FOUR STYLIZED FACTS ABOUT COVID-19

Andrew Atkeson
Karen Kopecky
Tao Zha

Four Stylized Facts about COVID-19
Andrew Atkeson, Karen Kopecky, and Tao Zha
NBER Working Paper No. 27719
August 2020
JEL No. C01,C02,C11,I1

ABSTRACT

We document four facts about the COVID-19 pandemic worldwide relevant for those studying the impact of non-pharmaceutical interventions (NPIs) on COVID-19 transmission. First: across all countries and U.S. states that we study, the growth rates of daily deaths from COVID-19 fell from a wide range of initially high levels to levels close to zero within 20-30 days after each region experienced 25 cumulative deaths. Second: after this initial period, growth rates of daily deaths have hovered around zero or below everywhere in the world. Third: the cross section standard deviation of growth rates of daily deaths across locations fell very rapidly in the first 10 days of the epidemic and has remained at a relatively low level since then. Fourth: when interpreted through a range of epidemiological models, these first three facts about the growth rate of COVID deaths imply that both the effective reproduction numbers and transmission rates of COVID-19 fell from widely dispersed initial levels and the effective reproduction number has hovered around one after the first 30 days of the epidemic virtually everywhere in the world. We argue that failing to account for these four stylized facts may result in overstating the importance of policy mandated NPIs for shaping the progression of this deadly pandemic.

Andrew Atkeson
Bunche Hall 9381
Department of Economics
UCLA
Box 951477
Los Angeles, CA 90095-1477
and NBER
andy@atkeson.net

Karen Kopecky
Research Department
Federal Reserve Bank of Atlanta
1000 Peachtree ST NE
atlanta, ga 30309
karen.kopecky@atl.frb.org

Tao Zha
Department of Economics
Emory University
Rich Memorial Building
1602 Fishburne Drive
Atlanta, GA 30322-2240
and Federal Reserve Bank of Atlanta
and also NBER
tzha@emory.edu

1 Introduction

In the month of March 2020, in many of the hotspots for COVID-19 around the world, daily deaths from the disease grew very fast — doubling every 2-3 days in the most severely impacted locations. In contrast, in other locations, early on in this pandemic, daily deaths from the disease grew much more slowly. This pattern of high and highly dispersed growth rates of daily deaths from COVID-19 ended very rapidly. Relatively slow growth or even shrinkage of daily deaths from the disease was observed in every location that we study 20-30 days after that location first experienced 25 cumulative deaths, and the dispersion in growth rates of daily deaths across locations fell even more rapidly. Now, in late July 2020, while different locations suffer widely different levels of daily deaths from COVID-19, relatively slow growth or even shrinkage of daily deaths from the disease continues to be observed worldwide and the dispersion in growth rates of daily deaths across locations remains at a low level.

In this paper, we document these facts regarding COVID deaths using both simple data smoothing procedures and a Bayesian estimation procedure that allows us to construct probability bands around our estimates of the growth of COVID deaths. We then use an SIR epidemiological model based on that in Kermack and McKendrick (1927) to interpret these data on the growth rate of COVID deaths as indicating that the effective reproduction number of COVID-19, here denoted by $\mathcal{R}(t)$, has fallen from initial values of 3 or more in many locations to levels close to one or even lower everywhere in the world. We invert a calibrated version of our simple SIR model to argue that the vast majority of the observed decline in the level and cross sectional variation of the effective reproduction number has been due to a decline in the transmission rate of the disease and not due to the progression of the population towards herd immunity.[1] We also report results using more complex epidemiological

---

[1]The *transmission rate* of the disease is defined as the average rate at which an infected individual contacts others and spreads the virus to those contacted. The *effective reproduction number* $\mathcal{R}(t)$ is defined as the average number of secondary cases infected by a single infected individual during his or her period of infectiousness when there are both susceptible and immune people among

models including one similar to that considered by Fernandez-Villaverde and Jones (2020) and show that our conclusions about the worldwide decline in the transmission rate for COVID-19 are not much affected by the choice of epidemiological model.[2]

This finding of a rapid decline in the transmission of COVID-19 within the first 30 days of the progression of the pandemic in widely heterogeneous countries worldwide and the persistence of these low transmission rates over the past few months has important implications for studies of the impact of policy interventions on the progression of this epidemic. Several prominent studies, including Dehning et al. (2020), Hsiang et al. (2020), and Flaxman et al. (2020), have studied empirically the role of government-mandated non-pharmaceutical interventions (NPIs) in reducing the transmission of COVID-19, and many of these studies argue that these NPI's had a large impact on the transmission rate of the disease in the early phase of the pandemic.[3] Given the observation that transmission rates for COVID-19 fell virtually everywhere in the world during this early pandemic period, we are concerned that these studies may substantially overstate the role of government-mandated NPI's in reducing disease transmission due to an omitted variable bias. Moreover, given the observation that disease transmission rates have remained low with relatively low dispersion across locations worldwide for the past several months as NPI's have been lifted, we are concerned that estimates of the effectiveness of NPI's in reducing disease transmission from the earlier period may not be relevant for forecasting the impact of the relaxation of those NPI's in the current period, due to some unobserved switch in regime.

What might this omitted variable or variables be? While we cannot answer this

---

those contacted by this infected individual. Note that since a person who is no longer susceptible to the disease does not become infected even if he or she takes in the virus from contact with an infected individual, the effective reproduction number can vary due to changes in the transmission rate and changes in the portion of susceptible people among those contacted by infectious individuals.

[2]See also Okell et al. (2020) for a discussion of whether European countries experienced slowing growth rates of deaths due to herd immunity.

[3]See also Allcott et al. (2020), Fang et al. (2020), Deb et al. (2020), Askitas et al. (2020), Bonardi et al. (2020), Weber (2020), Conyon et al. (2020), Chernozhukov et al. (2020), Wieland (2020), Rubin et al. (2020), Donohue and Miller (2020) and many others.

question ourselves, the literatures in both epidemiology and economics offer several candidates. COVID-19 is not the first epidemic for which transmission rates have fallen faster than would be predicted by simple epidemiological models. Chowell et al. (2016) and Eksin et al. (2019) are two of many studies indicating that this rapid drop in transmission rates is a pervasive feature of human epidemics. These studies and many others point to two hypotheses as to why transmission rates for human epidemics might fall rapidly that we argue deserve further study when applied to COVID-19.

