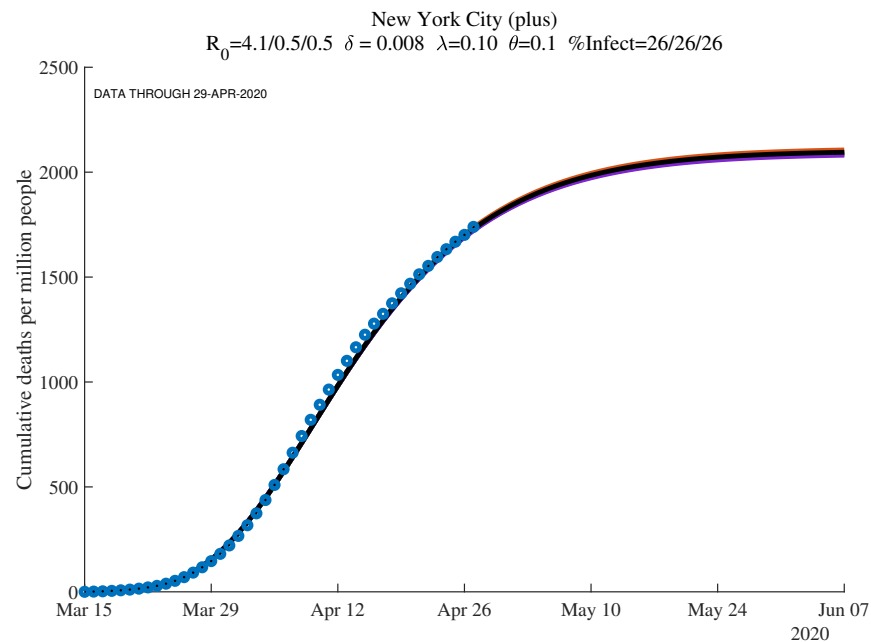Figure 11: New York City (7 days): Daily Deaths per Million People

Figure 12: New York City (7 days): Cumulative Deaths per Million (Future)

These very different paths cumulative into very different forecasts for the total num-
ber of deaths in California, as shown in Figure 15. By May 29, simulation results suggest

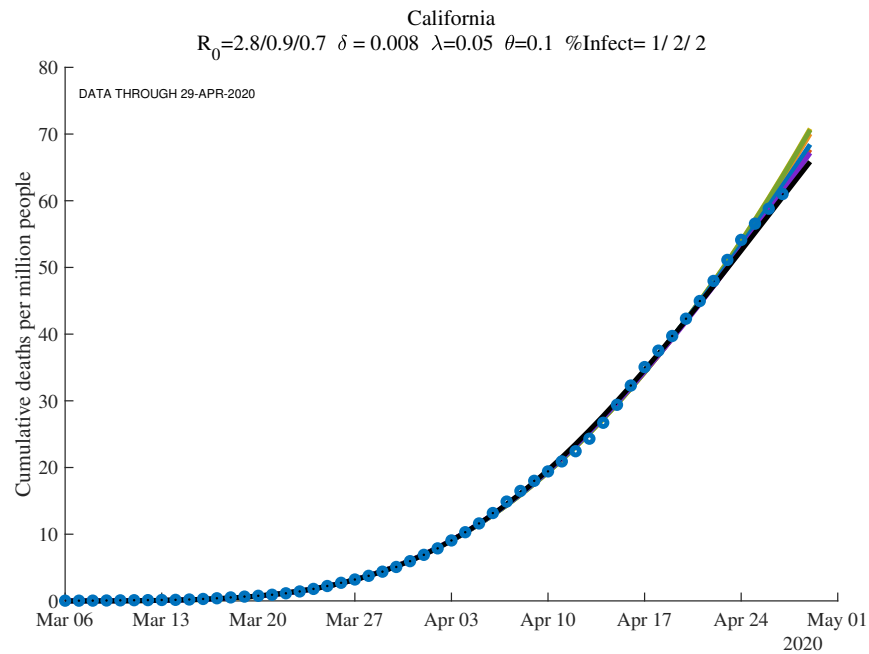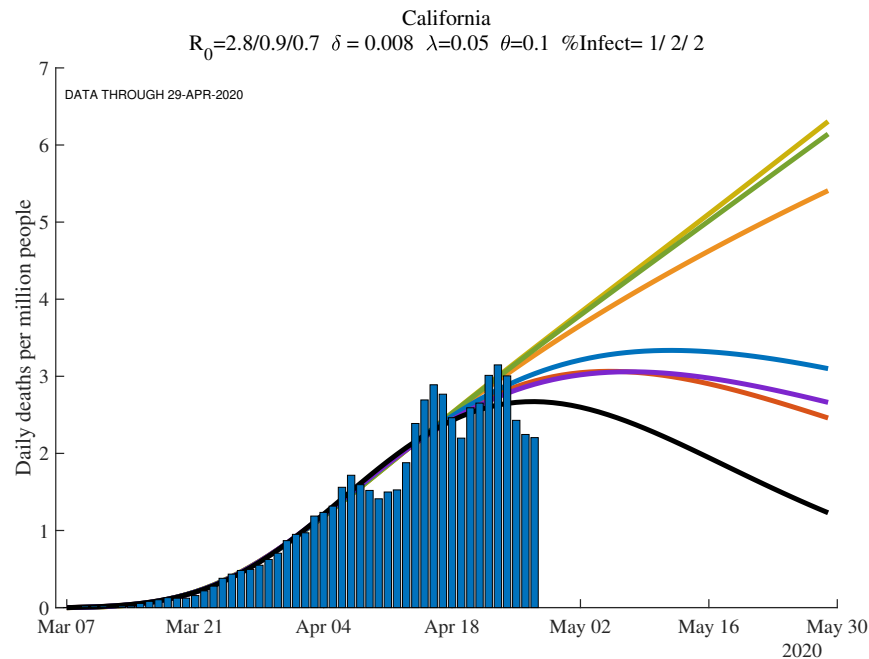Figure 13: California (7 days): Cumulative Deaths per Million

California
R_0=2.8/0.9/0.7 $\delta = 0.008$ λ=0.05 θ=0.1 %Infect= 1/ 2/ 2

Figure 14: California (7 days): Daily Deaths per Million People

California
R_0=2.8/0.9/0.7 $\delta = 0.008$ λ=0.05 θ=0.1 %Infect= 1/ 2/ 2

anywhere from 100 to 250 deaths per million. This compares to a current stock of deaths of around 50 per million.

Figure 15: California (7 days): Cumulative Deaths per Million, Log Scale
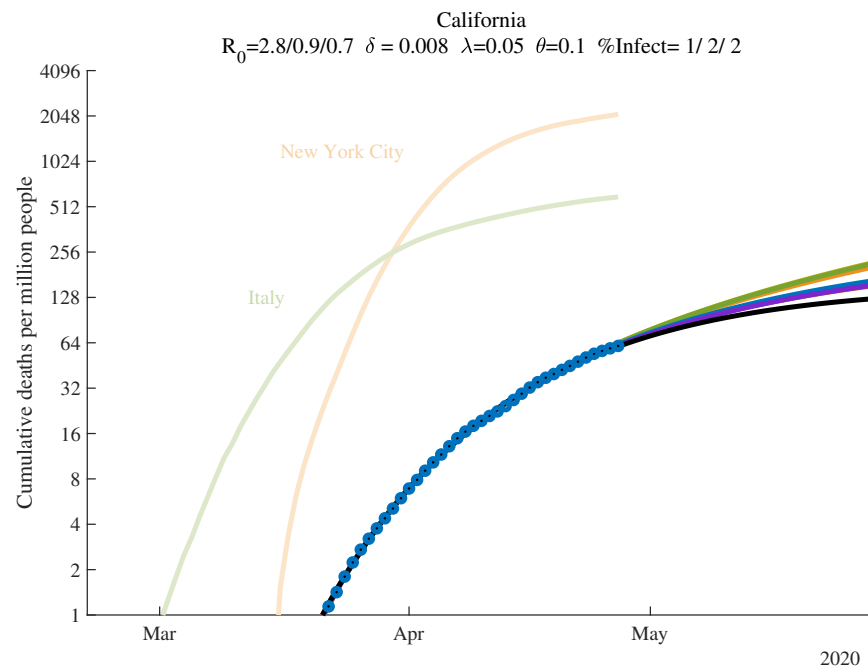
United Kingdom.   A similar illustration of this uncertainty is present in the graphs for the United Kingdom. Figure 16 shows the changing line as different death numbers come in. The percent ever infected in the U.K. is estimated to be 8 percent as of April 29.

Figure 17 shows the fluctuations in the forecasts of daily deaths in the United Kingdom as new data come in. When we first made these graphs two weeks ago, the U.K. forecasts were very sensitive to each new day of data. Now, however, the simulations have converged nicely. Notice that the estimates of R_0 are 4.0 initially and 0.6 for April 29.

The cumulative forecast is then shown in Figure 18. First the number of predicted deaths is high, 700 per million — very similar to Italy. Second, the forecasts have converged — there are 7 lines in the figure, but they all lie beneath the single black simulation line.

