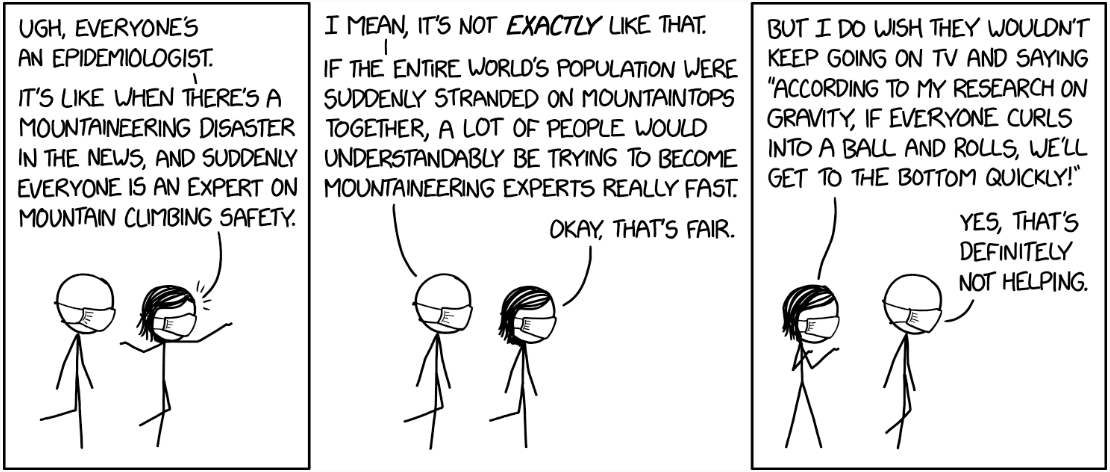
Estimating and Simulating a SIRD Model of COVID-19 for Many Countries, States, and Cities

Jesús Fernández-Villaverde[1] Chad Jones[2]

May 27, 2020

[1]University of Pennsylvania

[2]Stanford GSB


UGH, EVERYONE'S AN EPIDEMIOLOGIST.
IT'S LIKE WHEN THERE'S A MOUNTAINEERING DISASTER IN THE NEWS, AND SUDDENLY EVERYONE IS AN EXPERT ON MOUNTAIN CLIMBING SAFETY.
I MEAN, IT'S NOT EXACTLY LIKE THAT.
IF THE ENTIRE WORLD'S POPULATION WERE SUDDENLY STRANDED ON MOUNTAINTOPS TOGETHER, A LOT OF PEOPLE WOULD UNDERSTANDABLY BE TRYING TO BECOME MOUNTAINEERING EXPERTS REALLY FAST.
OKAY, THAT'S FAIR.
BUT I DO WISH THEY WOULDN'T KEEP GOING ON TV AND SAYING "ACCORDING TO MY RESEARCH ON GRAVITY, IF EVERYONE CURLS INTO A BALL AND ROLLS, WE'LL GET TO THE BOTTOM QUICKLY!"
YES, THAT'S DEFINITELY NOT HELPING.

Outline

- We want to take a basic SIRD model to the data for many countries, states, and cities:
  - Exploit variation across time and space.
  - Measure effects of social distancing via a time-varying β.
  - Make a more general point about structural vs. reduced-form parameters in SIRD models.
- Estimation and simulation:
  - Different countries, U.S. states, and cities.
  - Robustness to parameters and problem of underidentification.
  - "Forecasts" from each of the last 7 days.
  - Extended results available at: `https://web.stanford.edu/~chadj/Covid/Dashboard.html`.

Outline (continued)

- Re-opening and herd immunity: How much can we relax social distancing?
- How do we make more progress in understanding time-varying parameters and their relation to observed policies?
- Heterogeneous agents SIRD models.
- Example of policy counterfactuals with heterogeneous agents SIRD models: introducing a vaccine.

Basic model

Notation

- Stocks of people who are:

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

- Constant population size is N:

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

- Only one group. Why? I will return to this point repeatedly.

SIRD model: Overview

- Susceptible get infected at rate $\beta_t I_t / N$.
  - New infections $= \beta_t I_t / N \cdot S_t$.
- Infectiousness resolve at Poisson rate γ, so the average number of days that a person is infectious is $1/\gamma$. E.g., $\gamma = .2 \Rightarrow 5$ days.
- Post-infectious cases then resolve at Poisson rate θ. E.g., $\theta = .1 \Rightarrow 10$ days.
- Resolution happens in one of two ways:
  - Death: fraction δ.
  - Recovery: fraction $1 - \delta$.

SIRD model: Laws of motion

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

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

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

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

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

with $D_0 = 0$ and I_0.

Social distancing

- What about the time-varying infection rate β_t?
  - Disease characteristics – fixed, homogeneous (exceptions?).
  - Regional factors (NYC vs. Montana) – fixed, heterogeneous.
  - Social distancing – varies over time and space.
- Reasons why β_t may change over time:
  - Policy changes on social distancing.
  - Individuals voluntarily change behavior to protect themselves and others.
  - Superspreaders get infected quickly, but then recover and "burn out" early.
  - Spatial aggregation: SIRD model is highly non-linear.

Recovering β_t and $\mathcal{R}_{0t}$, I

- Recovering β_t, a latent variable, from the data is straightforward.
- D_{t+1}: stock of people who have died as of the *end* of date $t+1$.
- $\Delta D_{t+1} \equiv d_{t+1}$: number of people who died on date $t+1$.
- After some manipulations, we can "invert" the model and get:

$$\beta_t = \frac{N}{S_t}\left(\gamma + \frac{\frac{1}{\theta}\Delta\Delta d_{t+3} + \Delta d_{t+2}}{\frac{1}{\theta}\Delta d_{t+2} + d_{t+1}}\right)$$

and:

$$S_{t+1} = S_t\left(1 - \beta_t \frac{1}{\delta\gamma N}\left(\frac{1}{\theta}\Delta d_{t+2} + d_{t+1}\right)\right)$$

Recovering β_t and $\mathcal{R}_{0t}$, II

- With these two equations, a time series for d_t, and an initial condition $S_0/N \approx 1$, we iterate forward in time and recover β_t and S_{t+1}.
- We are using *future* deaths over the subsequent 3 days to tell us about β_t today.
- While this means our estimates will be three days late (if we have death data for 30 days, we can only solve for β for the first 27 days), we can still generate an informative estimate of β_t.
- More general point about SIRD models: state-space representation that we can exploit.

Recovering β_t and $\mathcal{R}_{0t}$, III

- We can also recover the basic reproduction number:

$$\mathcal{R}_{0t} = \beta_t \times 1/\gamma$$

  and the effective reproduction number:

$$\mathcal{R}_{et} = \mathcal{R}_{0t} \cdot S_t/N$$

- Now we can simulate the model forward using the most recent value of β_T and gauge where a region is headed in terms of the infection and current behavior.
- And we can correlate the β_t with other observables to evaluate the effectiveness of certain government policies such as mandated lock downs.

An endogenous $\mathcal{R}_{0t}$

- Individuals react endogenously to risk.
- Indeed, much of the reaction is not even government-mandated.
- We could solve a complex dynamic programming problem.
- Instead, Cochrane (2020) has suggested:

$$\mathcal{R}_{0t} = \text{Constant} \cdot d_t^{-\alpha}$$

where d_t is daily deaths per million people.

Estimates and simulations

Parameters assumed fixed and homogeneous, I

- $\gamma = 0.2$: the average length of time a person is infectious is $1/\gamma$, so 5 days in our baseline. We also consider $\gamma = 0.15$ (7 day duration).

- $\theta = 0.1$: 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 that the average case takes a total of 15 days to resolve. The implied exponential distribution includes a long tail capturing that some cases take longer to resolve.

- $\alpha = 0.05$. We estimate α_i from data for each location i. Tremendous heterogeneity across locations in these estimates, so a common value is not well-identified in our data.
  We report results with $\alpha = 0$ and $\alpha = .05$.

Parameters assumed fixed and homogeneous, II

- $\delta = 1.0\%$.
  - Case fatality rates not helpful: no good measure of how many people are infected.
  - Evidence from a large seroepidemiological national survey in Spain: $\delta = 1.0\%$ in Spain is between 1% and 1.1%. Because many of the early deaths in the epidemic were linked with mismanagement of care at nursing homes in Madrid and Barcelona, we pick 1% as our benchmark value.
  - Correction by demographics to other countries. For most of the countries, mortality rate clusters around 1%. For the U.S.: 0.76% without correcting for life expectancy and 1.05% correcting by it.
  - Other studies suggest similar values of δ. New York City data suggests death rates of around 0.8%-1%.
  - CDC has release a lower estimate (0.26%). I just do not see it.

Estimation based entirely on death data

- Johns Hopkins University CSSE data plus a few extra sources for regions/cities.
- Excess death issue:
  - New York City added 3,000+ deaths on April 15 $\approx$ 45% more.
  - *The Economist* and *NYT* increases based on vital records.
  - Example: Spain, where we have a national civil registry: 43,034 excess deaths vs. 27,117 at CSSE (18%).
  - We adjust all NYC deaths before April 15 by this 45% and non-NYC deaths upwards by 33%.
- We use an HP filter to death data.
  - Otherwise, very serious "weekend effects" in which deaths are underreported.
  - Even zero sometimes, followed by a large spike.