The first of these hypotheses is that humans spontaneously take action to avoid disease transmission once an epidemic breaks out.[4] For COVID-19, a great deal of real-time mobility and economic data indicate that human social and economic interactions have fallen substantially across a large number of locations.[5] Further research is required, however, to determine whether this decline in human interactions is sufficiently large and widespread to account for the apparent decline in transmission rates for COVID-19 worldwide.[6]

The second of these hypothesis is that the network structure of human interactions naturally leads to a slowdown in disease transmission faster than would be predicted from a simple epidemiological model in which the population interacts uniformly with each other.[7] Further research is required to determine whether a network structure of

---

[4]For expositions of this hypothesis, see Eksin et al. (2019), Keppo et al. (2020) (slides available at https://www.lonessmith.com/wp-content/uploads/2020/04/pandemic-slides. pdf), John Cochrane's discussion at https://johnhcochrane.blogspot.com/2020/05/ an-sir-model-with-behavior.html, Farboodi et al. (2020), Eichenbaum et al. (2020), Guerrieri et al. (2020), Kaplan et al. (2020), Toxvaerd (2020), Phillipson and Posner (1993) and many others.

[5]See, for example, Cronin and Evans (2020), Goolsbee and Syverson (2020), Fetzer et al. (2020), Gupta et al. (2020).

[6]The data in Lecocq et al. (2020) indicate that the decline in human mobility has indeed been global.

[7]For expositions of this hypothesis, see Chowell et al. (2016), Ellison (2020), Bisin and Moro (2020), and Britton et al. (2020). For an application of this hypothesis to the spread of COVID-19 in cities, see Akbarpour et al. (2020) and the associated website at https:// reopenmappingproject.com/ as well as Fajgelbaum et al. (2020) and Glaeser et al. (2020). For an application of this hypothesis to the worldwide spread of COVID-19, see Siwiak et al. (2020) and https://www.gleamproject.org/.

human interactions can indeed explain the global decline in transmission of COVID-19.

Finally, we must consider the possibility that unexplained natural forces might account for the observed decline in transmission rates for COVID-19. As discussed in Moore et al. (2020), the transmission of COVID-19 bears resemblance to that for pandemic influenza. These authors observe that of eight major influenza pandemics that have occurred since the early 1700's (including the Spanish Flu of 1918-19), seven had an early peak that disappeared over the course of a few months without significant human intervention. Unfortunately, each of those seven had a second substantial peak approximately six months after the first. This final hypothesis suggests that some unobserved natural factor may have driven the decline to date in the transmission of COVID-19. Clearly, the existence of such an unobserved factor would complicate empirical studies of the causal driving forces behind COVID-19 transmission.

In this paper, we characterize the empirical regularities in the regional data on deaths for COVID-19 with a focus on countries and states of the United States that have experienced more than 1000 cumulative deaths from the disease late July 2020. This gives us a list of 23 countries and 25 U.S. states.

The data on daily deaths are quite noisy, with substantial day-of-the-week effects, and large jumps due to revisions in the criteria used to attribute COVID-19 as the cause of death. We consider two procedures for estimating deaths from these noisy data.[8] The first uses a simple seven-day moving average of daily deaths and a finite difference to estimate the growth rate of daily deaths.[9] In our second procedure, we fit a mixture of Weibull distributions to the daily deaths data in each region using a Bayesian estimation procedure described here and in greater detail in Atkeson et al.

---

[8]See Baqaee et al. (Forthcoming) for a third procedure for smoothing the data on daily deaths.

[9]One can plot a seven-day moving average of daily deaths on a logarithmic scale (useful for measuring the growth rate of this statistic) for a large number of countries and U.S. states here https://ig.ft.com/coronavirus-chart. One can readily identify virus hotspots early on in the epidemic as steep curves in these charts. One can also readily see that, more recently, there are no regions with such rapid growth in daily deaths anywhere in the world from these charts.

(2020).[10] This second procedure allows us to compute probability bands around our estimates of daily deaths as well as the higher order derivatives of daily deaths that are needed to invert the more complex epidemiological models considered in our appendix.

We use a simple SIR epidemiological model to interpret these observations on deaths due to COVID-19 in terms of the effective reproduction number and transmission rate of the disease. In doing so, we invert the model with a time-varying transmission rate to fit the estimated deaths data exactly.[11] In our simple SIR model, the effective reproduction number is a linear transformation of the growth rate in logarithms of daily deaths, with the model implied effective reproduction number equal to one when this growth rate is equal to zero and the slope of the relationship equal to the average number of days during which an infected individual is infectious to others. With our model inversion procedure, we are also able to recover an estimate of the evolution of the fraction of the population remaining susceptible over time.[12] With this fraction estimated, we show that the decline in the effective reproduction number (and its cross sectional variation) is largely due to a decline in the transmission rate (and a decline in the cross sectional variation in transmission rates).

In our empirical work, we focus on data on deaths for our estimates of the transmission rates for COVID-19. In this regard, we follow others such as Murray (2020), Hay (2020), Korevaar et al. (2020), and Flaxman et al. (2020) who argue that data on COVID related deaths is more accurate than testing data on cases.[13] Related work by Buckman et al. (2020) and Beenstock and Dai (2020) find a large decline in the

---

[10]This procedure bears some similarity to that used by Murray (2020) who use a mixture of error functions to model cumulative COVID deaths.

[11]See also Atkeson (2020) and Atkeson et al. (2020). Baqaee et al. (Forthcoming) use this procedure in their estimation of the effective reproduction number in the United States.

[12]We confirm our model estimates of this fraction by comparing them with findings from widespread serology studies for New York, Connecticut, Louisiana, and Spain.

[13]See also Aspelund et al. (2020) and https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/about-serology-surveillance.html for a discussion of possible measurement errors in confirmed cases data.

effective reproduction number for COVID-19 across countries when estimated using data on confirmed cases.[14] The websites https://rt.live/ and https://covidestim.org provide estimates of the effective reproduction number for U.S. states using data on confirmed cases.[15] These estimates also show a large and widespread decline in the effective reproduction number for the disease for all U.S. states.

The remainder of the paper is organized as follows. In section 2 we present our procedure for interpreting COVID deaths data with an SIR epidemiological model. We then present our estimation procedures in section 3. We present our findings in section 4. Finally, we conclude.

2   Interpreting deaths data with an SIR model

The SIR epidemiological model we use to interpret the data on deaths from COVID-19 is as follows.

The population is set to N. At each moment of time, the population is divided into four categories (states) that sum to the total population. These states are susceptible S, infected I, resistant R, and dead D. Agents that are susceptible are at risk of getting the disease. Agents that are infected are contagious and may pass it on to others through some form of interaction with susceptible agents. Agents that are resistant are not at risk of getting the disease, either because they have immunity built up from a vaccine or from previous experience with this or similar diseases. Likewise, those who have died from the disease are no longer at risk of getting the disease. We normalize the total population $N = 1$, so all results regarding S, I, R and D should be interpreted as fractions of the relevant population.

We use $\mathcal{R}(t)$ to denote the *effective reproduction number* of the disease at date t. This effective reproduction number is the ratio of the rate at which infected agents

---

[14]See also Liu et al. (2020).