United States.   Figure 19 shows the daily death numbers for the United States as a country. Once again, when we made this graph a week ago, there was a fair amount of instability in the 7 days of simulations for the United States. Now, however, the seven

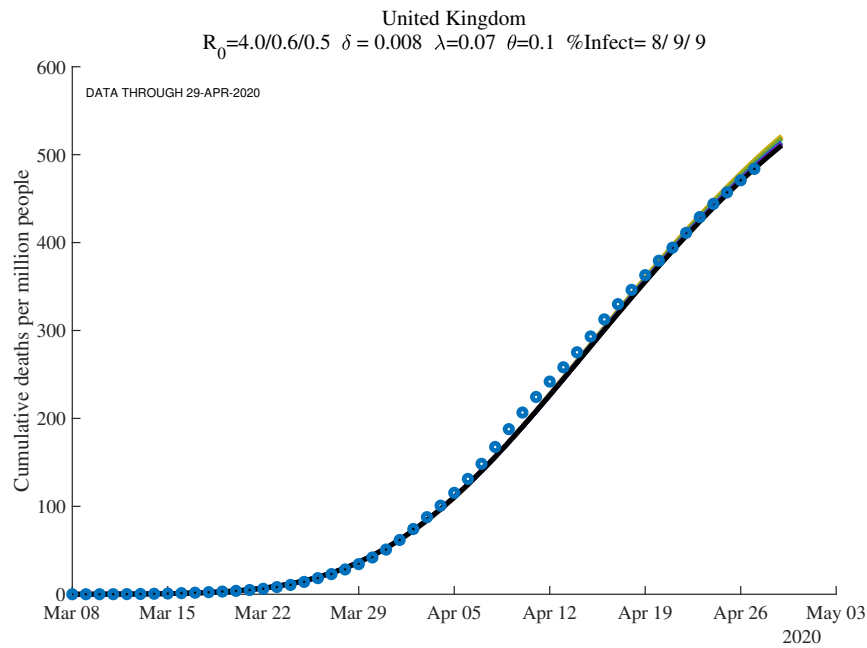Figure 16: United Kingdom (7 days): Cumulative Deaths per Million
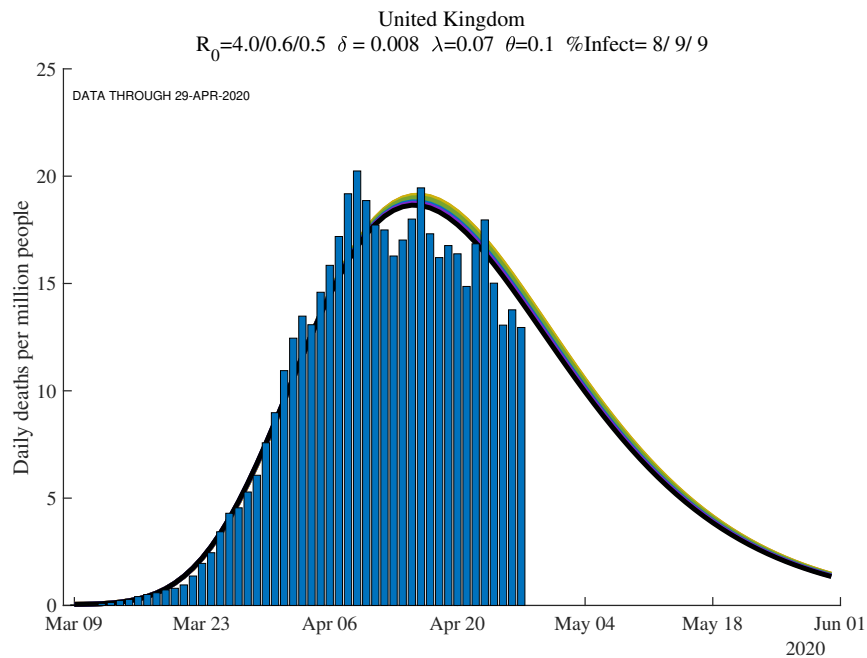

Figure 17: United Kingdom (7 days): Daily Deaths per Million People

lines lie on top of each other.

The cumulative forecast is then shown in Figure 20. The number of predicted deaths

Figure 18: United Kingdom (7 days): Cumulative Deaths per Million, Log Scale
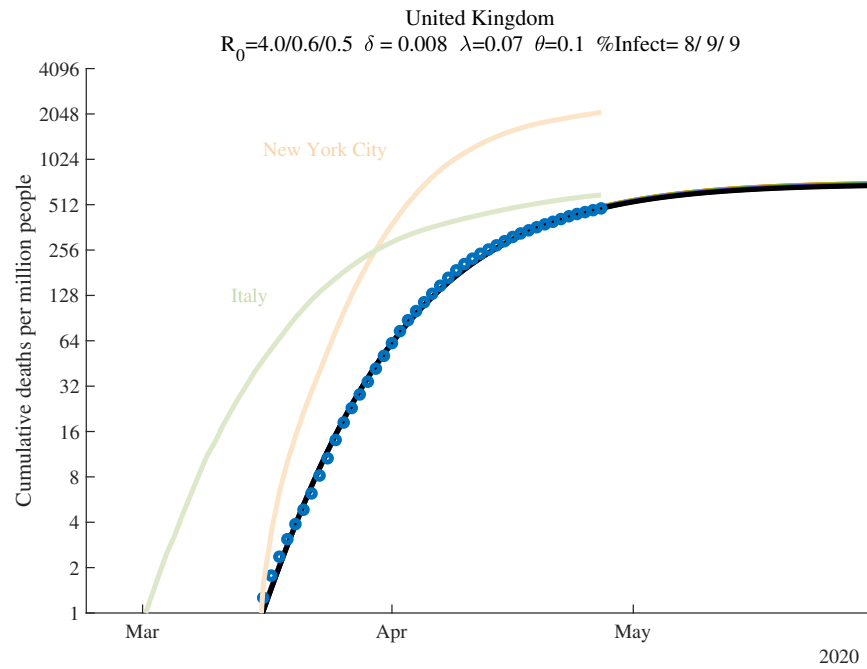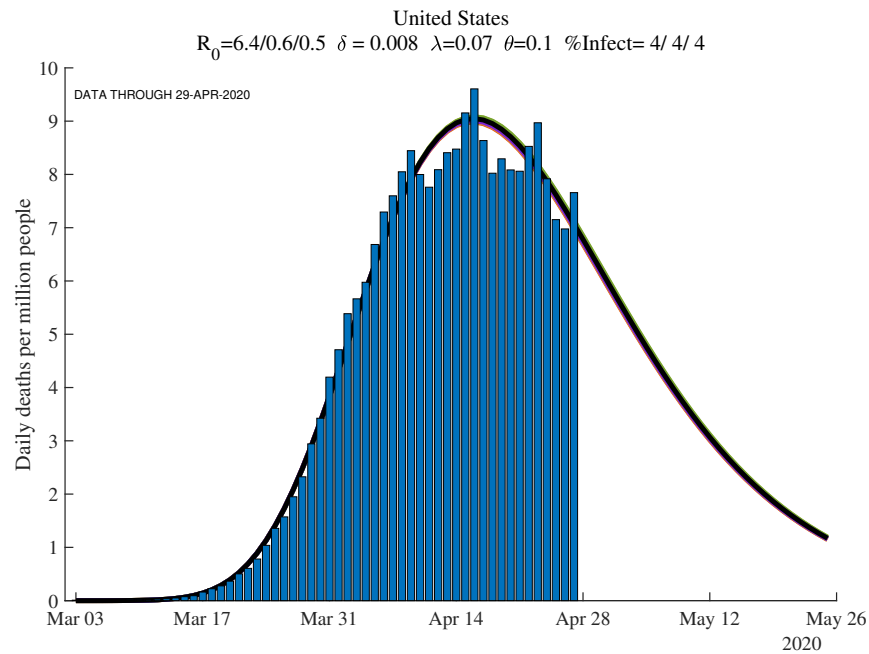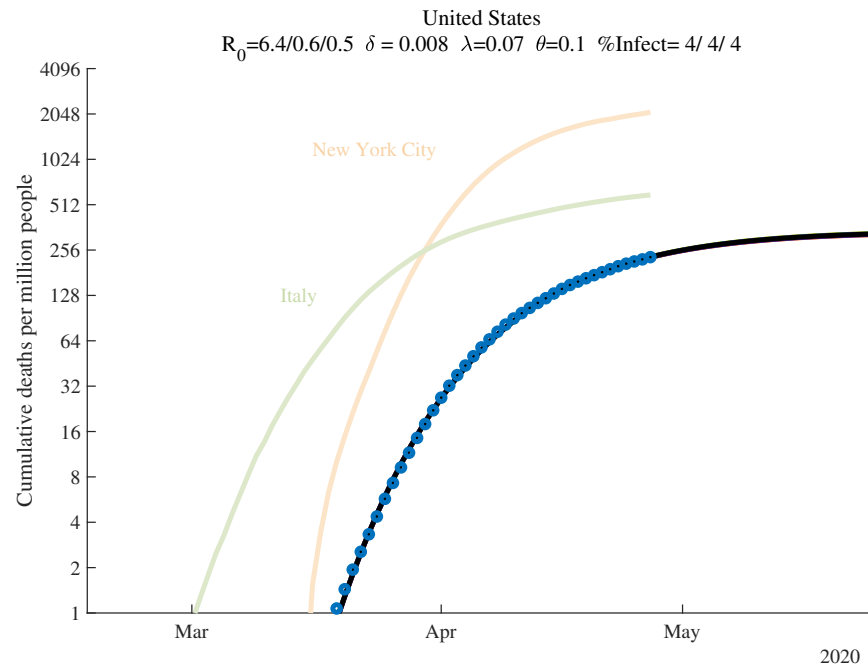

Figure 19: United States (7 days): Daily Deaths per Million People

is relatively high, with the lowest number being around 400 per million. Notice that even this low-end death rate implies 120,000 deaths for the United States as a whole.