Spain

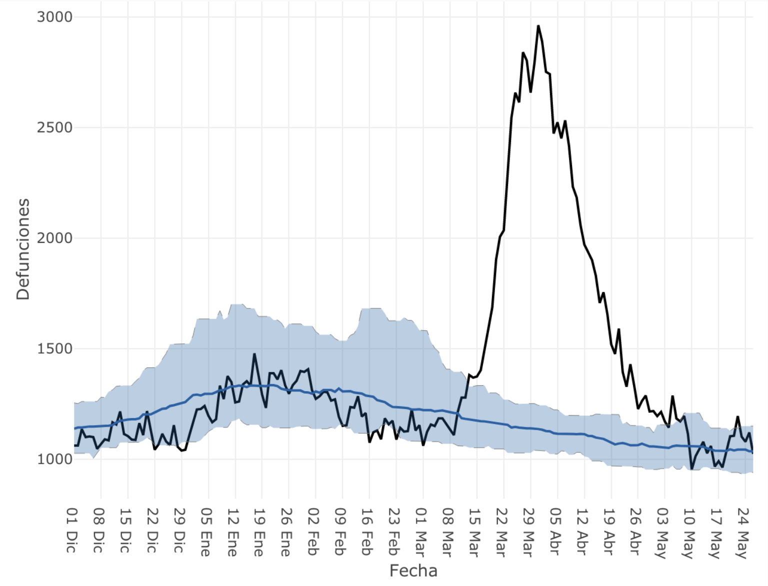
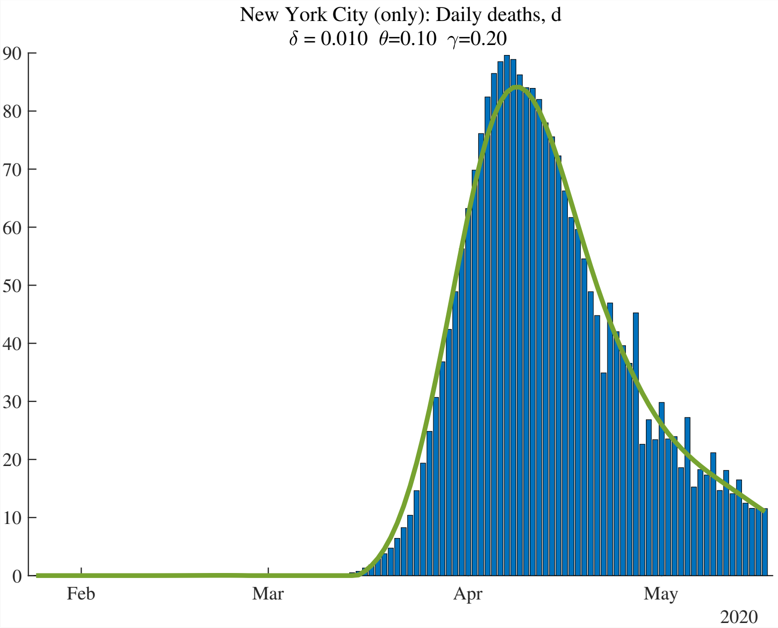

3000
2500
2000
1500
1000
Defunciones
01 Dic
08 Dic
15 Dic
22 Dic
29 Dic
05 Ene
12 Ene
19 Ene
26 Ene
02 Feb
09 Feb
16 Feb
23 Feb
01 Mar
08 Mar
15 Mar
22 Mar
29 Mar
05 Abr
12 Abr
19 Abr
26 Abr
03 May
10 May
17 May
24 May
Fecha

Figure 2: New York City: Daily Deaths and HP Filter



Guide to graphs

- 7 days of forecasts: Rainbow color order!

  ROY-G-BIV (old to new, low to high)

  - Black = current.
  - Red = oldest, Orange = second oldest, Yellow =third oldest....
  - Violet (purple) = one day earlier.

- For robustness graphs, same idea:

  - Black = baseline (e.g. $\delta = 0.8\%$).
  - Red = lowest parameter value (e.g. $\delta = 0.8\%$).
  - Green = highest parameter value (e.g. $\delta = 1.2\%$).

Figure 1: New York City: Estimates of $\mathcal{R}_{0t} = \beta_t/\gamma$

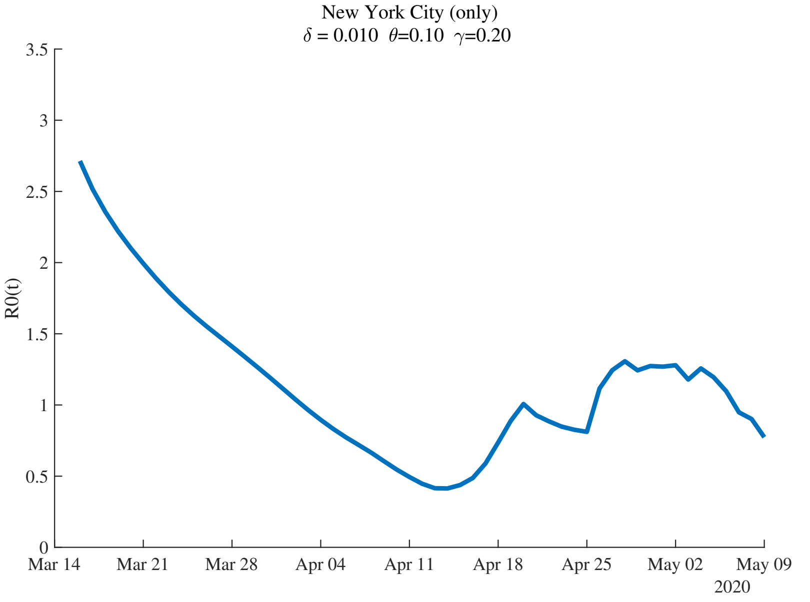


Figure 5: New York City: Percent of the Population Currently Infectious

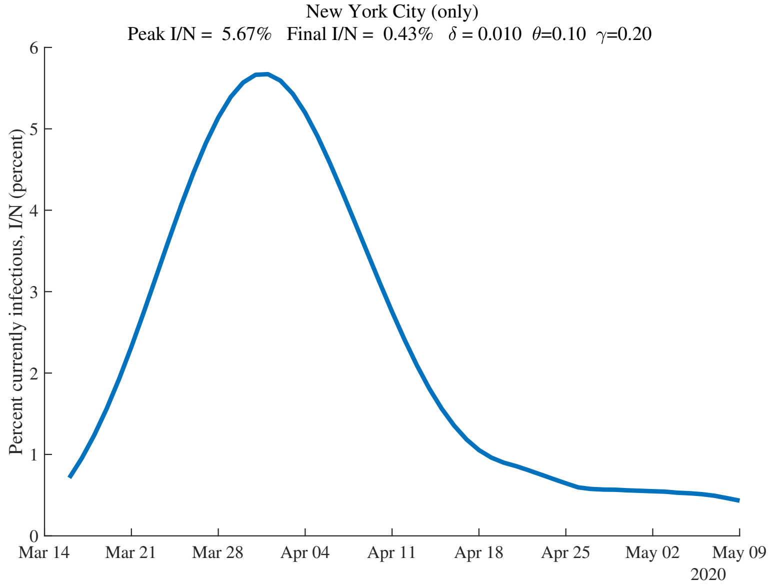


Figure 7: Percent of the Population Currently Infectious

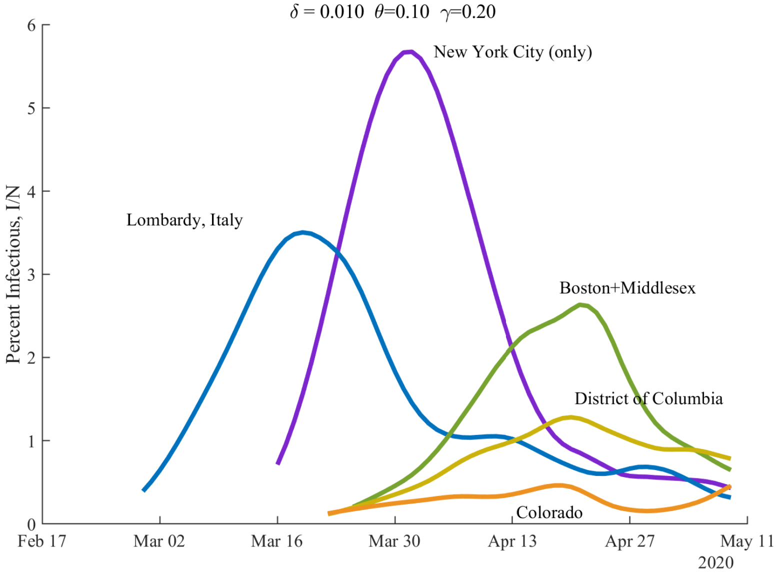


Figure 9: New York City: Daily Deaths per Million People ($\delta = 1.0\%/0.8\%/1.2\%$)