[15]The commonly used method of Cori et al. (2013) for estimating reproduction numbers requires accurate data on new cases.

infect susceptible agents to the recovery rate of infected agents from the disease at date t.

The equations of the model can be stated in terms of the effective reproduction number as

$$dS(t)/dt = -\mathcal{R}(t)\gamma I(t) \tag{1}$$

$$dI(t)/dt = (\mathcal{R}(t) - 1)\gamma I(t) \tag{2}$$

$$dR(t)/dt = (1 - \nu)\gamma I(t) \tag{3}$$

$$dD(t)/dt = \nu\gamma I(t) \tag{4}$$

The parameter γ governs the rate at which agents who are infected stop being infectious and hence stop transmitting the disease. We refer to this parameter as the *recovery rate*. This parameter is considered a fixed parameter determined by the biology of the disease. We denote the fatality rate from the disease by ν.

The parameter $\beta(t)$ is the rate at which infected agents spread the virus to others that they encounter at date t. We refer to this parameter as the *transmission rate*. We define the ratio $\beta(t)/\gamma$ to be the *normalized transmission rate*. It is standard to refer to the value of the normalized transmission rate at the start of the pandemic 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 $\mathcal{R}_0 \equiv \beta(0)/\gamma$.

We assume that infected agents interact randomly with other agents in a uniform manner so 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:

$$\mathcal{R}(t) = \frac{\beta(t)}{\gamma} \frac{S(t)}{1 - D(t)}. \tag{5}$$

We see from equation (5) that the effective reproduction number can fall either due to changes in the normalized transmission rate or changes in the fraction of the population remaining susceptible to the disease.

To invert this model to interpret data on deaths note that from (4), we have

$$I(t) = \frac{1}{\nu\gamma} dD(t)/dt. \tag{6}$$

Using (3) and (4) together and the assumption that $R(0) = D(0) = 0$, we have that

$$R(t) = \frac{1 - \nu}{\nu} D(t). \tag{7}$$

Using that the states must sum to one, we have

$$S(t) = 1 - \frac{1}{\nu} D(t) - \frac{1}{\nu\gamma} \frac{dD(t)}{dt}. \tag{8}$$

To obtain the effective reproduction number implied by deaths data, note that from equations (2), (6), and the time derivative of this second equation, we have

$$\mathcal{R}(t) = 1 + \frac{1}{\gamma} \frac{\frac{d^2 D(t)}{dt^2}}{\frac{dD(t)}{dt}} \tag{9}$$

where the last term in this equation can be interpreted as the time derivative of the logarithm of daily deaths.

Note that this equation (9) implies that there is a linear relationship between the growth rate of daily deaths (here measured as the time derivative to the logarithm of daily deaths) and the model-implied effective reproduction number. This reproduction number is equal to one when the growth rate of daily deaths is equal to zero.

The slope of this relationship is given by $1/\gamma$ corresponding to the number of days on average that an infected individual remains infectious. To compute estimates of the effective reproductive number that are consistent with our estimated paths for the growth rate of daily deaths, we set $\gamma = 0.2$. This value implies that if the growth rate of daily deaths is 30 percent initially, the basic reproduction number, the value of the effective number at date $t = 0$, is 2.5. These values are inline with CDC estimates.

Given these equations, one can obtain an estimate of the normalized transmission rate of the disease from equations (5) and (8). Thus, one can use this estimate to determine the extent to which the model-implied effective reproduction number has changed due to changes in the transmission rate versus a reduction in the fraction of the population remaining susceptible to the disease. For this exercise we also need to set a value for ν. We use the CDC's preferred estimate of this parameter and set $\nu = 0.004$.

3   Death Data and Estimation

The data sources for daily deaths are New York Times for U.S. states and Johns Hopkins University for other countries. We select U.S. states and other countries that have experienced more than 1000 cumulative deaths due to COVID-19 by July 22, 2020. For each location i, our estimation period begins at the location specific date when cumulative deaths reached 25 in that location and ends on July 22, 2020. The 24 countries that we study are Argentina, Belgium, Brazil, Canada, Chile, France, Germany, India, Iran, Ireland, Italy, Japan, Mexico, Netherlands, Panama, Peru, Portugal, Russia, Spain, Sweden, Switzerland, Denmark, Turkey, the United Kingdom, and the United States excluding the states listed next. The 25 U.S. states that we study are Alabama, Arizona, California, Colorado, Connecticut, Florida, Georgia, Illinois, Indiana, Louisiana, Maryland, Massachusetts, Michigan, Mississippi, Minnesota, Missouri, New Jersey, New York, North Carolina, Ohio, Pennsylvania,

9

South Carolina, Texas, Virginia, and Washington. The rest of the U.S. is counted as another region.

One problem that we face is that the data on daily deaths due to COVID-19 is noisy. In many regions, there are both substantial day-of-the-week effects and occasional large spikes in reported deaths due to, among other things, changes in the criteria used to classify deaths as being due to COVID-19. To estimate the *trend* growth of daily deaths from these noisy data, we assume that the trend path of daily deaths in each location is given by a mixture of Weibull density functions which we estimate using a Bayesian procedure.[16] We assume that observed daily deaths are the sum of a mixture of Weibull density functions and a residual whose magnitude is regime-specific. The regime-switching residuals allow us to effectively deal with erratic noises in the data. The Bayesian procedure allows us to construct posterior probability bands around the estimate. It also allows us to derive smooth estimates of the first and higher derivatives of daily deaths that we need to recover estimates of the effective reproduction numbers and transmission rates of the disease from several structural SIR models.

Our estimation methodology begins by scaling the cumulative death data in location i so that $D_i(t)/(1 + d_i)D_i(T)$ lies between zero and one, where $D_i(T)$ is the cumulative number of deaths in location i at the end of the estimation period and $d_i > 0$ is a scale parameter to be estimated as described in the appendix. Let $\Delta D_{i,t}^{\text{Data}}$ be the daily measured object for $dD_i(t)/dt$ and denote

$$\Delta \widetilde{D}_{i,t}^{\text{Data}} = \frac{\Delta D_{i,t}^{\text{Data}}}{(1 + d_i)D_{i,T}^{\text{Data}}}.$$

Given the daily death data and the value of d_i, we run a non-linear regression with

---

[16]Due to its flexibility, the Weibull distribution is commonly used in survival and failure analysis to model the time until an event such as death or mechanical failure. See, for instance, Bučar et al. (2004), Doksum and Lindqvist (2003), Pinder III et al. (1978), and Tsionas (2002).