Figure 20: United States (7 days): Cumulative Deaths per Million, Log Scale

United States
R_0=6.4/0.6/0.5 $\delta = 0.008$ λ=0.07 θ=0.1 %Infect= 4/ 4/ 4

Massachusetts and Sweden.   To show additional examples where this convergence has not yet occurred, consider Massachusetts and Sweden, shown in Figures 21 through 24. The outcome over the next month or two in these places is still quite uncertain.

Other countries.   The appendix and the dashboard contain additional graphs for other countries and regions. The patterns, uncertainties, and biases we have highlighted above are apparent in these graphs as well. These graphs will be updated frequently.

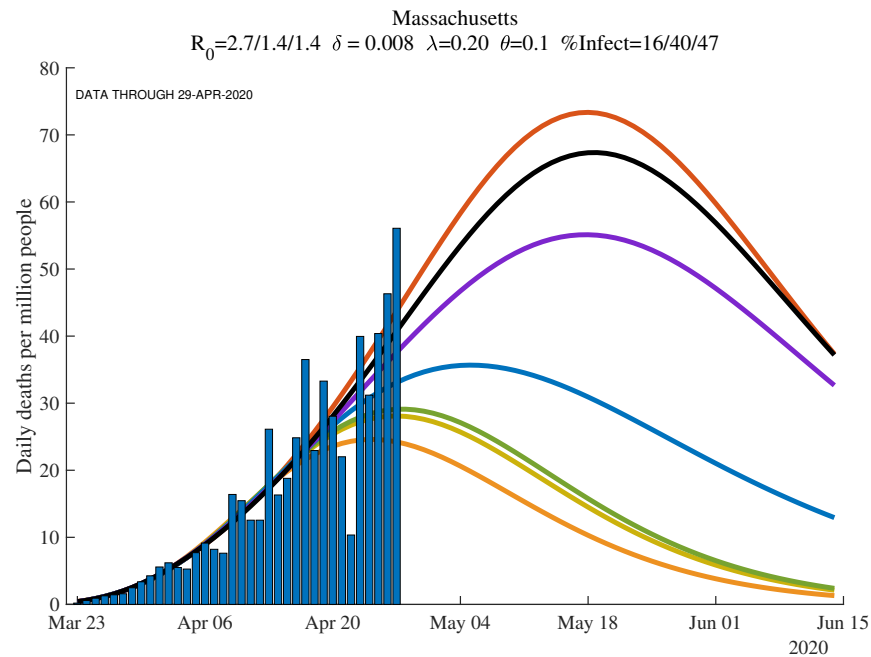Figure 21: Massachusetts (7 days): Daily Deaths per Million People

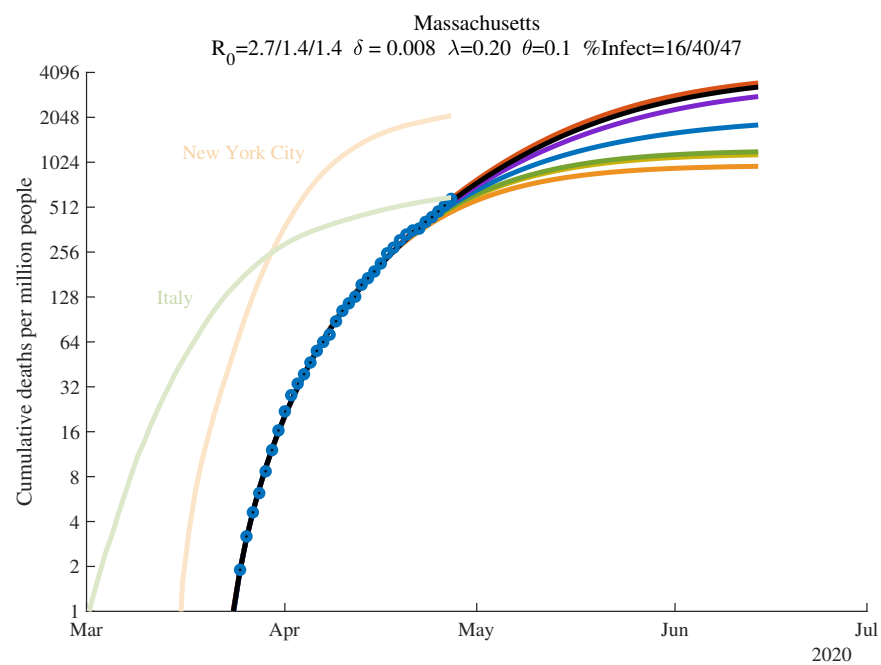Figure 22: Massachusetts (7 days): Cumulative Deaths per Million, Log Scale

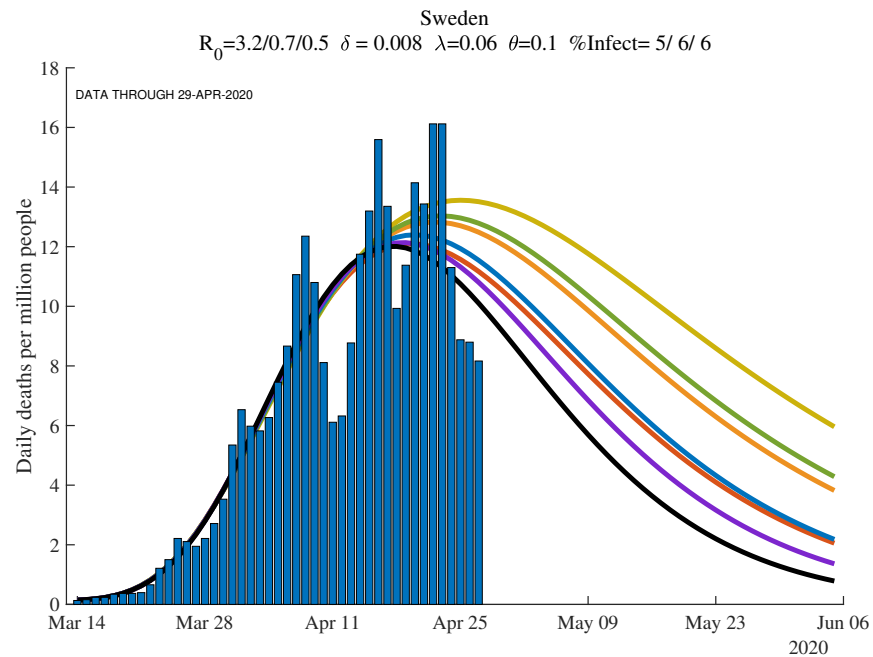Figure 23: Sweden (7 days): Daily Deaths per Million People
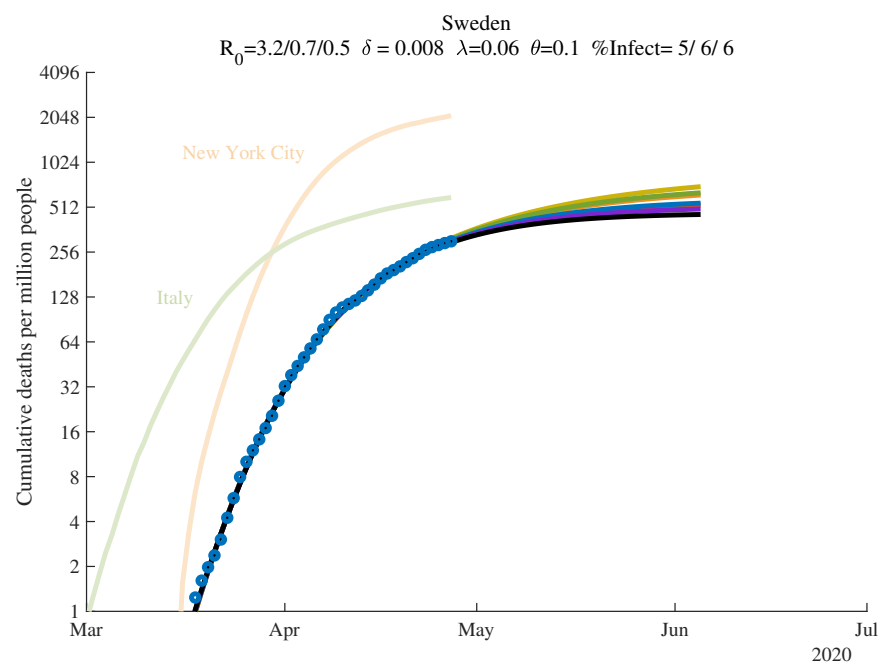

Figure 24: Sweden (7 days): Cumulative Deaths per Million, Log Scale

5.   Herd Immunity and Re-Opening the Economy

An important question at this stage of the pandemic is how we go about re-opening the economy. The estimation we've conducted has something helpful to contribute on this point.