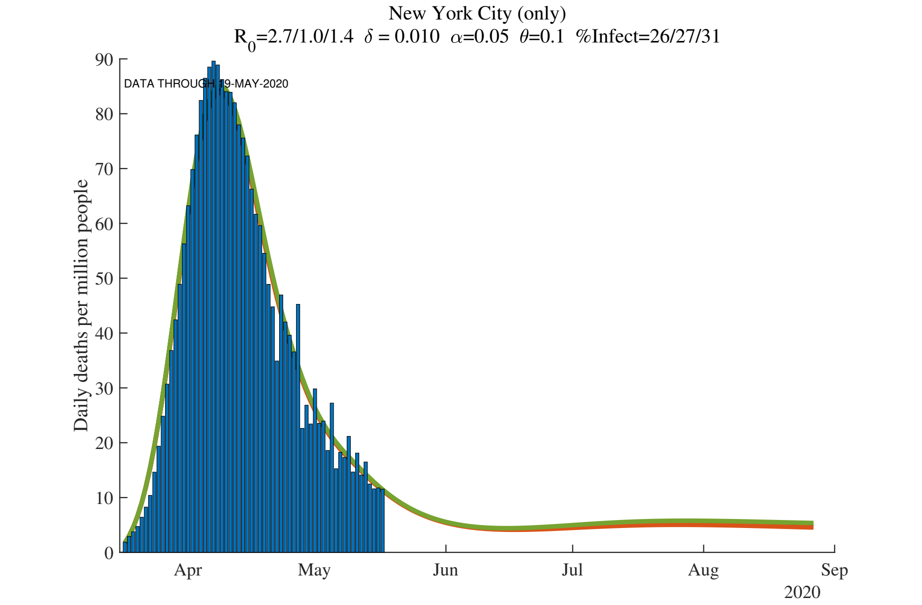


Figure 11: Spain: Cumulative Deaths per Million People ($\gamma = .2/.1$)

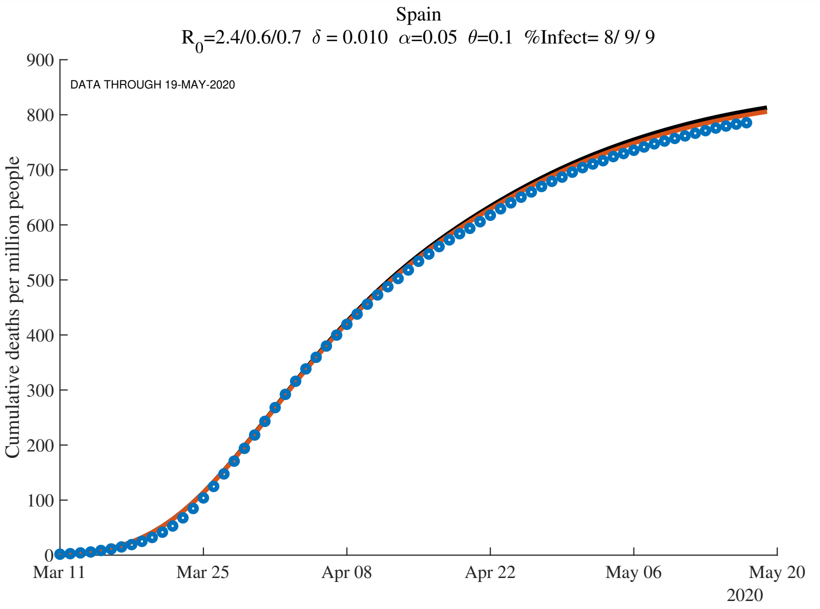


Figure 13: Spain: Cumulative Deaths per Million (Future, $\gamma = .2/.1$)

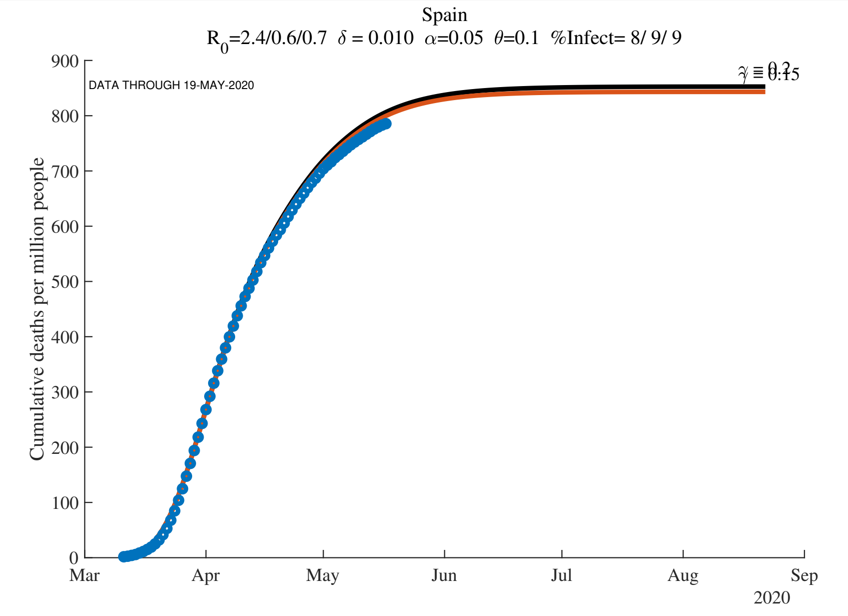


Figure 15: Italy: Cumulative Deaths per Million (Future, $\theta = .1/.07/.2$)

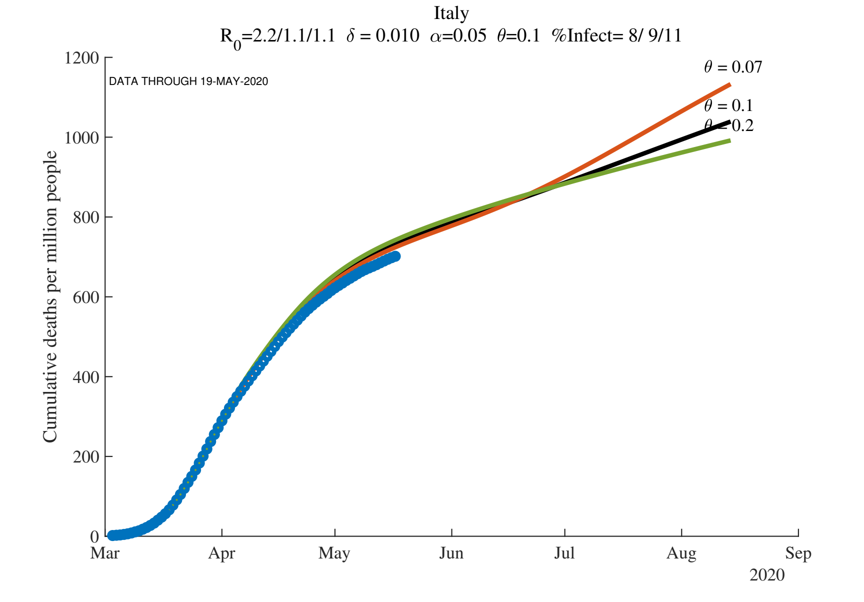


Figure 17: Lombardy, Italy (7 days): Daily Deaths per Million People

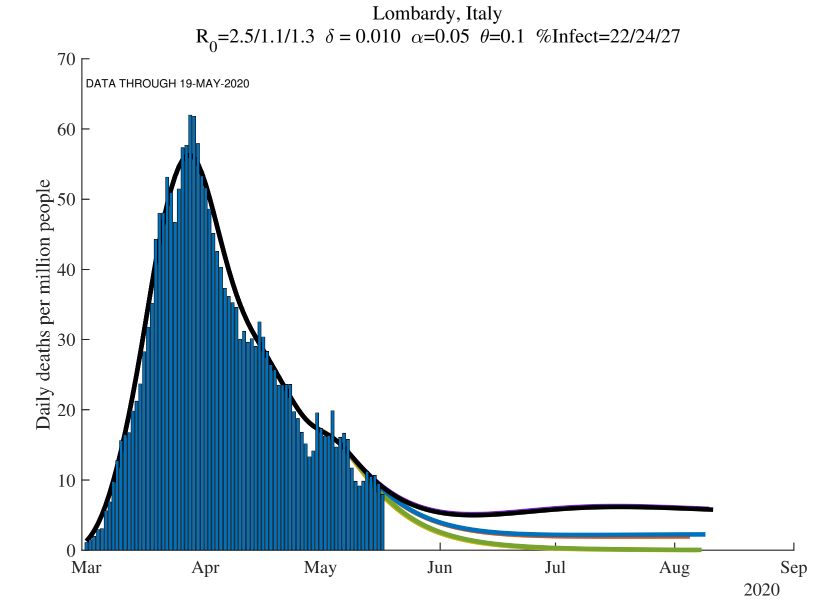


Figure 19: New York City (7 days): Cumulative Deaths per Million (Future)