a mixture of Weibull density functions and regime-switching heteroskedastic errors:

$$\Delta \widetilde{D}_{i,t}^{\text{Data}} = \sum_{j=1}^{\mathcal{J}} w_{i,j} \, f\left(t - t_{0,i} - c_{i,j}, a_{i,j}, b_{i,j}\right) + \sigma_{i,k_t} \varepsilon_{i,t},$$

where $\varepsilon_{i,t}$ is an iid standard normal random residual, weights $w_{i,j}$ are non-negative and sum to one across j, and

$$f\left(t - t_{0,i} - c_{i,j}, a_{i,j}, b_{i,j}\right) = \frac{b_{i,j}}{a_{i,j}} \left(\frac{t - t_{0,i} - c_{i,j}}{a_{i,j}}\right)^{b_{i,j}-1} \exp\left(-\left(\frac{t - t_{0,i} - c_{i,j}}{a_{i,j}}\right)^{b_{i,j}}\right)$$

when $t - t_{0,i} - c_{i,j} \geq 0$ and $t_{0,i}$ is the time when the cumulative death toll reached 25 for location i. The switching state $k_t \in \{1, \ldots, \mathcal{K}\}$ follows a Markov-switching process and can accommodate both an expectedly large surge in daily deaths and a low death volatility typically associated with a low number of deaths.[17]

Given the estimates of our death model parameters, we construct the estimates of death growth rates by taking the time derivative of our estimated mixture of Weibull densities

$$\hat{g}_i(t) = \frac{\sum_{j=1}^{\mathcal{J}} w_{i,j} \frac{\partial}{\partial t} f\left(t - t_{0,i} - c_{i,j}, a_{i,j}, b_{i,j}\right)}{\sum_{j=1}^{\mathcal{J}} w_{i,j} \, f\left(t - t_{0,i} - c_{i,j}, a_{i,j}, b_{i,j}\right)}.$$

The left panels of Figure 1 shows the fit of our Weibull estimation to the noisy daily death data for four selected locations from four different continents: New York State, Sweden, Panama, and Iran. These locations span the range of daily death patterns observed in the data and illustrate that our flexible specification is able to fit disparate patterns. New York had a steep initial rise in deaths followed by a decline that slowed considerably over time. Sweden experienced a flatter wave of deaths. Panama's daily deaths were nearly flat for several weeks and then began to rise, a similar pattern to that observed in some U.S. states like Texas and Florida. Iran, which has struggled to contain the pandemic, was well into a second and larger wave of daily deaths by the end of our estimation period. The x-axis of these panels

---

[17]For the selection of $\mathcal{K}$ and estimation details, see Appendix A and Atkeson et al. (2020).

11

also show the starting date at which each of these regions first reached 25 cumulative deaths.

The right panels of Figure 1 shows the corresponding estimates for the growth rate of daily deaths (on the left axis) and the of the effective reproduction number implied by equation (9) (on the right axis). In these panels, we see the rapid decline of the growth rates of daily deaths in these locations in the 30 days after each of these regions reached 25 cumulative deaths.

The smoothed daily death series in the left column of Figure 1 are based on 7-day moving averages following Baqaee et al. (Forthcoming) using a procedure described in Appendix A). We present these non-parametric estimates of the trend in daily deaths as a check on our estimation procedure. As one can see, the 0.90 probability bands capture most of the movements in the smoothed series. However, the smoothed series are heavily influenced by large spikes in the noisy daily death data. As a result, the growth rates of smoothed daily deaths can fluctuate wildly (see Figure A.1 in Appendix A).

4   Findings

Our estimation results yield the following four stylized facts about the COVID-19 epidemic.

Fact 1. *The growth rate of daily deaths from COVID-19 fell rapidly everywhere within the first 30 days after each region reached 25 cumulative deaths.*

Fact 1 is shown in the top panel of Figure 2. The solid black line shows that the median estimated growth rate of daily deaths fell rapidly from an initial level of about 40 percent to zero within the first 30 days of the estimation period. Notice that the 0.90 posterior probability intervals computed across all locations that we consider, the dashed lines in the figure, follow a similar pattern.[18] The estimated

---

[18]The posterior probability intervals in Figure 2 include both location uncertainty and sampling

12

half-life of the growth rate during this period is 2.7 days with a 0.90 probability interval ranging from 1.3 to 7.1.

Fact 2. *After this first period of rapid decline, the growth rate of daily deaths in all regions has hovered around zero or slightly below.*

As Figure 2 shows, after the initial 30 day period, the median growth rate of daily deaths has been nearly flat and slightly below zero. The 0.90 posterior probability interval has also been low relative to its initial range. Notice in the right column of Figure 1 that the growth rate of daily deaths has followed a pattern consistent with Facts 1 and 2 in all four of our example regions despite the heterogeneity in both their geographic locations and daily death patterns.

Fact 3. *The cross-regional standard deviation of growth rates of deaths fell rapidly in the first 10 days of the epidemic and has, subsequently, remained low relative to its initial level.*

Fact 3 can also be seen in Figure 2. The shrinking of the two-thirds and 0.90 posterior probability intervals shows that the dispersion in death growth rates across locations fell sharply within the initial 10 days of the epidemic. While there has been some increase in dispersion towards the end of the estimation period, overall levels of dispersion remain small relative to their initial values.

Facts 1 - 3, when taken together, demonstrate important differences in the patterns of daily death growth rates early on versus later on in the epidemic. In particular, the facts indicate that there is a structural break in the COVID-19 death data occurring roughly 30 days after 25 cumulative deaths occur. Prior to the structural break is a period characterized by rapidly declining growth rates of daily deaths and high cross-regional dispersion. The period after the structural break features growth rates of daily deaths that mostly hover near but slightly below zero. This latter period also features substantially less cross-regional dispersion compared to the early period.

---

uncertainty. However, most of the cross-sectional dispersion in growth rates in the figure is driven by location uncertainty as sampling uncertainty within a location is small (see the right column of Figure 1).

Fact 4. *When interpreted through a range of epidemiological models, Facts 1 - 3 imply that both the effective reproduction numbers and transmission rates of COVID-19 fell rapidly from widely dispersed initial levels during the 30 days after cumulative deaths reached 25. After this initial period of rapid decline, the effective reproduction number has hovered around one everywhere in the world.*

Fact 4 is obtained by deriving the implications of our estimated daily death paths for the paths of the effective reproduction number and transmission rates using four variants of an SIR model. The baseline version is described above, and the effective reproduction number and normalized transmission rates implied by this model are obtained from the deaths data using equations (8), (5), and (9). The parameter γ is set to 0.2 such that a growth rate of daily deaths of 30 percent corresponds to a basic reproduction number of 2.5. The right y-axis in Figure 2 shows that the median effective reproduction number fell from an initial level of nearly 3 to 1 in 30 days and has, subsequently, remained slightly below 1.

Equation (5) illustrates that the effective reproduction number can fall due to both declines in the normalized transmission rate and declines in the fraction of the population remaining susceptible to the disease. We use our estimated paths for daily deaths and the equations of the SIR model to determine the relative contributions of each. We find that the rapid decline in daily death growth rates early on in the epidemic is primarily due to a rapid fall in the transmission rate of the disease as illustrated by the bottom panel of Figure 2. Disease transmission rates, like the effective reproduction numbers, fell early on in the epidemic from widely dispersed initial levels and and have since remained close to 1.