First, Table 1 reports the estimated fraction of the population that has ever been infected as of April 29 for different countries and regions. Values for three different values of δ are also reported, with the baseline case of $\delta = 0.8\%$ in the center column. Two key things stand out in the table. First, consider the baseline. As we discussed above, we estimate that 31% of New Yorker City have ever been infected. This is consistent with the Cuomo-reported 21% number because it takes 3-4 weeks for people to show antibodies, so that number should be interpreted as saying that 21% of New York City residents had been infected as of April 1 or so. Our estimation suggests that the extra 10% between 21% and 31% are the new infections accumulated since April 1.

In contrast, only 3 percent of people in New York state outside of New York City and only 1 percent of Californians have ever been infected. There is enormous heterogeneity in ever-infected rates. Where do these numbers come from? In our model, the fraction δ of those infected eventually die, with the timing determined by γ and θ, but essentially suggesting that deaths today reflect infections from 15 days earlier. With an assumed death rate of $\delta = 0.8\%$, for each death, there are approximately 125 other people who have been infected. The large differences in the number of deaths per million in New York versus California then translate into these differences in infection rates. Interestingly, rates in Norway and South Korea are similarly very low, while ever-infected rates in Italy, Spain, and France are estimated to be around 7 to 9 percent.

The second point is that these numbers are — in an obvious way — very sensitive to the assumed value of δ. If you double the death rate, you (roughly) halve the ever-infected rate. If you halve the death rate, you (roughly) double the infected rate. And as we discuss in more detail next, in thinking about herd immunity and re-opening the economy, knowing the fraction ever-infected is crucial, at least under the important assumption that antibodies give rise to immunity for an extended period of time.

Notice that there is an important complementarity here. We would like the death rate to be low, not just because it means that fewer people die, but also because it means that lots of people will already have been infected. For example, if the true death rate

Table 1: Why Random Testing Would Be So Valuable

| | — Percent Ever Infected (today) — | | |
	$\delta = 0.3\%$	$\delta = 0.8\%$	$\delta = 1.0\%$
New York City (only)	81	31	25
New York	70	27	22
Detroit	56	22	18
New Jersey	50	20	16
Connecticut	41	16	13
Spain	25	9	8
Massachusetts	39	16	13
Italy	23	8	7
Michigan	27	10	8
United Kingdom	22	8	7
France	19	7	6
Chicago	19	7	6
District of Columbia	24	9	7
Sweden	14	5	4
United States	11	4	3
New York excluding NYC	9	3	3
Georgia	6	2	2
Washington	6	2	2
Los Angeles	6	2	2
Denmark	4	1	1
Florida	3	1	1
Ecuador	3	1	1
California	3	1	1

is 3 in 1000 rather than 8 in 1000, it means that 81 percent of New Yorkers have already been infected and the herd immunity effects would be very strong. In this sense, the finding that only 21 percent of New York City as ever infected as of April 1 is doubly bad news: it pushes up the death rate and means we are far from herd immunity, even in the place with the largest number of infections.

As Atkeson (2020), Stock (2020), and others have emphasized, random testing would be extremely helpful in identifying which of these cases is relevant. Moreover, the table suggests that if we are going to do 1000 random tests, you would much rather do them in New York City than in California. So few people are likely infected in California that it would be very hard to distinguish statistically between the different death rates, whereas even 1000 random tests would be very informative in New York City.

5.1   How far can we relax social distancing?

This brings us to the next reason why knowing the percent ever infected would be so useful. The complement of this number is the percent of the population that is still susceptible to the virus. Call this fraction $s(t) \equiv S(t)/N$ (or better might be $S(t)/(N - D(t))$ but $D(t)$ is so low that it makes no difference).[8]

We know from the basic SIR model that the virus will die out as long as

$$R_0(t)s(t) < 1$$

That is, if $R_0(t) \equiv \beta_t/\gamma$ is smaller than $1/s(t)$. The term $s(t)$ is the herd immunity term. The fewer people who are susceptible and the more people who are recovered and hence immune, the less our random interactions result in infections. In particular, we can relax social distancing — increase β_t and $R_0(t)$ — to the critical value such that $R_0(t)s(t)$ is just below one. That would mean that today's infected people infect fewer than one person on average, so the herd immunity keeps the virus from re-surging.

Table 2 shows these calculations for one month from now $(t + 30)$ given the baseline estimates from the model. For example, from the middle column, it is estimated that in one month's time, 73 percent of New York City (plus surrounding counties) will still be susceptible. This means we could relax social distancing to the point where R_0 would

---

[8]Notice, however, that our very stylized SIRD model is silent about how you map concrete policy decisions (i.e., should we or not open non-essential businesses) into changes in $R_0(t)$.

Table 2: Using Percent Susceptible to Estimate Herd Immunity, $\delta = 0.8\%$

	R_0	$R_0(t)$	Percent Susceptible t+30	R_0(t+30) with no outbreak	Percent way back to normal
New York City (plus)	4.1	0.5	73.7	1.4	23.0
Detroit	3.9	1.0	73.1	1.4	12.5
New Jersey	4.6	0.8	78.0	1.3	13.6
Spain	5.2	0.5	90.5	1.1	12.2
Massachusetts	2.7	1.4	59.5	1.7	24.2
Italy	4.5	0.7	91.2	1.1	9.8
Michigan	2.8	0.8	88.0	1.1	14.9
Louisiana	4.0	0.8	90.2	1.1	8.1
United Kingdom	4.0	0.6	91.3	1.1	13.8
France	11.0	0.6	92.4	1.1	4.6
Chicago	3.6	0.9	90.8	1.1	8.5
District of Columbia	2.1	1.2	78.3	1.3	8.2
Netherlands	4.8	0.7	94.3	1.1	9.1
Sweden	3.2	0.7	94.2	1.1	15.6
Pennsylvania	3.0	0.7	95.1	1.1	14.7
United States	6.4	0.6	95.8	1.0	7.3
Maryland	4.8	1.1	87.6	1.1	-0.2
New York excluding NYC	3.7	0.7	96.3	1.0	11.5
Georgia	3.5	0.6	97.7	1.0	14.7
Washington	1.8	0.7	97.7	1.0	27.3
Germany	4.0	0.6	98.3	1.0	12.5
Denmark	3.1	0.7	98.4	1.0	13.1
Iran	3.9	0.8	98.6	1.0	8.2
Kentucky	2.6	0.7	98.3	1.0	17.7
Wisconsin	2.5	0.5	98.8	1.0	23.6
Ecuador	3.2	1.0	98.4	1.0	1.3
California	2.7	0.9	98.3	1.0	7.6
Norway	2.6	0.5	99.3	1.0	23.3
Brazil	2.5	1.3	95.2	1.1	-19.6

rise to $1/0.73 = 1.4$. This compares to the current estimate for New York City of 0.5 and the initial estimate of 4.1. In other words, a month from now, in this simulation (which may not adequately capture the real world!), New York City could move 23% ([1.4-0.5]/[4.1-0.5]) of the way back to normal and see no re-surgence of the virus.

The rest of the state of New York, in contrast, is estimated to still have 96% of the population susceptible a month from now. So outside of the city, New York needs to maintain its R_0 at 1.0 versus its current level of 0.7 to keep the virus from spreading. New York City and the rest of New York state need very different policies if the fraction of the population that remains susceptible is as different as these estimates imply.

Places with current values of $R_0 < 1$ can relax somewhat and still keep the virus in check. But the basic news from this table is that with a death rate of 0.8%, there is very little herd immunity that has been accumulated and our scope for relaxing social distancing is limited.[9]

On the other hand, if somehow this 0.8% death rate is overstated and the true death rate is only 0.3%, the picture changes. The herd immunity effects for this case are shown in Table 3. In this case, the greater New York City area has only about 32% of its population susceptible because ever-infected rates are much higher. New York City could move 100 percent of the way back to normal, raising R_0 to 3.2, and still not see a resurgence of the virus. Other places would also have more scope to relax social distancing, at least somewhat.

This is why random testing in several places with a relatively high deaths per million is so important: it will help us learn whether the numbers in Table 2 or 3 are more relevant.

5.2 What happens if we re-open the economy?

We now show simulations of what happens if we relax social distancing and increase the value of R_0 (i.e. β). The next set of graphs assume that more generous policies are adopted on the last date of the preceding simulations (which varies by country), i.e. on something like May 29. Four simulation results are shown in each graph:

- The black line assumes the R_0 from April 29 remains in place forever.

---

[9]Notice, also, that these computations assume that individuals stay within their territories, and do not move among them, mixing infection rates across areas.