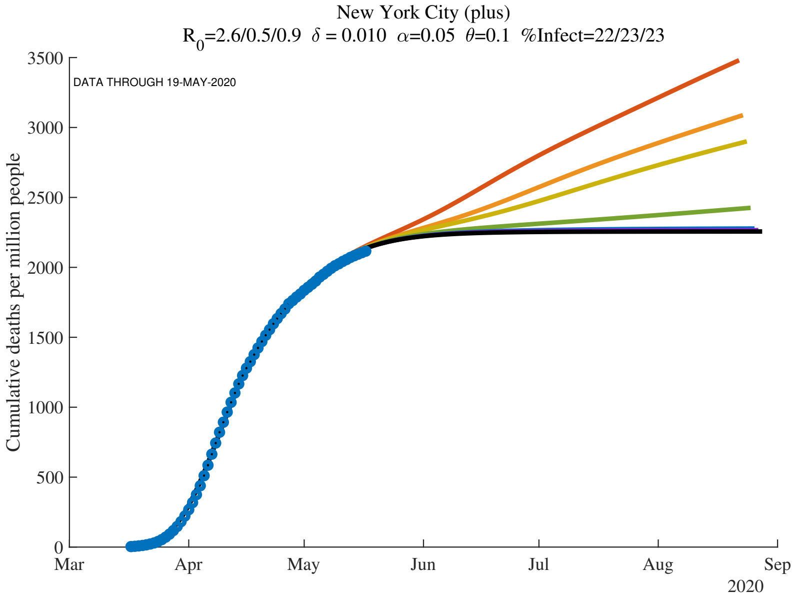


Figure 21: California (7 days): Daily Deaths per Million People

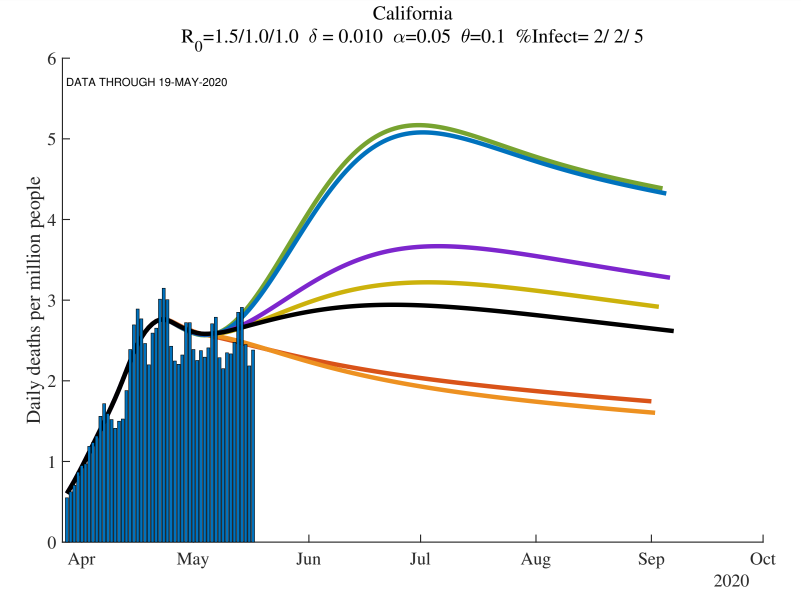
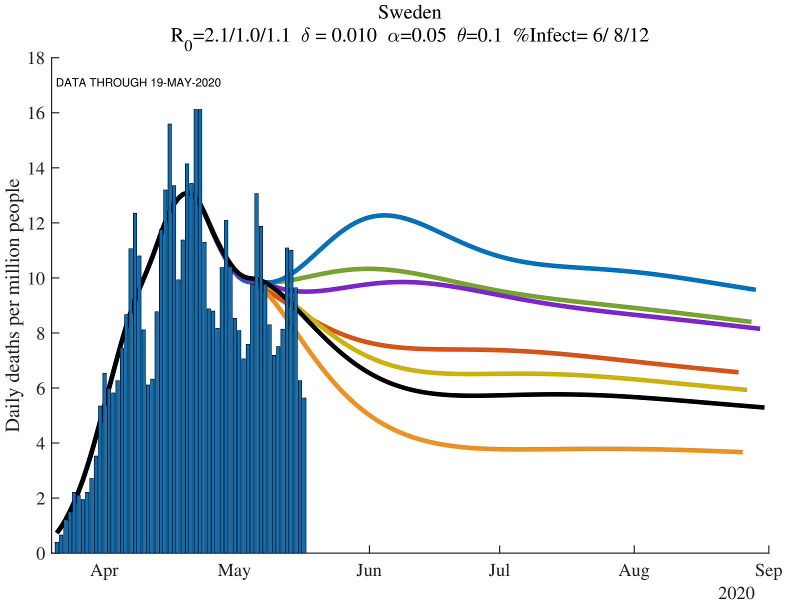


Figure 30: Sweden (7 days): Daily Deaths per Million People



Reopening and herd immunity

Reopening and herd immunity

- The disease will die out as long as:

$$\mathcal{R}_{0t} \cdot S_t / N < 1$$

- That is, if the "new" $\mathcal{R}_{0t}$ is smaller than $1/s(t)$.
- Today's infected people infect fewer than 1 person on average.
- We can relax social distancing to raise $\mathcal{R}_{0t}$ to $1/s(t)$.
- Note, however, the importance of "momentum."

Why random testing is so valuable

	— Percent Ever Infected (today) —		
	$\delta = 0.5\%$	$\delta = 1.0\%$	$\delta = 1.2\%$
New York City (only)	51	26	22
Lombardy, Italy	43	22	19
New York	31	16	13
Madrid, Spain	36	18	15
Detroit	36	18	15
New Jersey	37	19	16
Stockholm, Sweden	36	18	15
Connecticut	33	17	14
Boston+Middlesex	29	15	12
Massachusetts	29	15	12
Paris, France	21	11	9
Philadelphia	23	12	10
Michigan	18	9	8
Spain	17	8	7
Italy	15	8	7
Illinois	13	7	6
Sweden	12	6	5
Pennsylvania	12	6	5
United States	9	5	4
New York excluding NYC	8	4	3
Los Angeles	5	3	2
Florida	3	2	1
California	3	2	1

Using percent susceptible to estimate herd immunity, $\delta = 1.0\%$

	$\mathcal{R}_0$	$\mathcal{R}_{0t}$	Percent Susceptible t+30	$\mathcal{R}_{0t+30}$ with no outbreak	Percent way back to normal
New York City (only)	2.7	0.8	73.5	1.4	30.3
Lombardy, Italy	2.5	0.9	77.5	1.3	23.4
New York	2.6	0.7	83.8	1.2	26.4
Madrid, Spain	2.6	0.2	81.5	1.2	43.2
Detroit	2.4	0.5	81.6	1.2	37.6
New Jersey	2.6	1.1	78.3	1.3	11.4
Stockholm, Sweden	2.6	1.2	78.3	1.3	7.2
Boston+Middlesex	2.1	0.7	84.9	1.2	32.9
Massachusetts	2.1	1.0	83.3	1.2	21.3
Paris, France	2.4	0.2	89.4	1.1	42.0
Philadelphia	2.5	0.9	87.2	1.1	17.0
Spain	2.4	0.5	91.5	1.1	29.8
Chicago	2.2	0.9	87.0	1.1	18.0
Illinois	2.0	0.9	91.2	1.1	15.3
Sweden	2.1	0.9	92.7	1.1	15.2
Pennsylvania	2.1	0.8	93.0	1.1	19.5
United States	2.0	0.9	94.7	1.1	13.1
New York excluding NYC	2.0	1.1	92.8	1.1	-2.3
Los Angeles	1.6	1.0	96.2	1.0	5.4
Florida	1.6	0.9	98.0	1.0	15.3
California	1.5	1.0	97.5	1.0	-3.4

Simulations of re-opening

- Begin with the basic estimates shown already.
- Different policies are then adopted starting around May 20.
  - Black: assumes $\mathcal{R}_{0t}$(today) remains in place forever.
  - Red: assumes $\mathcal{R}_{0t}$(suppress)$= 1/s(today)$.
  - Green: we move 25% of the way from $\mathcal{R}_{0t} =$ "today" back to initial $\mathcal{R}_{0t} =$ "normal."
  - Purple: we move 50% of the way from $\mathcal{R}_{0t} =$ "today" back to initial $\mathcal{R}_{0t} =$ "normal."
- We assume these $\mathcal{R}_{0t}$ values stay in place forever.
  - In practice, over course, β_t would likely start to fall again as mortality rises.