To explore the robustness of our results to model uncertainty, we consider three variants of the baseline SIR model: an SEIR model which extends the SIR model by assuming agents first become exposed to the disease before becoming infectious, an SIHR model which allows for a longer period between infection and death by adding a hospitalized state, and an SEIHR model which extends the SIR model by adding both the exposed state and the hospitalized state. Following a similar procedure as

14

with the baseline SIR model, for each model extension we are able to express the effective reproduction number as a function of model parameters, as well as, daily deaths and its derivatives.[19]

Figure 3 shows that our findings are robust to variations in the baseline SIR model. The left panel of the figure shows that, across all four model variations, estimated effective reproduction numbers fell rapidly in the first period before the structural break. In the second period after the structural break, the median effective reproductive ratios implied by each model, hover slightly below 1. The first period is also characterized by more cross-region dispersion in reproductive ratios. The right column in the figure shows that, regardless of model, the cross-regional standard deviation of effective reproduction numbers fell substantially within the first 10 days of the estimation period and, subsequently, remained stable and relatively low.

5   Conclusion

One of the central policy questions regarding the COVID-19 pandemic is the question of which non-pharmeceutical interventions governments might use to influence the transmission of the disease. Our ability to identify empirically which NPI's have what impact on disease transmission depends on there being enough independent variation in both NPI's and disease transmission across locations as well as our having robust procedures for controlling for other observed and unobserved factors that might be influencing disease transmission. The facts that we document in this paper cast doubt on this premise.

Our finding in Fact 1 that early declines in the transmission rate of COVID-19 were nearly universal worldwide suggest that the role of region-specific NPI's implemented in this early phase of the pandemic is likely overstated. This finding instead suggests that some other factor(s) common across regions drove the early and rapid transmission rate declines. While all three factors mentioned in the introduction,

---

[19]Details on these model extensions are provided in Appendix B.

voluntary social distancing, the network structure of human interactions, and the nature of the disease itself, are natural contenders, disentangling their relative roles is difficult.

Our findings in Fact 2 and Fact 3 further raise doubt about the importance in NPI's (lockdown policies in particular) in accounting for the evolution of COVID-19 transmission rates over time and across locations. Many of the regions in our sample that instated lockdown policies early on in their local epidemic, removed them later on in our estimation period, or have have not relied on mandated NPI's much at all. Yet, effective reproduction numbers in all regions have continued to remain low relative to initial levels indicating that the removal of lockdown policies has had little effect on transmission rates.

The existing literature has concluded that NPI policy and social distancing have been essential to reducing the spread of COVID-19 and the number of deaths due to this deadly pandemic. The stylized facts established in this paper challenge this conclusion. We argue that research going forward should account for these facts when assessing how important NPI policy is in shaping the progression of COVID-19.
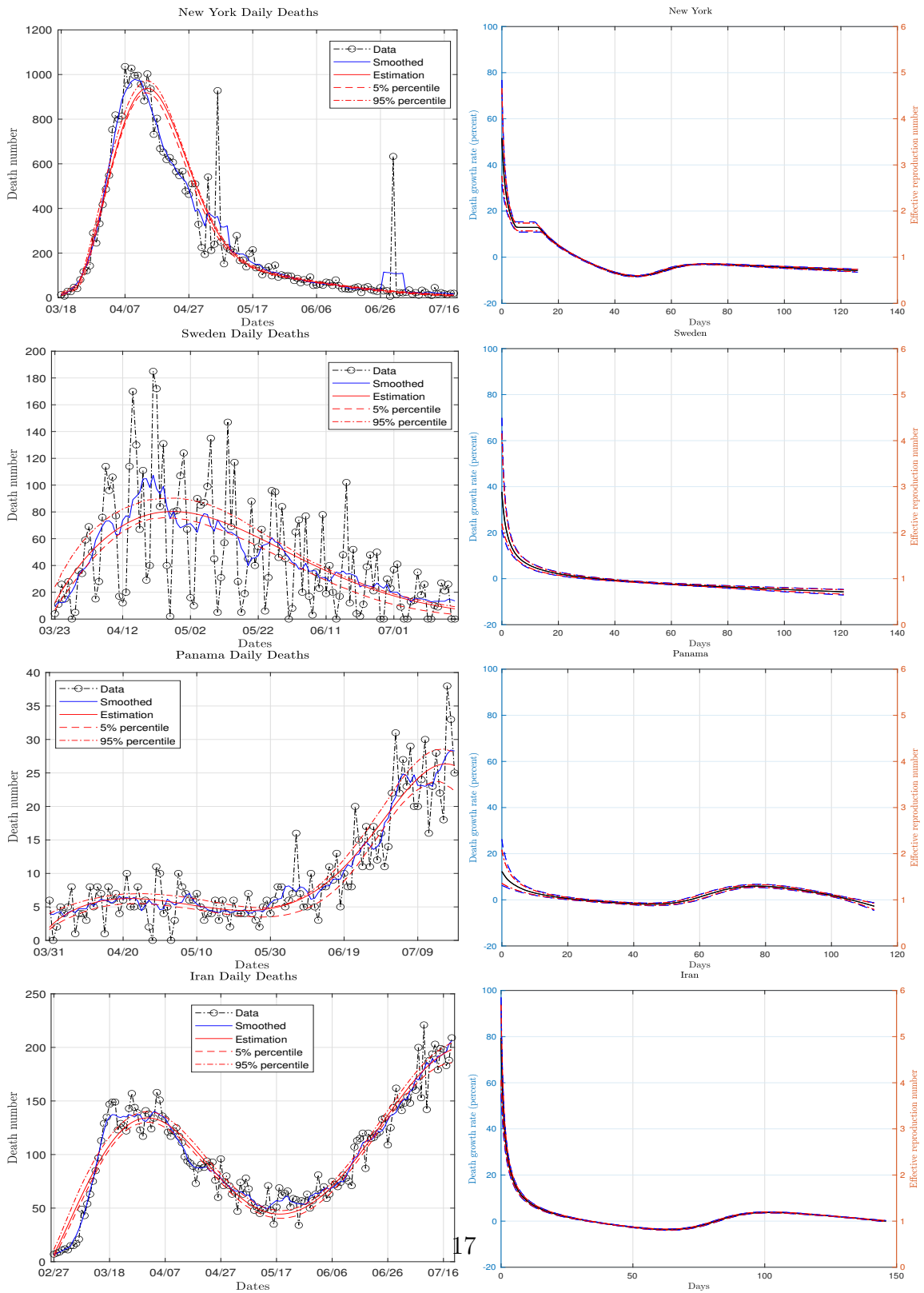
Figure 1: Left column: estimated trend of death with 0.90 probability (credible) bands. The smoothed death (the solid blue line) is a 7-day average. Right column: growth rates of death with 0.90 probability (credible) bands. Effective reproduction number is based on the SIR model.
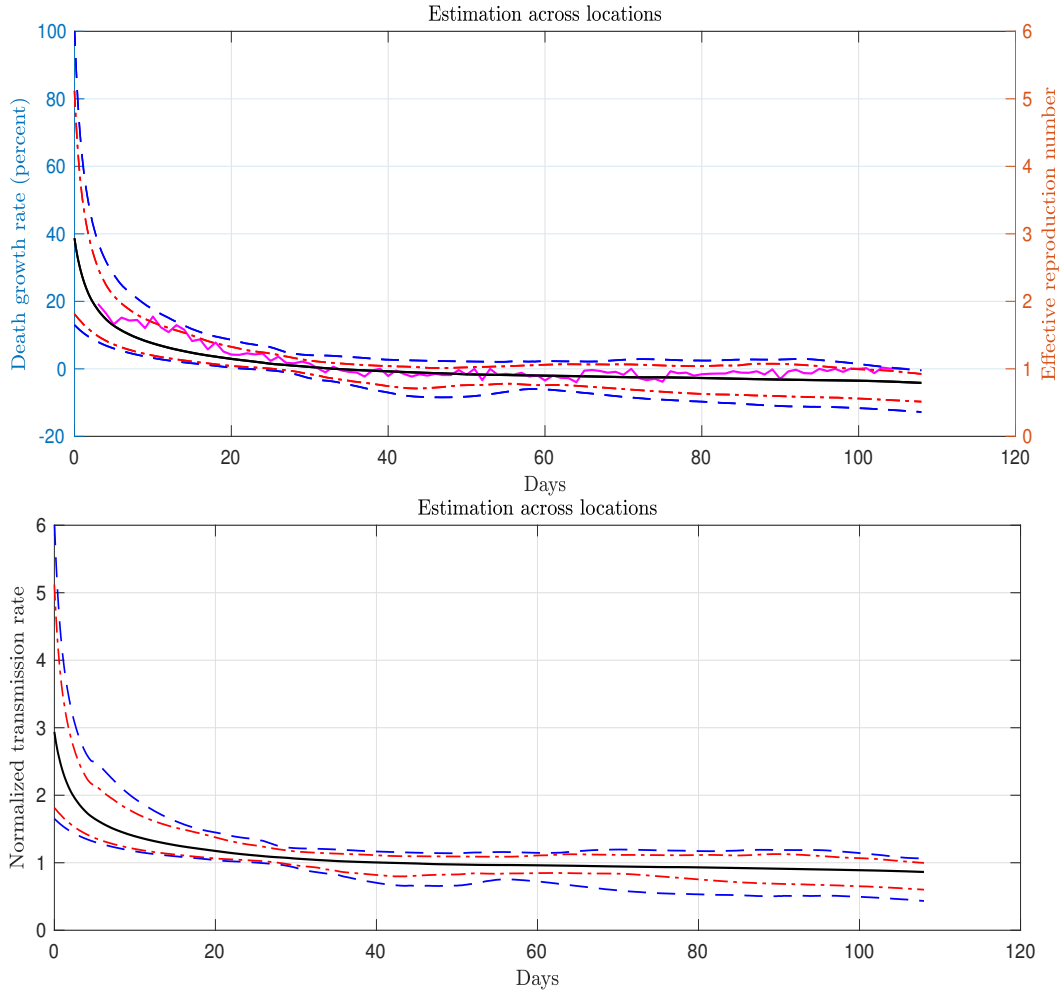
17