Table 3: Herd Immunity with a Much Lower Death Rate, $\delta = 0.3\%$

	R_0	$R_0(t)$	Percent Susceptible t+30	R_0(t+30) with no outbreak	Percent way back to normal
New York City (plus)	3.2	0.9	31.6	3.2	100.5
Detroit	3.5	1.4	39.8	2.5	52.6
New Jersey	4.1	0.9	47.1	2.1	37.4
Spain	5.1	0.7	74.6	1.3	15.3
Massachusetts	2.3	1.5	40.4	2.5	118.9
Italy	4.4	0.9	76.6	1.3	12.5
Michigan	2.9	1.0	68.8	1.5	24.7
Louisiana	4.1	1.0	74.6	1.3	11.8
United Kingdom	3.8	0.7	77.4	1.3	19.8
France	10.2	0.6	80.3	1.2	6.5
Chicago	3.6	1.0	76.4	1.3	13.0
District of Columbia	2.2	1.3	56.5	1.8	53.2
Netherlands	4.7	0.8	84.8	1.2	10.6
Sweden	3.2	0.8	84.2	1.2	17.4
Pennsylvania	3.0	0.7	87.3	1.1	18.2
United States	6.1	0.6	88.8	1.1	9.0
Maryland	4.7	1.2	73.7	1.4	4.7
New York excluding NYC	3.7	0.7	90.1	1.1	12.8
Georgia	3.5	0.6	93.9	1.1	16.1
Washington	1.8	0.7	94.0	1.1	30.1
Germany	3.9	0.6	95.6	1.0	13.8
Denmark	3.0	0.7	95.8	1.0	14.9
Iran	3.8	0.8	96.3	1.0	8.7
Kentucky	2.6	0.7	95.4	1.0	18.9
Wisconsin	2.5	0.5	96.7	1.0	24.6
Florida	3.9	0.6	96.3	1.0	11.9
Ecuador	3.2	1.0	95.7	1.0	2.2
California	2.8	0.9	95.4	1.0	8.1
Norway	2.6	0.5	98.0	1.0	23.9
Brazil	2.5	1.3	88.7	1.1	-14.2

- The red line assumes that R_0 rises slightly to $1/s(t)$, i.e. to the level that would just keep the virus from exploding based on the current number of susceptible people (as of April 29 again).

- The green line assumes we increase R_0 by 25% back toward normal, i.e. by 25% of the way back from the April 29 estimate toward the initial level of R_0.

- Finally, the purple line is similar except we increase R_0 by to 50% of the way back to normal.

For these simulations, we assume that the new value of R_0 remains in place forever. In practice, of course, one imagines that if deaths started to rise significantly, mitigation policies would be re-adopted.

For comparison, these graphs also show — in the light brown bars — the daily deaths per million in New York City. This is helpful in judging how severe a resurgence of the virus might be in different scenarios. For example, if Spain were to move 50 percent of the way back to normal (the purple line), the predicted resurgence would be much more severe than New York City experienced. Even with a 25 percent return to normal, the rise in deaths would be comparable to New York City, albeit with a lag into the late summer.
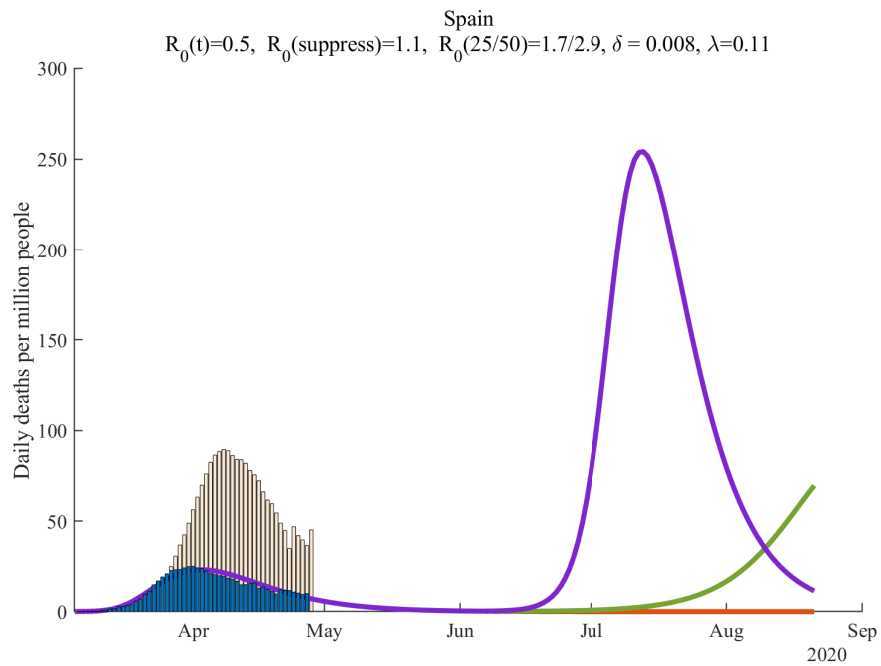
Figure 25: Spain: Re-opening

Spain
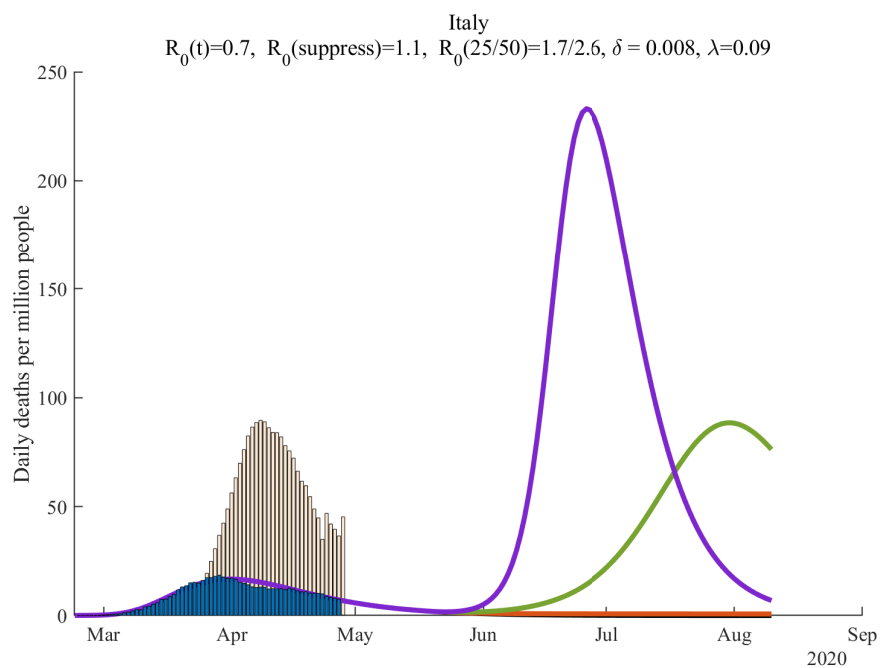$R_0(t)=0.5$, $R_0(\text{suppress})=1.1$, $R_0(25/50)=1.7/2.9$, $\delta = 0.008$, $\lambda=0.11$

Figure 26: Italy: Re-opening

Italy
$R_0(t)=0.7$, $R_0(\text{suppress})=1.1$, $R_0(25/50)=1.7/2.6$, $\delta = 0.008$, $\lambda=0.09$

Figure 27: New York City: Re-opening

New York City (plus)
$R_0(t)=0.5$, $R_0(\text{suppress})=1.4$, $R_0(25/50)=1.4/2.3$, $\delta = 0.008$, $\lambda=0.10$
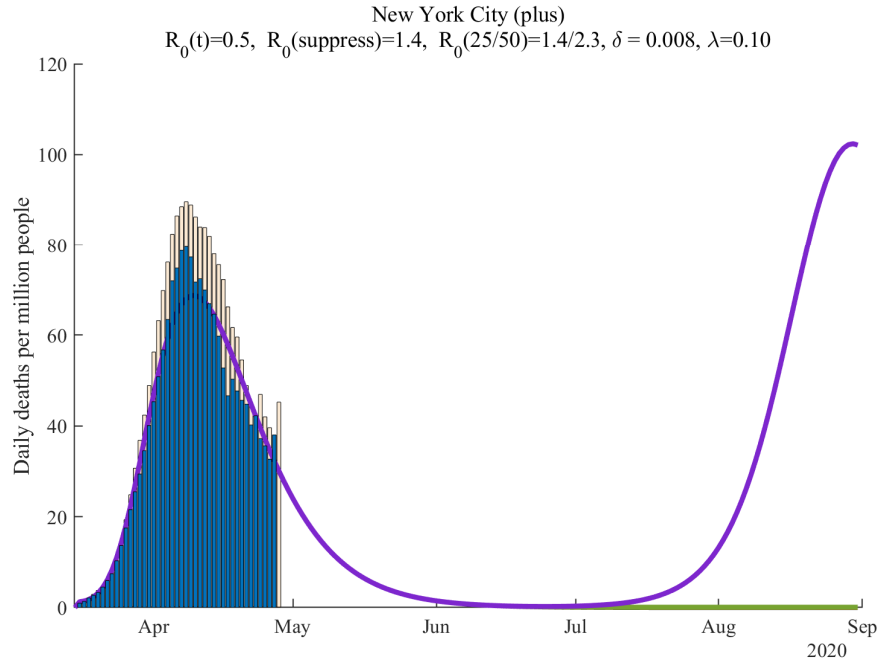
Figure 28: New York excluding NYC: Re-opening