Figure 35: Spain: Re-opening

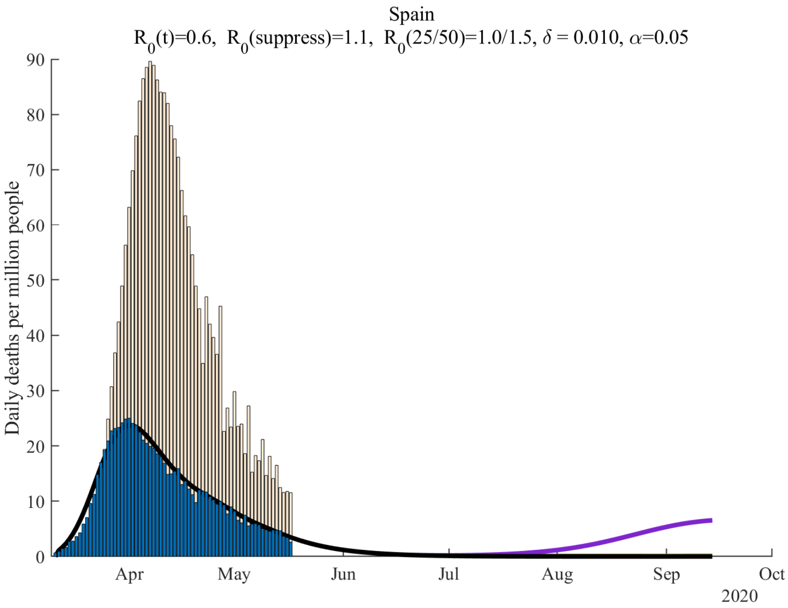


Figure 36: Italy: Re-opening

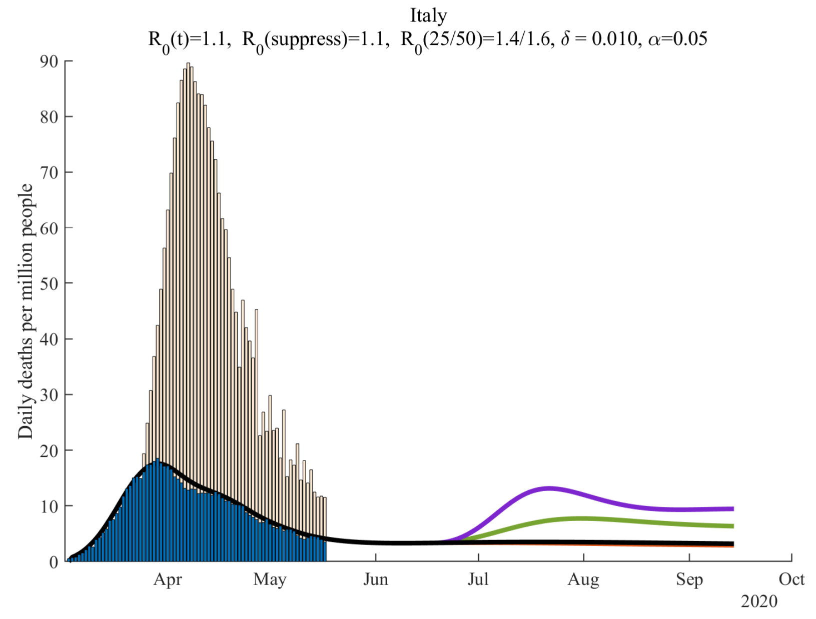


Figure 37: New York City: Re-opening

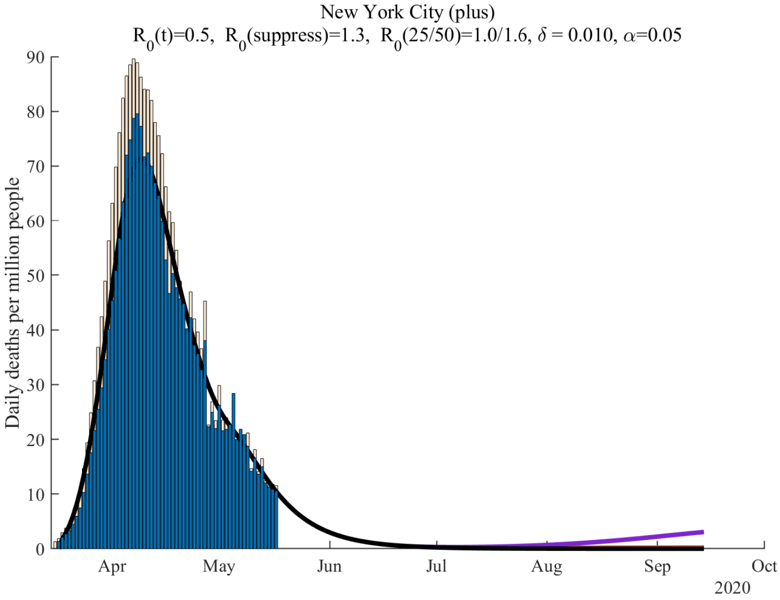


Figure 38: New York excluding NYC: Re-opening

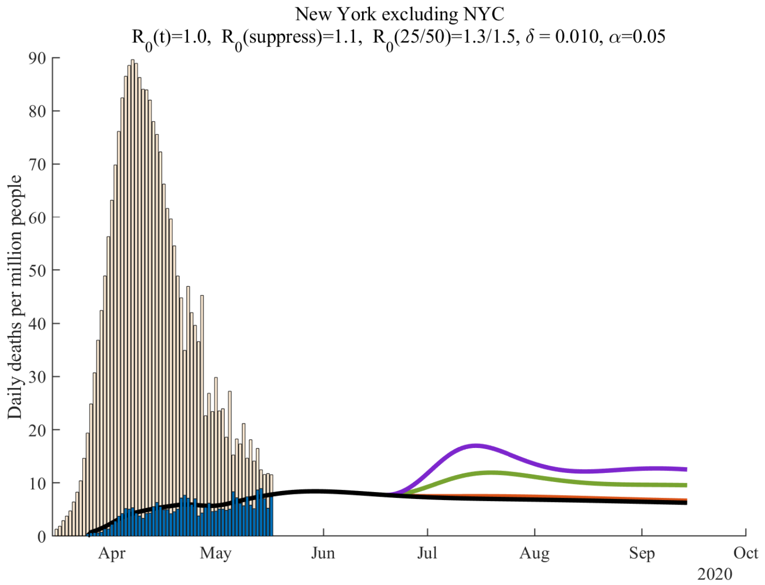


Figure 39: Los Angeles: Re-opening

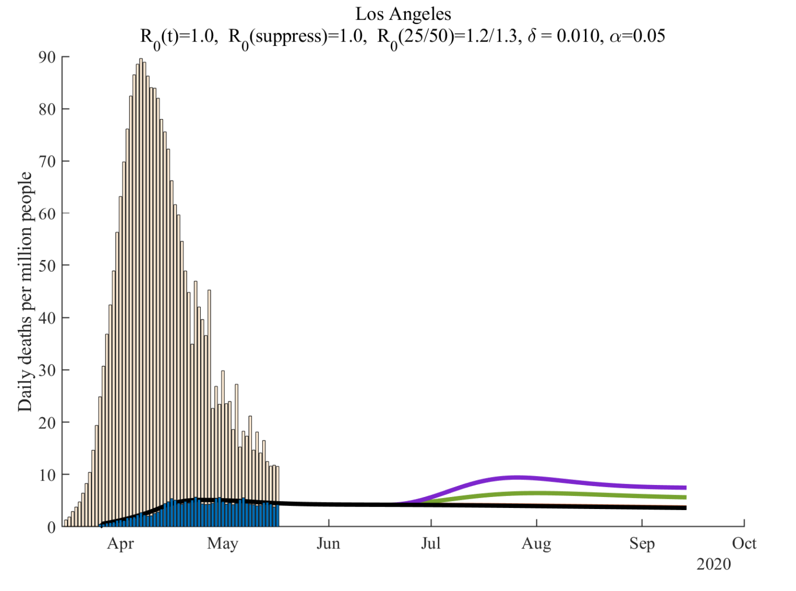


Figure 40: Stockholm, Sweden: Re-opening

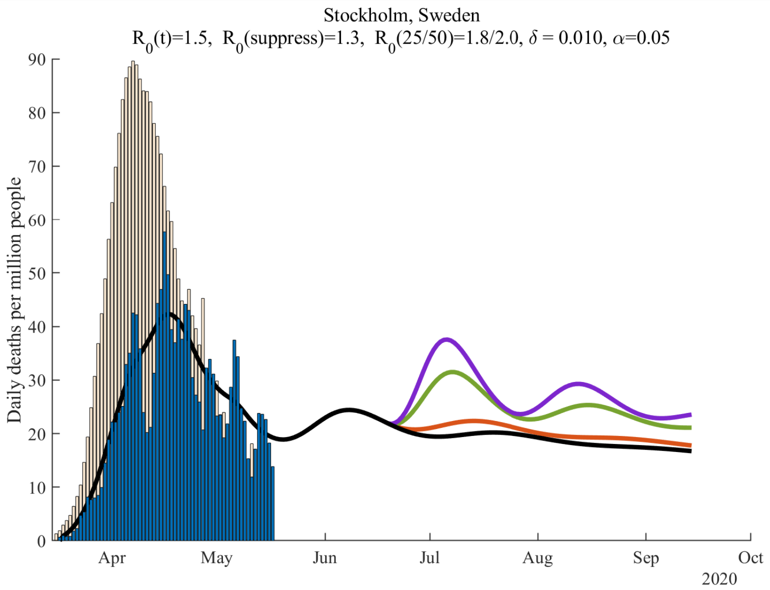


Figure 41: Chicago: Re-opening

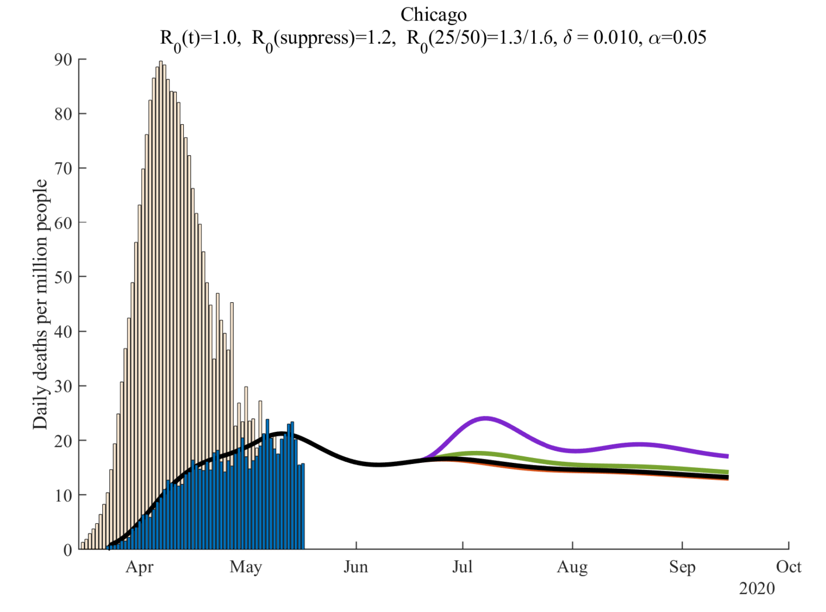
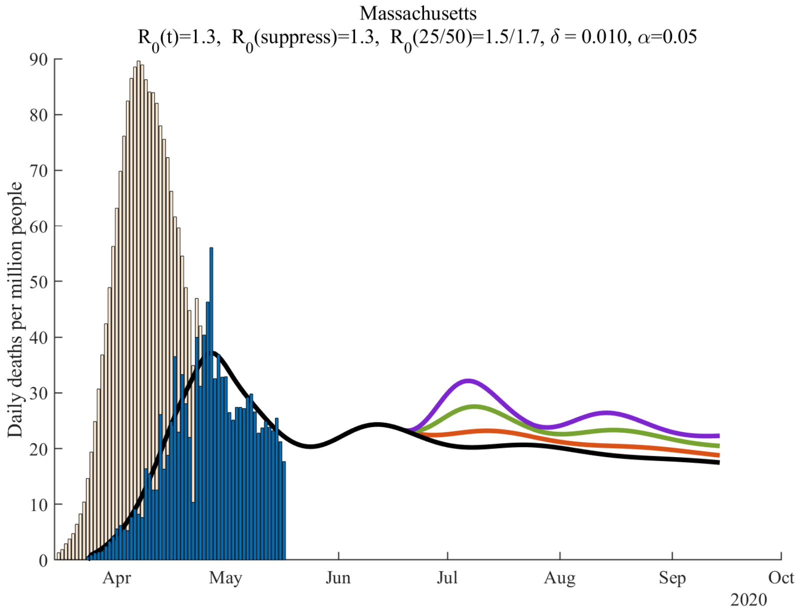


Figure 42: Massachusetts: Re-opening



More progress on β_t

- Can we understand the evolution of β_t (i.e., initial and final level, rate of decay)?
- This might help us to forecast its evolution.
- Also, it might help us map changes of β_t into concrete policies.
- Two points:
  1. Agents react endogenously to information: Cochrane (2020), Farboodi, Jarosch, and Shimer (2020), and Toxvaerd (2020).
  2. Economists like to think at the margin.