Figure 2: Location and sampling uncertainty. The black solid line in both charts represents the median posterior estimate. The solid magenta line in the top chart represents the median growth rate of 7-day smoothed daily deaths for all 50 locations and corresponds only to the left scale. The two dash-dotted bands in both charts contain two thirds of the posterior probability at each point in time and the two dashed bands, 0.90 of the posterior probability. The growth rates of death is estimated according to the fitted Weibull function. Effective reproduction numbers and normalized transmission rates are based on the SIR model. Day 0 is the earliest date when the cumulative death toll reached 25 in each location.
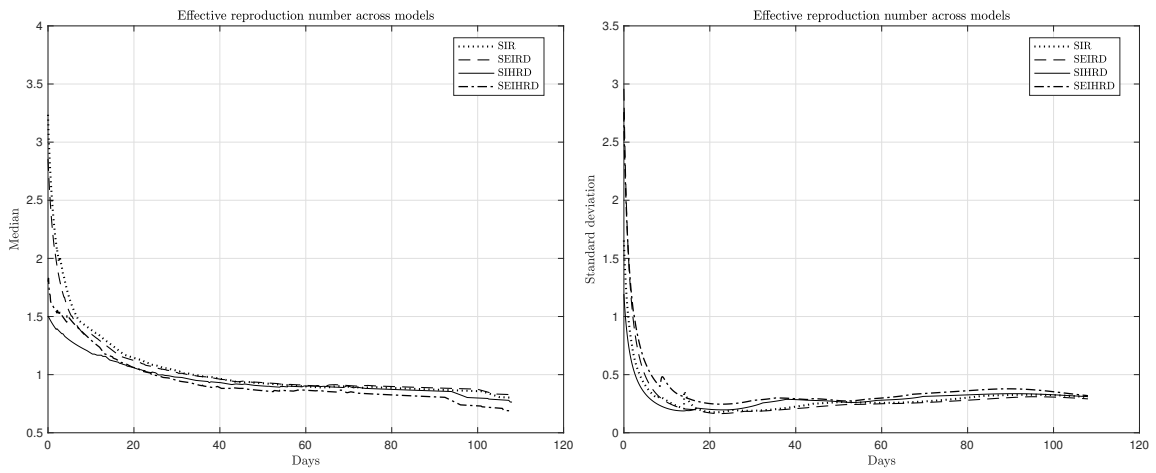
Figure 3: Model uncertainty: effective reproduction numbers estimated from the four epidemiological models. The estimated median and standard deviations for each model are based on all 50 locations. Day 0 is the earliest date when the cumulative death toll reached 25 in each location.

References

Mohammad Akbarpour, Cody Cook, Aude Marzuoli, Simon Mongey, Abhishek Nagaraj, Matteo Saccarola, Pietro Tebaldi, Shoshana Vasserman, and Hanbin Yang. Socioeconomic network heterogeneity and pandemic policy response. Working Paper 27374, National Bureau of Economic Research, June 2020. URL http://www.nber.org/papers/w27374.

Hunt Allcott, Levi Boxell, Jacob Conway, Billy Ferguson, Matthew Gentzkow, and Benny Goldman. Economic and health impacts of social distancing policies during the coronavirus pandemic. May 2020.