New York excluding NYC
$R_0(t)=0.7$, $R_0(\text{suppress})=1.0$, $R_0(25/50)=1.4/2.2$, $\delta = 0.008$, $\lambda=0.11$
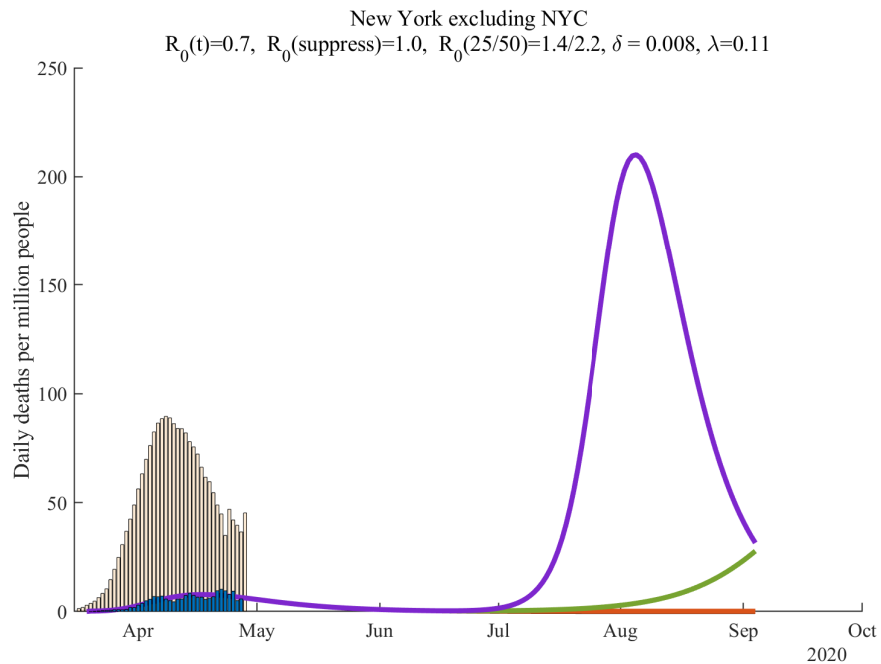
Figure 29: California: Re-opening

California
$R_0(t)$=0.9, R_0(suppress)=1.0, R_0(25/50)=1.3/1.8, δ = 0.008, λ=0.05
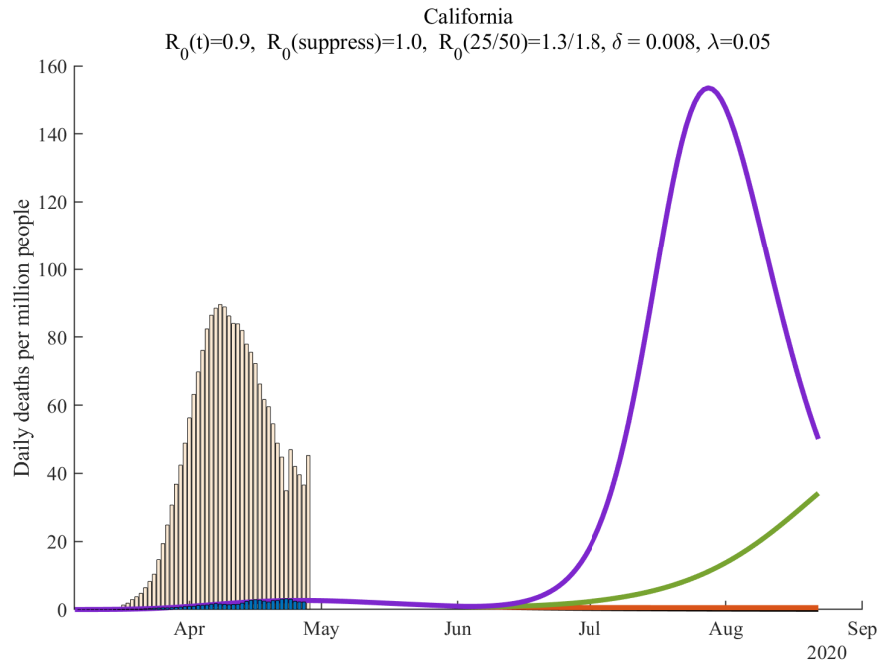
Figure 30: Sweden: Re-opening

Sweden
$R_0(t)$=0.7, R_0(suppress)=1.1, R_0(25/50)=1.3/1.9, δ = 0.008, λ=0.06
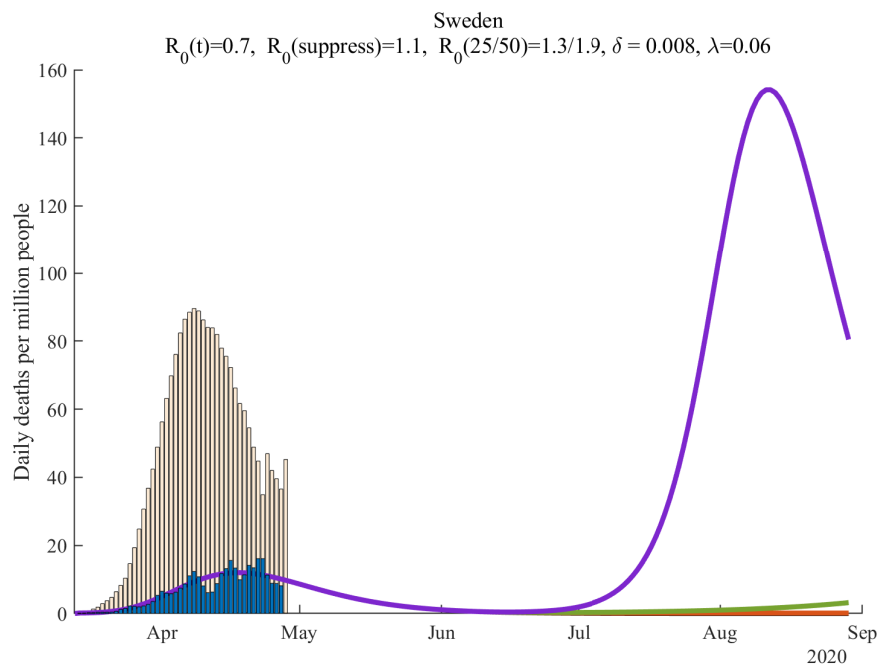
Figure 31: Detroit: Re-opening

Detroit
$R_0(t)=1.0$, R_0(suppress)=1.4, R_0(25/50)=1.7/2.5, $\delta = 0.008$, $\lambda=0.20$

Figure 32: Massachusetts: Re-opening

Massachusetts
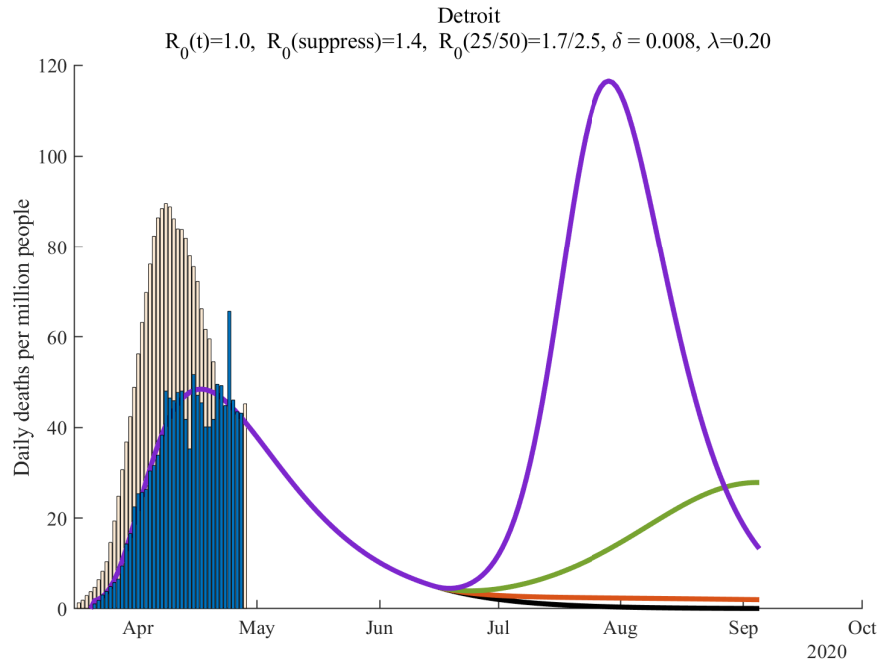$R_0(t)=1.4$, R_0(suppress)=1.9, R_0(25/50)=1.7/2.0, $\delta = 0.008$, $\lambda=0.20$

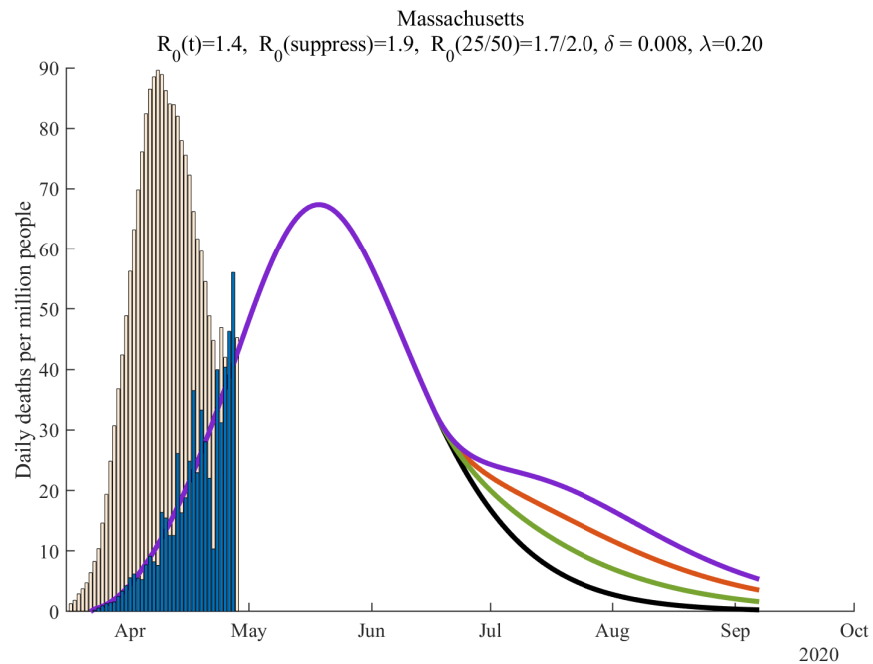## 6. Conclusions

The first thing to (re)emphasize about these simulations is that we are economists, not epidemiologists. These are speculations based on estimation and simulation of a relatively simple and standard model. We find them helpful for thinking about the dynamics of the pandemic, but great care should be taking in using these numbers to inform any policy discussion.