More progress on β_t (continued)

- We looked at:

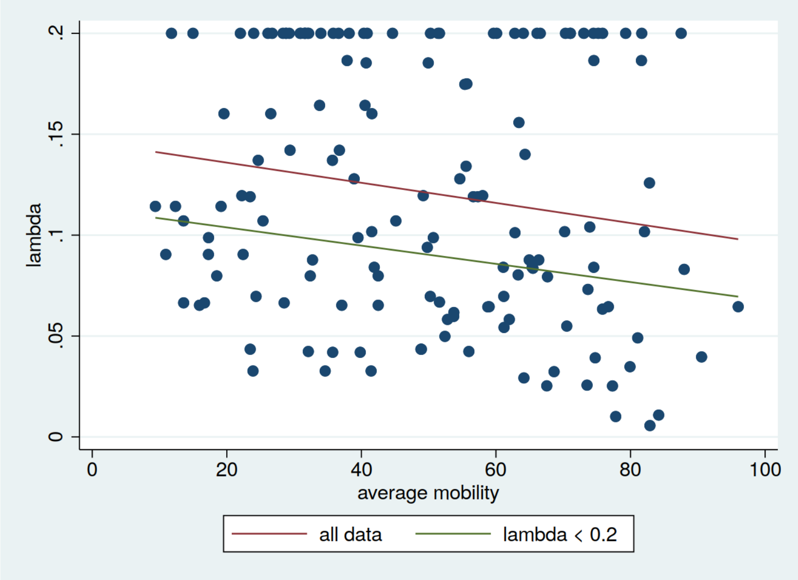
    1. Fraction of housing units located in an urban environment.
    2. Population density per square kilometer.
    3. Average annual temperature in degrees Celsius.
    4. log real GDP/personal income per capita.

- Urbanization and income are significant, but both marginally and, for income, with a surprising sign.
- We do not take this results are anything but a suggestion there are no obvious patterns there.

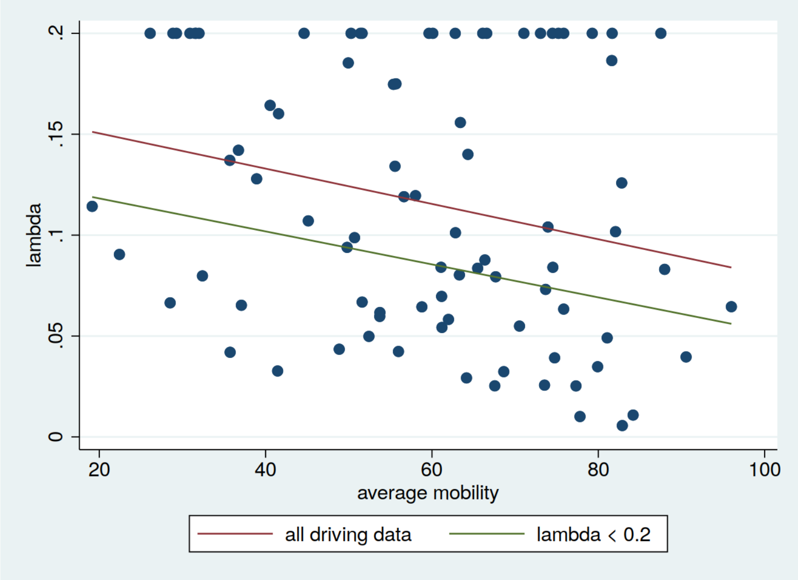
More progress on β_t (continued)

- We map changes of β_t into measures of policies.
- A proxy of the effects of policies: Mobility Trends Reports from Apple Maps.
- However, this proxy mixes voluntary and compulsory reductions in mobility and causality is hard to ascertain.
- Significant correlation between λ and reductions in average mobility (with and without additional controls).
- Correlation triggered by driving. Walking and mass transit per se are not significant.