Nikos Askitas, Kontantinos Tatsiramos, and Bertrand Verheyden. Lockdown strategies, mobility patterns and covid-19. Covid Economics 23, CEPR, 28 May 2020.

Karl M Aspelund, Michael C Droste, James H Stock, and Christopher D Walker. Identification and estimation of undetected covid-19 cases using testing data from iceland. Working Paper 27528, National Bureau of Economic Research, July 2020. URL http://www.nber.org/papers/w27528.

Andrew Atkeson. On using sir models to model disease scenarios for covid-19. *Federal Reserve Bank of Minneapolis Quarterly Review*, (4111), June 2020.

Andrew Atkeson, Karen Kopecky, and Tao Zha. Estimating and forecasting disease scenarios for covid-19 with an sir model. Working Paper 27335, National Bureau of Economic Research, June 2020. URL http://www.nber.org/papers/w27335.

David Baqaee, Emmanuel Farhi, Michael Mina, and James H. Stock. Policies for a second wave. *Brookings Papers on Economic Activity*, July Forthcoming.

Michael Beenstock and Xieer Dai. The natural and unnatural histories of covid-19 contagion. Technical report, CEPR Covid Economics, April 27 2020.

Alberto Bisin and Andrea Moro. Learning epidemiology by doing: The empirical implications of a spatial-sir model with behavioral responses. Working Paper

27590, National Bureau of Economic Research, July 2020. URL http://www.nber.org/papers/w27590.

Jean-Philippe Bonardi, Quentin Gallea, Dimtrija Kalanoski, and Rafael Lalive. Fast and local: How lockdown policies affect the spread and severity of covid-19. Covid Economics 23, CEPR, 28 May 2020.

Tom Britton, Frank Ball, and Pieter Trapman. A mathematical model reveals the influence of population heterogeneity on herd immunity to sars-cov-2. *Science*, 2020. ISSN 0036-8075. doi: 10.1126/science.abc6810. URL https://science.sciencemag.org/content/early/2020/06/22/science.abc6810.

Tomaž Bučar, Marko Nagode, and Matija Fajdiga. Reliability approximation using finite weibull mixture distributions. *Reliability Engineering & System Safety*, 84(3): 241–251, 2004.

Shelby R. Buckman, Reuven Glick, Kevin J. Lansing, Nicolas Petrosky-Nadeu, and Lily M. Seitelman. Replicating and projecting the path of covid-19 with a model-implied reproduction number. Working Paper 2020-24, Federal Reserve Bank of San Francisco, July 2020.

Victor Chernozhukov, Hiroyuki Kasahara, and Paul Schrimpf. Causal impact of masks, policies, behavior on early covid-19 pandemic in the us. Covid Economics 35, CEPR, 7 July 2020.

Gerado Chowell, Lisa Sattenspiel, Shweta Bansal, and Cecile Viboud. Mathematical models to characterize early epidemic growth: A review. *Physical Life Review*, 18: 66–97, September 2016.

Martin J Conyon, Lerong He, and Steen Thomsen. Lockdowns and covid-19 deaths in scandinavia. Covid Economics 26, CEPR, 5 June 2020.

Anne Cori, Neil M Ferguson, Christophe Fraser, and Simon Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, 178(9), 2013.

Christopher J Cronin and William N Evans. Private precaution and public restrictions: What drives social distancing and industry foot traffic in the covid-19 era? Working Paper 27531, National Bureau of Economic Research, July 2020. URL http://www.nber.org/papers/w27531.

Pragyan Deb, Davide Furceri, Jonathan D. Ostroy, and Nour Tawk. The effect of containment measures on the covid-19 pandemic. Covid Economics 19, CEPR, 18 May 2020.

Jonas Dehning, Johannes Zierenberg, F. Paul Spitzner, Michael Wibral, Joao Pinheiro Neto, Michael Wilczek, and Viola Priesemann. Inferring change points in the spread of covid-19 reveals the effectiveness of interventions. *Science*, May 2020.

Kjell A Doksum and Bo Lindqvist. *Mathematical and statistical methods in reliability*, volume 7. World Scientific, 2003.

Julie M. Donohue and Elizabeth Miller. COVID-19 and School Closures. *JAMA*, 07 2020. ISSN 0098-7484. doi: 10.1001/jama.2020.13092. URL https://doi.org/10.1001/jama.2020.13092.

Martin S Eichenbaum, Sergio Rebelo, and Mathias Trabandt. The macroeconomics of epidemics. Working Paper 26882, National Bureau of Economic Research, March 2020. URL http://www.nber.org/papers/w26882.

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

Glenn Ellison. Implications of heterogeneous sir models for analyses of covid-19. Working Paper 27373, National Bureau of Economic Research, June 2020. URL http://www.nber.org/papers/w27373.

Pablo Fajgelbaum, Amit Khandelwal, Wookun Kim, Cristiano Mantovani, and Edouard Schaal. Optimal lockdown in a commuting network. Working Paper 27441, National Bureau of Economic Research, June 2020. URL http://www.nber.org/papers/w27441.

Hanming Fang, Long Wang, and Yang Yang. Human mobility restrictions and the spread of the novel coronavirus (2019-ncov) in china. Working Paper 26906, National Bureau of Economic Research, March 2020. URL http://www.nber.org/papers/w26906.

Maryam Farboodi, Gregor Jarosch, and Robert Shimer. Internal and external effects of social distancing in a pandemic. Working Paper 27059, National Bureau of Economic Research, April 2020. URL http://www.nber.org/papers/w27059.

Jesus Fernandez-Villaverde and Charles I. Jones. Estimating and simulating a SIRD model of COVID-19 for many countries, states, and cities. Working Paper 27128, NBER, April 2020.

Thiemo R Fetzer, Marc Witte, Lukas Hensel, Jon Jachimowicz, Johannes Haushofer, Andriy Ivchenko, Stefano Caria, Elena Reutskaja, Christopher P Roth, Stefano Fiorin, Margarita Gómez, Gordon Kraft-Todd, Friedrich M Götz, and Erez Yoeli. Global behaviors and perceptions at the onset of the covid-19 pandemic. Working Paper 27082, National Bureau of Economic Research, May 2020. URL http://www.nber.org/papers/w27082.