Relative to the standard SIRD model in the literature, we include a time-varying β, and therefore a time-varying R_0. We see this as important for capturing behavioral changes by individuals in response to the pandemic as well as policy changes related to social distancing. We also include an additional "recovering" state that is consistent with the medical evidence that cases seem to be infectious for four to five days while taking a total of several weeks or more to resolve. These changes connect the model better to the epidemiology of the virus and are important in improving the model's ability to fit the data.

We look forward to evidence in the coming weeks from random sampling in order to help us pin down the true death rate. Based on evidence from Sutton, Fuchs, D'Alton and Goffman (2020) for new mothers in New York City and from the preliminary results announced by Governor Cuomo for random antibody testing in New York, the fraction ever infected in New York City may have been around 21% on April 1 and may be around 30 percent today (May 1), suggesting mortality rates of about 0.8%. Alternatively, other evidence suggested the possibility that death rates were lower. Which of these scenarios is correct matters tremendously for the extent to which the economy can relax social distancing without a second wave of deaths.

References

Acemoglu, Daron, Victor Chernozhukov, Iván Werning, and Michael D. Whinston, "A Multi-Risk SIR Model with Optimally Targeted Lockdown," Technical Report, MIT 2020.

Álvarez, Fernando E., David Argente, and Francesco Lippi, "A Simple Planning Problem for COVID-19 Lockdown," Working Paper 26981, National Bureau of Economic Research April 2020.

Atkeson, Andrew, "How Deadly Is COVID-19? Understanding The Difficulties With Estimation Of Its Fatality Rate," Working Paper 26965, National Bureau of Economic Research April 2020.

Avery, Christopher, William Bossert, Adam Clark, Glenn Ellison, and Sara Fisher Ellison, "Policy Implications of Models of the Spread of Coronavirus: Perspectives and Opportunities for Economists," Working Paper 27007, National Bureau of Economic Research April 2020.

Bar-On, Yinon M., Avi Flamholz, Rob Phillips, and Ron Milo, "SARS-CoV-2 (COVID-19) by the numbers," Technical Report March 2020.

Bendavid, Eran, Bianca Mulaney, Neeraj Sood, Soleil Shah, Emilia Ling, Rebecca Bromley-Dulfano, Cara Lai, Zoe Weissberg, Rodrigo Saavedra, James Tedrow, Dona Tversky, Andrew Bogan, Thomas Kupiec, Daniel Eichner, Ribhav Gupta, John Ioannidis, and Jay Bhattacharya, "COVID-19 Antibody Seroprevalence in Santa Clara County, California," *medRxiv*, 2020.

Berger, David W, Kyle F Herkenhoff, and Simon Mongey, "An SEIR Infectious Disease Model with Testing and Conditional Quarantine," Working Paper 26901, National Bureau of Economic Research March 2020.

Bethune, Zachary A and Anton Korinek, "Covid-19 Infection Externalities: Trading Off Lives vs. Livelihoods," Technical Report, National Bureau of Economic Research 2020.

Bodenstein, Martin, Giancarlo Corsetti, and Luca Guerrieri, "Social Distancing and Supply Disruptions in a Pandemic," Technical Report 2020-031 04 2020.

Brauer, F., C. Castillo-Chavez, and Z. Feng, *Mathematical Models in Epidemiology*, Springer New York, 2019.

Eichenbaum, Martin S, Sergio Rebelo, and Mathias Trabandt, "The Macroeconomics of Epidemics," Working Paper 26882, National Bureau of Economic Research March 2020.

Farboodi, Maryam, Gregor Jarosch, and Robert Shimer, "Internal and External Effects of Social Distancing in a Pandemic," Technical Report 2020-47 2020. University of Chicago, Becker Friedman Institute for Economics Working Paper.

Garriga, Carlos, Rody Manuelli, and Siddhartha Sanghi, "Optimal Management of an Epidemic: An Application to COVID-19. A Progress Report," Technical Report April 2020. Federal Reserve Bank of St. Louis manuscript.

Hethcote, Herbert W., "The Mathematics of Infectious Diseases," *SIAM Review*, 2000, *42* (4), 599–653.

Hornstein, Andreas, "Social Distancing, Quarantine, Contact Tracing, and Testing: Implications of an Augmented SEIR-Mode," Technical Report, Federal Reserve Bank of Richmond 2020.

Hortasu, Ali, Jiarui Liu, and Timothy Schwieg, "Estimating the Fraction of Unreported Infections in Epidemics with a Known Epicenter: an Application to COVID-19," Working Paper 27028, National Bureau of Economic Research April 2020.

Hurwicz, Leonid, "On the Structural Form of Interdependent Systems," in Ernest Nagel, Patrick Suppes, and Alfred Tarski, eds., *Logic, Methodology and Philosophy of Science*, Vol. 44 of *Studies in Logic and the Foundations of Mathematics*, Elsevier, 1962, pp. 232 – 239.

Imperial College COVID-19 Response Team, "Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand," Technical Report, Imperial College 2020.

Johns Hopkins University CSSE, "2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository," 2020. Center for Systems Science and Engineering.

Karin, Omer, Yinon M Bar-On, Tomer Milo, Itay Katzir, Avi Mayo, Yael Korem, Boaz Dudovich, Eran Yashiv, Amos J Zehavi, Nadav Davidovich et al., "Adaptive cyclic exit strategies from lockdown to suppress COVID-19 and allow economic activity," *medRxiv*, 2020.

Katz, Josh and Margot Sanger-Katz, "N.Y.C. Deaths Reach 6 Times the Normal Level, Far More Than Coronavirus Count Suggests," *The New York Times*, April 2020.

Kermack, William Ogilvy and A. G. McKendrick, "A contribution to the mathematical theory of epidemics, part I," *Proceedings of the Royal Society of London. Series A*, 1927, *115* (772), 700–721.

_ and _ , "Contributions to the mathematical theory of epidemics. II – The problem of endemicity," *Proceedings of the Royal Society of London. Series A*, 1932, *138* (834), 55–83.

Korolev, Ivan, "Identification and Estimation of the SEIRD Epidemic Model for COVID-19," Technical Report, Binghamton University 2020.

_ , "What Does the Case Fatality Ratio Really Measure?," Technical Report, Binghamton University 2020.

Kucinskas, Simas, "Tracking R of COVID-19," *Available at SSRN 3581633*, 2020.

Linton, Oliver, "When will the Covid-19 pandemic peak?," Technical Report, University of Cambridge 2020.

New York Department of Health, "The NYSDOH Wadsworth Center's Assay for SARS-CoV-2 IgG," Technical Report April 2020.

Sanche, S., Y.T. Lin, C. Xu, E. Romero-Severson, N. Hengartner, and R. Ke, "High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2," *Emerging Infectious Diseases*, 2020, *26*.

Stock, James H, "Data Gaps and the Policy Response to the Novel Coronavirus," Working Paper 26902, National Bureau of Economic Research March 2020.

Sutton, Desmond, Karin Fuchs, Mary D'Alton, and Dena Goffman, "Universal Screening for SARS-CoV-2 in Women Admitted for Delivery," *New England Journal of Medicine*, April 2020.

The Economist, "Tracking covid-19 excess deaths across countries," April 2020.

Toda, Alexis Akira, "Susceptible-Infected-Recovered (SIR) Dynamics of COVID-19 and Economic Impact," 2020. U.C.S.D. manuscript.

A.   Graphs for Other Countries

Graphs for many countries and regions are shown below. We report detailed and extended results for even more countries, states, and cities on our web pages and plan to update the dashboard frequently.