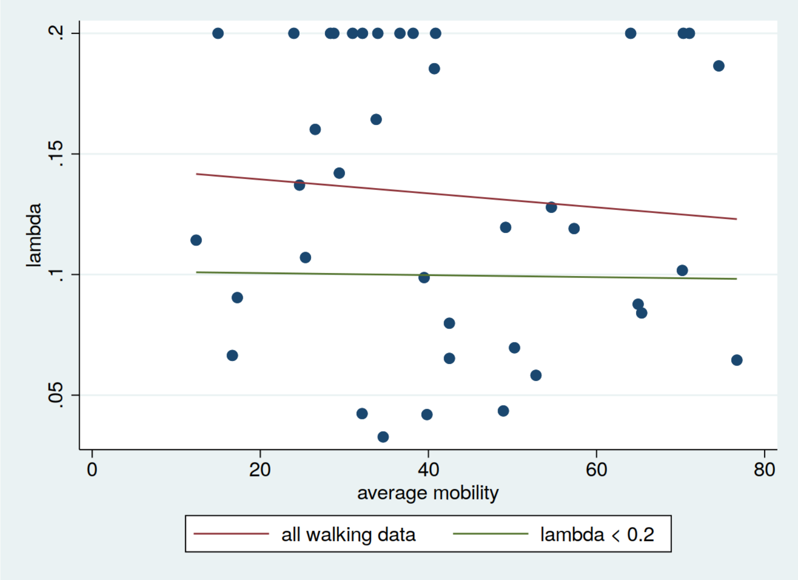
λ vs. mobility



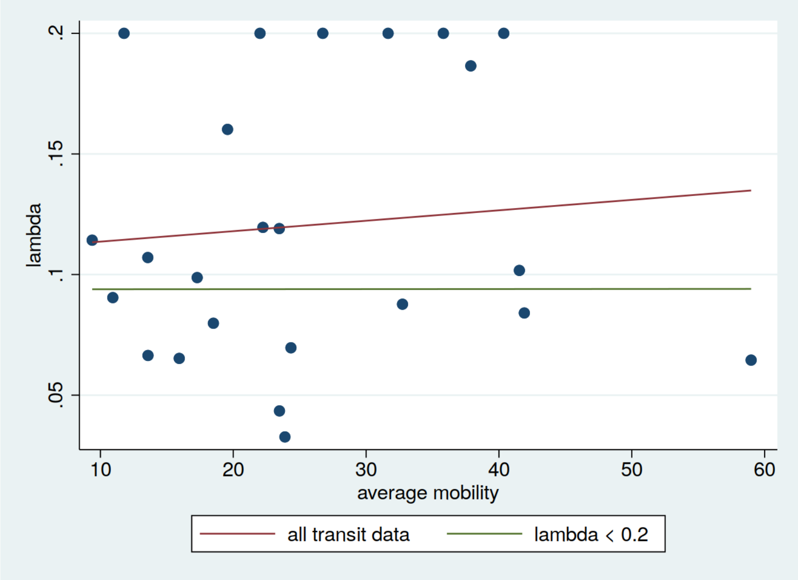
λ vs. mobility



λ vs. mobility



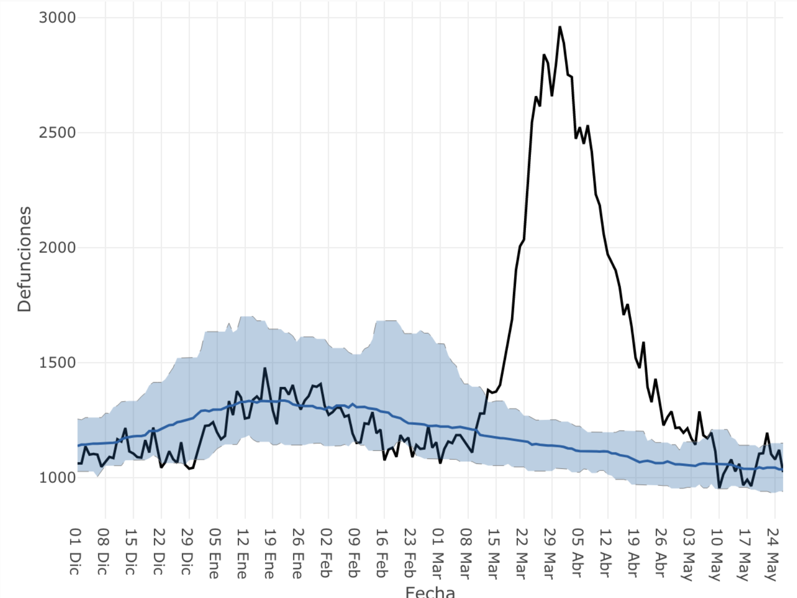
λ vs. mobility



Heterogeneity

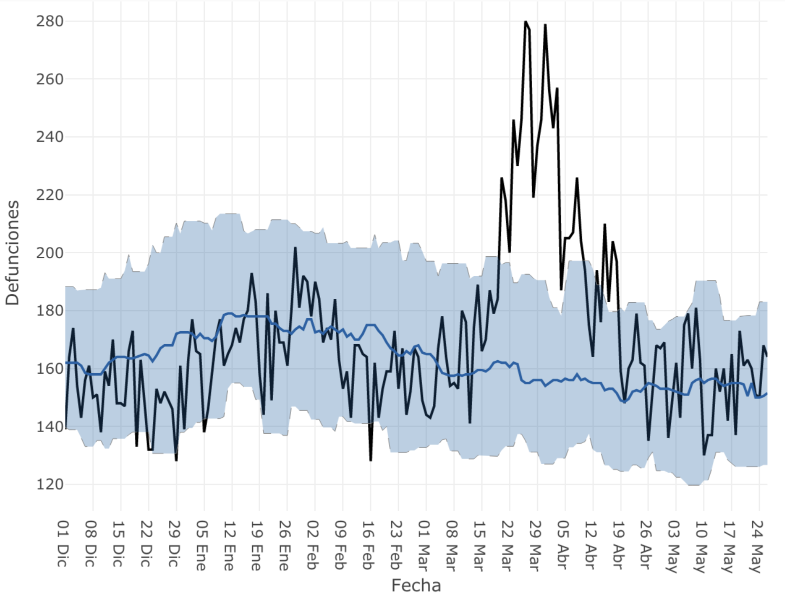
- We know heterogeneity is key, for instance, for mortality (age, pre-existing conditions).
- Also, for patterns of behavior and social contact.
- Role of super-spreaders and nursing homes.
- Introduction movement across territories.
- Heterogeneous-agents SIRD model. Among many others, Acemoglu et al. (2020) and Berger, Herkenhoff, and Mongey (2020).

Spain (all)

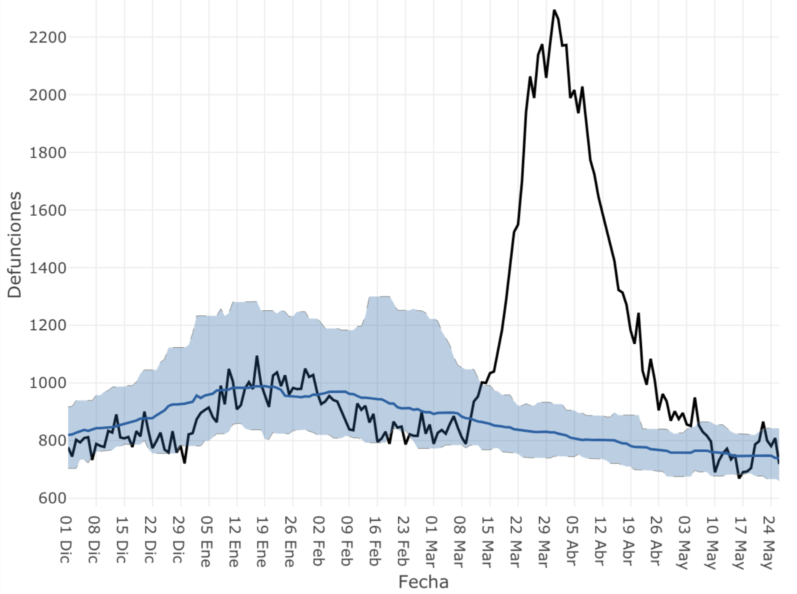


3000
2500
2000
1500
1000
Defunciones
01 Dic
08 Dic
15 Dic
22 Dic
29 Dic
05 Ene
12 Ene
19 Ene
26 Ene
02 Feb
09 Feb
16 Feb
23 Feb
01 Mar
08 Mar
15 Mar
22 Mar
29 Mar
05 Abr
12 Abr
19 Abr
26 Abr
03 May
10 May
17 May
24 May
Fecha

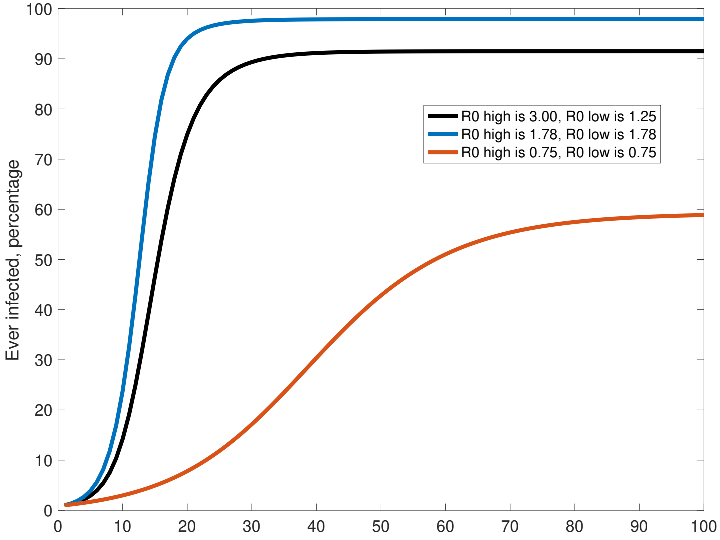
Spain (under 65)



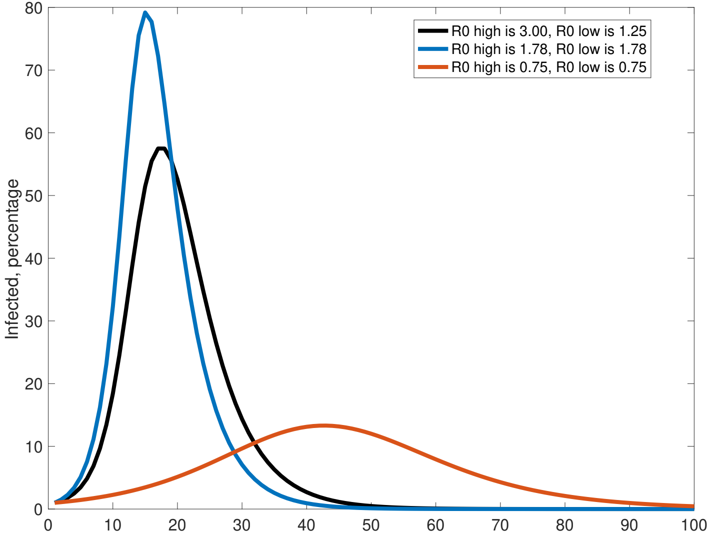
Spain (over 75)