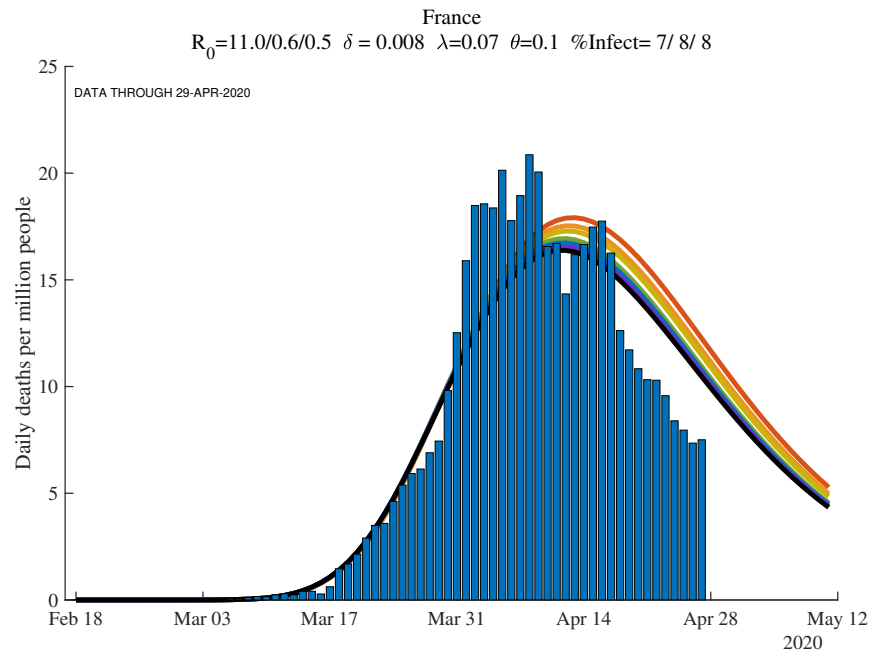### Figure 33: France (7 days): Daily Deaths per Million People

France
R_0=11.0/0.6/0.5 δ = 0.008 λ=0.07 θ=0.1 %Infect= 7/ 8/ 8

DATA THROUGH 29-APR-2020

Figure 34: South Korea (7 days): Daily Deaths per Million People

New Jersey
R_0=4.6/0.8/0.5 δ = 0.008 λ=0.06 θ=0.1 %Infect=20/22/22

DATA THROUGH 29-APR-2020
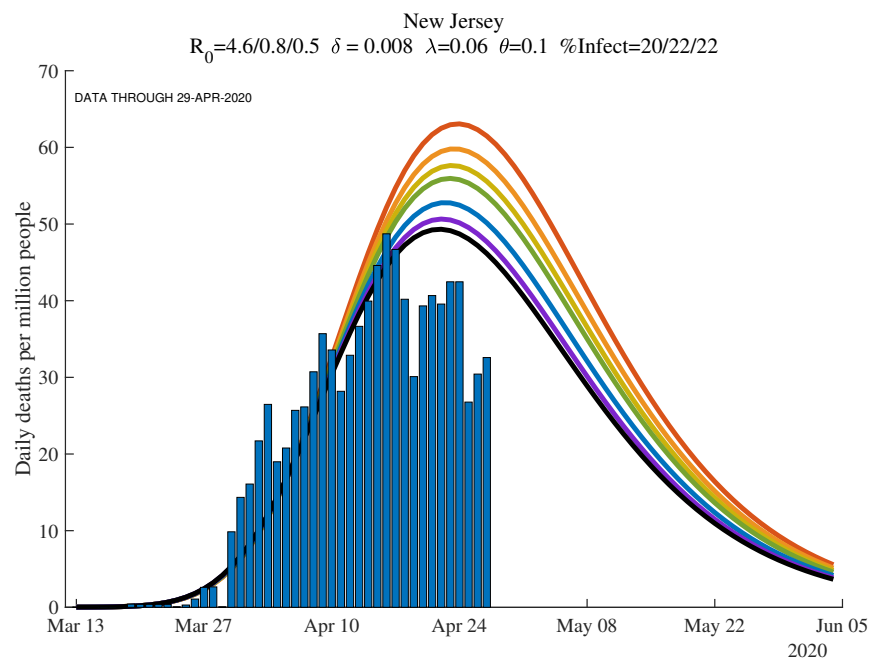
Figure 35: Washington (7 days): Daily Deaths per Million People

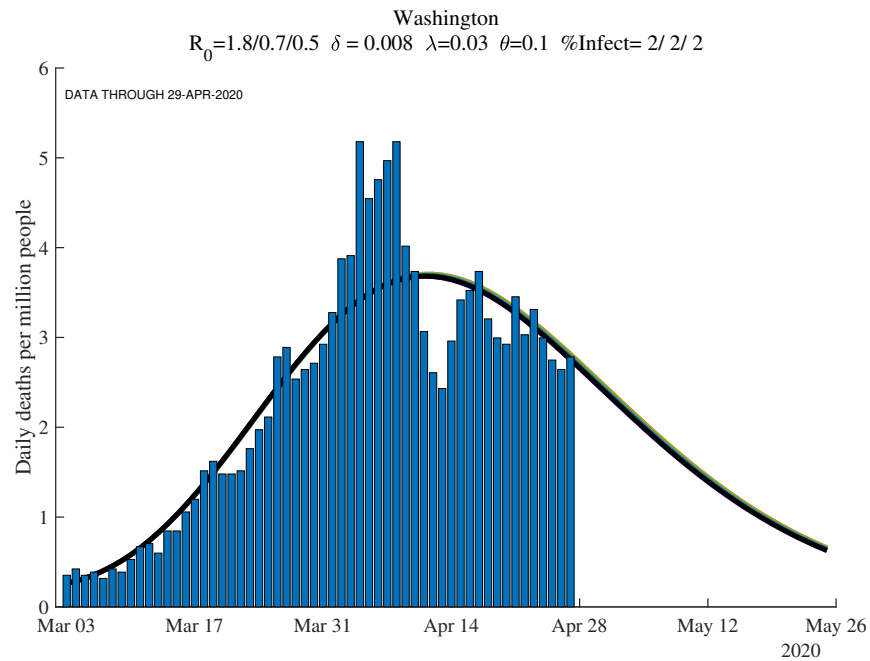

Figure 36: Louisiana (7 days): Daily Deaths per Million People

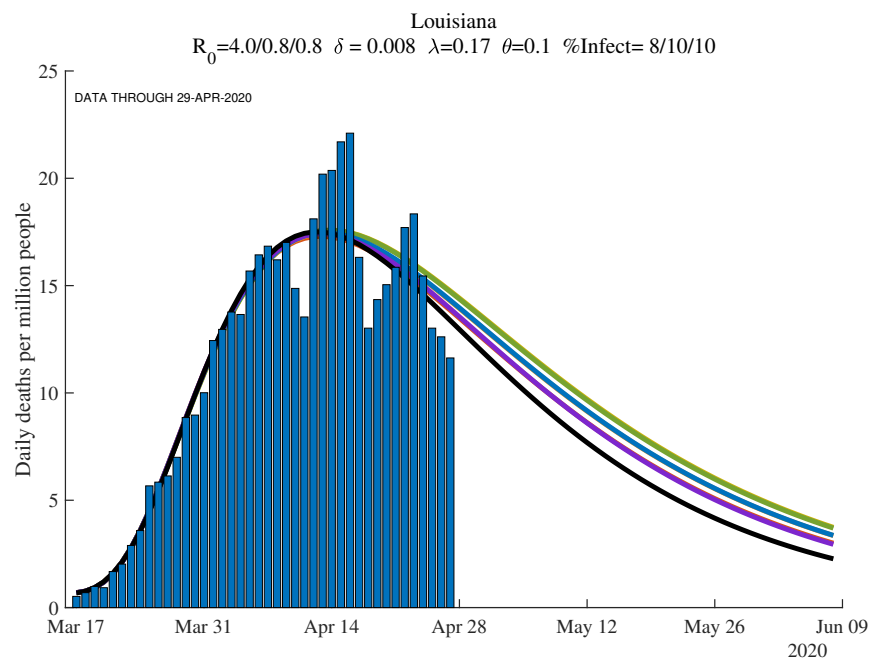Figure 37: Florida (7 days): Daily Deaths per Million People

Florida
R_0=4.0/0.6/0.5 δ = 0.008 λ=0.07 θ=0.1 %Infect= 1/ 1/ 1
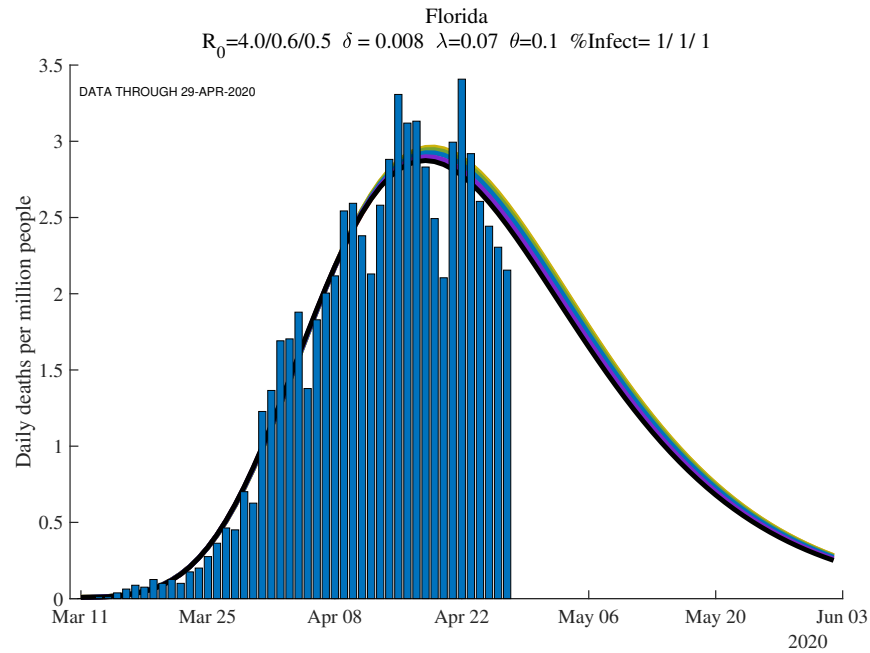
DATA THROUGH 29-APR-2020

Figure 38: Michigan (7 days): Daily Deaths per Million People

Michigan
R_0=2.8/0.8/0.8 δ = 0.008 λ=0.13 θ=0.1 %Infect=10/12/12

DATA THROUGH 29-APR-2020