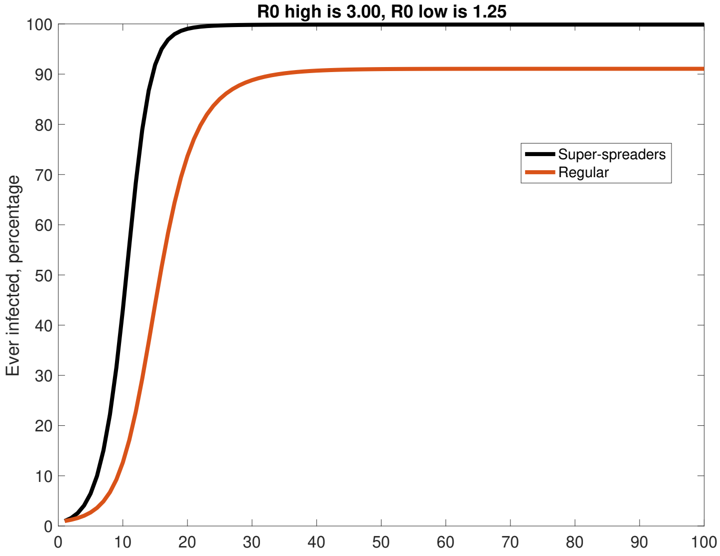
Super-spreaders



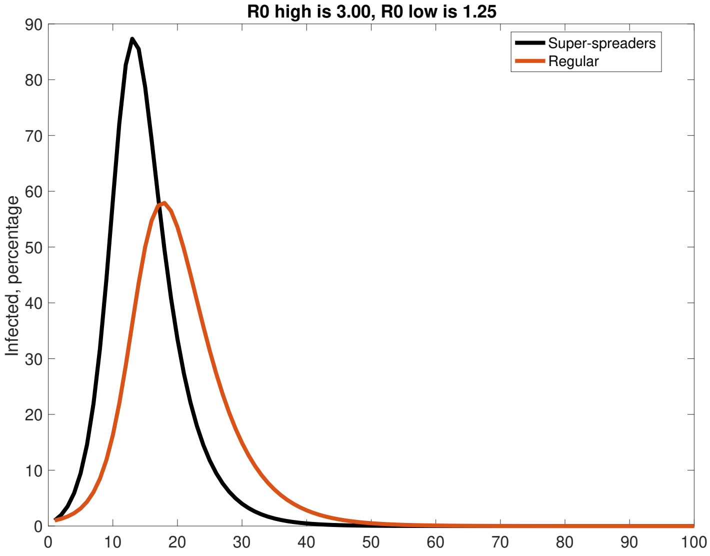
Super-spreaders (continued)



Super-spreaders (continued)



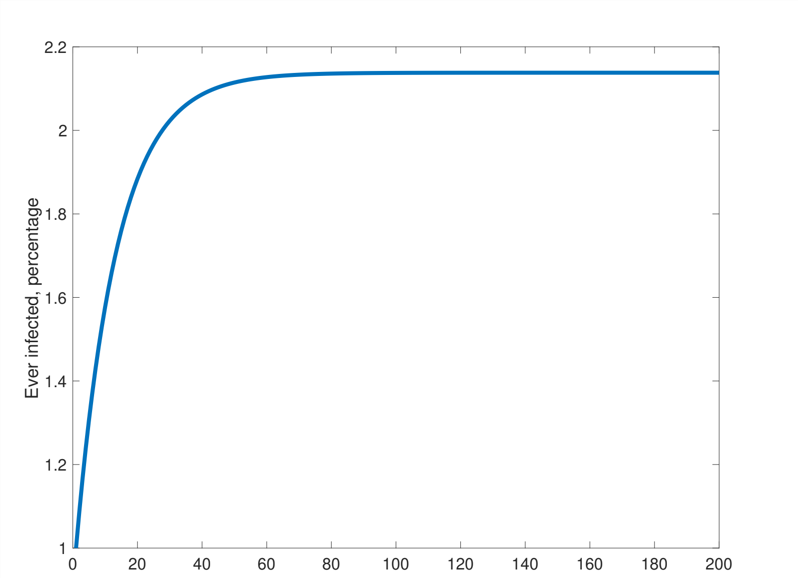
Super-spreaders (continued)



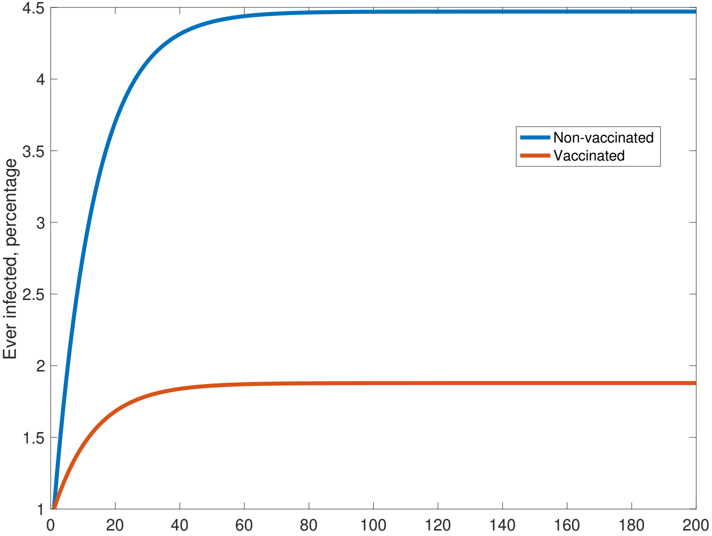
Other exercises

- We can take the results that we get from our estimation and undertake policy exercises.
- Take, for instance, the point estimates for NYC, including a $R_0 = 4.1$.
- We model a vaccine.
- Success rate of the vaccine: 75% of vaccinated do not get infected and, of the 25% who do get infected, only 25% can transmit it (relatively conservative assumption given the clinical success of other vaccines; recall: COVID-19 has a very different structure than the Influenza virus).
- 90% vaccination rate.
- You can effectively control the epidemic.

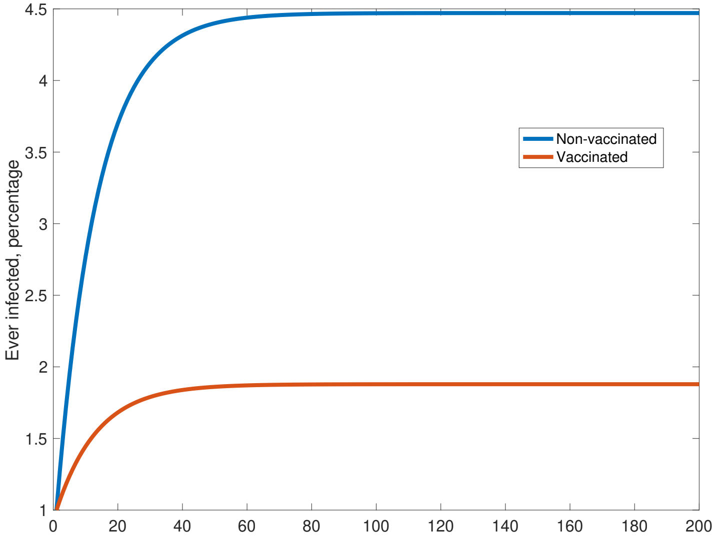
Ever infected



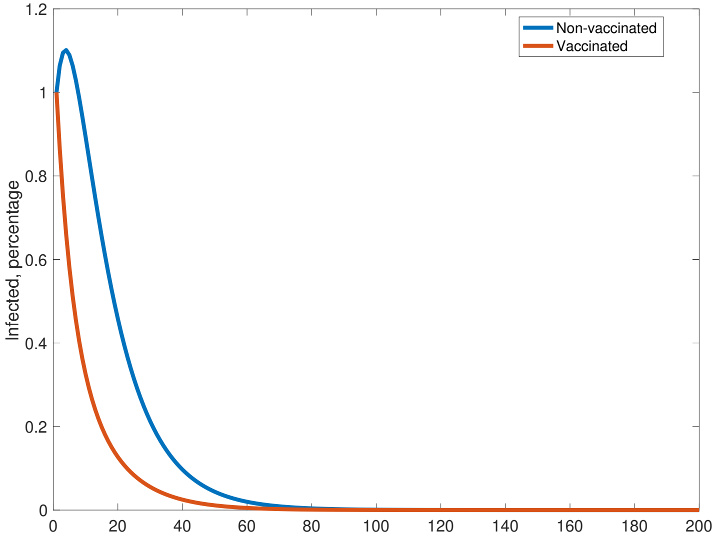
Ever infected (continued)



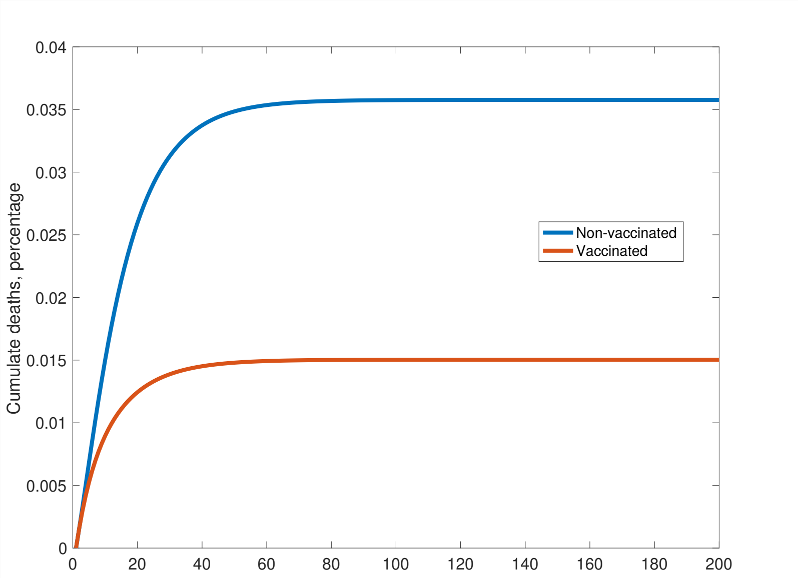
Currently infected



Currently infected (continued)



Cumulate deaths



Conclusions

- Time-varying β (or $\mathcal{R}_{0t}$) needed to capture social distancing, by individuals or via policy.
- Is the death rate 0.8% or 1.0% or ??? Random sampling!
- "One size fits all" will not work for re-opening.
- Susceptible rates are heterogeneous.
- We can employ rich models for policy analysis.
- But one needs to be cautious.