Seth Flaxman, Swapnil Mishra, Axel Gandy, H. Juliette T. Unwin, Thomas A. Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W. Eaton, Mélodie Monod, Pablo N. Perez-Guzman, Nora Schmit, Lucia Cilloni, Kylie E. C. Ainslie, Marc Baguelin, Adhiratha Boonyasiri, Olivia Boyd, Lorenzo Cattarino, Laura V. Cooper, Zulma Cucunubá, Gina Cuomo-Dannenburg, Amy Dighe, Bimandra Djaafara, Ilaria Dorigatti, Sabine L. van Elsland, Richard G. FitzJohn, Katy A. M. Gaythorpe, Lily Geidelberg, Nicholas C. Grassly, William D. Green, Timothy Hallett, Arran Hamlet, Wes Hinsley, Ben Jeffrey, Edward Knock, Daniel J. Laydon, Gemma Nedjati-Gilani, Pierre Nouvellet, Kris V. Parag, Igor Siveroni, Hayley A. Thompson, Robert Verity, Erik Volz, Caroline E. Walters, Haowei Wang, Yuanrong Wang, Oliver J. Watson, Peter Winskill, Xiaoyue Xi, Patrick GT Walker, Azra C. Ghani, Christl A. Donnelly, Steven M. Riley, Michaela A. C. Vollmer, Neil M. Ferguson, Lucy C. Okell, Samir Bhatt, and Imperial College COVID-19 Response Team.

Estimating the effects of non-pharmaceutical interventions on covid-19 in europe. *Nature*, 2020.

Edward L. Glaeser, Caitlin S. Gorback, and Stephen J. Redding. How much does covid-19 increase with mobility? evidence from new york and four other u.s. cities. Working Paper 27519, NBER, July 2020.

Austan Goolsbee and Chad Syverson. Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020. Working Paper 27432, National Bureau of Economic Research, June 2020. URL http://www.nber.org/papers/w27432.

Veronica Guerrieri, Guido Lorenzoni, Ludwig Straub, and Iván Werning. Macroeconomic implications of covid-19: Can negative supply shocks cause demand shortages? Working Paper 26918, National Bureau of Economic Research, April 2020. URL http://www.nber.org/papers/w26918.

Sumedha Gupta, Thuy D Nguyen, Felipe Lozano Rojas, Shyam Raman, Byungkyu Lee, Ana Bento, Kosali I Simon, and Coady Wing. Tracking public and private responses to the covid-19 epidemic: Evidence from state and local government actions. Working Paper 27027, National Bureau of Economic Research, April 2020. URL http://www.nber.org/papers/w27027.

James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, 1989.

Simon I Hay. Covid-19 scenarios for the united states. *medRxiv*, 2020. doi: 10. 1101/2020.07.12.20151191. URL https://www.medrxiv.org/content/early/2020/07/14/2020.07.12.20151191.

Solomon Hsiang, Daniel Allen, Sébastien Annan-Phan, Kendon Bell, Ian Bolliger, Trinetta Chong, Hannah Druckenmiller, Luna Yue Huang, Andrew Hultgren, Emma Krasovich, Peiley Lau, Jaecheol Lee, Esther Rolf, Jeanette Tseng, and Tiffany Wu. The effect of large-scale anti-contagion policies on the covid-19 pandemic. *Nature*, 2020.

Greg Kaplan, Benjamin Moll, and Gianluca Violante. Pandemics according to hank. May 2020.

Jussi Keppo, Elena Quercioli, Marianna Kudlyak, Andrea Wilson, and Lones Smith. The behavioral sir model, with applications to the swine flu and covid-19 pandemics. April 2020.

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

Hannah M Korevaar, Alexander David Becker, Ian F Miller, Bryan T Grenfell, C. Jessica E. Metcalf, and Michael J Mina. Quantifying the impact of us state non-pharmaceutical interventions on covid-19 transmission. *medRxiv*, 2020. doi: 10.1101/2020.06.30.20142877. URL https://www.medrxiv.org/content/early/2020/07/01/2020.06.30.20142877.

Thomas Lecocq, Stephen P. Hicks, Koen Van Noten, Kasper van Wijk, Paula Koelemeijer, Raphael S. M. De Plaen, Frédérick Massin, Gregor Hillers, Robert E. Anthony, Maria-Theresia Apoloner, Mario Arroyo-Solórzano, Jelle D. Assink, Pinar Büyükakpınar, Andrea Cannata, Flavio Cannavo, Sebastian Carrasco, Corentin Caudron, Esteban J. Chaves, David G. Cornwell, David Craig, Olivier F. C. den Ouden, Jordi Diaz, Stefanie Donner, Christos P. Evangelidis, Läslo Evers, Benoit Fauville, Gonzalo A. Fernandez, Dimitrios Giannopoulos, Steven J. Gibbons, Társilo Girona, Bogdan Grecu, Marc Grunberg, György Hetényi, Anna Horleston, Adolfo Inza, Jessica C. E. Irving, Mohammadreza Jamalreyhani, Alan Kafka, Mathijs R. Koymans, Celeste R. Labedz, Eric Larose, Nathaniel J. Lindsey, Mika McKinnon, Tobias Megies, Meghan S. Miller, William Minarik, Louis Moresi, Víctor H. Márquez-Ramírez, Martin Möllhoff, Ian M. Nesbitt, Shankho Niyogi, Javier Ojeda, Adrien Oth, Simon Proud, Jay Pulli, Lise Retailleau, Annukka E. Rintamäki, Claudio Satriano, Martha K. Savage, Shahar Shani-Kadmiel, Reinoud Sleeman, Efthimios Sokos, Klaus Stammler, Alexander E. Stott, Shiba Subedi, Mathilde B. Sørensen, Taka'aki Taira, Mar Tapia, Fatih Turhan, Ben van der Pluijm, Mark Vanstone, Jerome Vergne,

Tommi A. T. Vuorinen, Tristram Warren, Joachim Wassermann, and Han Xiao. Global quieting of high-frequency seismic noise due to covid-19 pandemic lockdown measures. *Science*, 2020. ISSN 0036-8075. doi: 10.1126/science.abd2438. URL https://science.sciencemag.org/content/early/2020/07/22/science.abd2438.

Laura Liu, Hyungsik Roger Moon, and Frank Schorfheide. Panel forecasts of country-level covid-19 infections. Working Paper 27248, National Bureau of Economic Research, May 2020. URL http://www.nber.org/papers/w27248.

Kristine A. Moore, Marc Lipsitch, John M. Barry, and Michael T. Osterholm. The future of the covid-19 pandemic: Lessons learned from pandemic influenza. Viewpoint Part 1, CIDRAP, the University of Minnesota, April 2020.

Christopher JL Murray. Forecasting the impact of the first wave of the covid-19 pandemic on hospital demand and deaths for the usa and european economic area countries. *medRxiv*, 2020. doi: 10.1101/2020.04.21.20074732. URL https://www.medrxiv.org/content/early/2020/04/26/2020.04.21.20074732.

L. C. Okell, R. Verity, O. J. Watson, S. Mishra, P. Walker, C. Whittaker, A. Katzourakis, C. A. Donnelly, S. Riley, A. C. Ghani, A. Gandy, S. Flaxman, N. M. Ferguson, and S. Bhatt. Have deaths from covid-19 in europe plateaued due to herd immunity? *Lancet*, 395(10241):e110–e111, June 2020.

Tomas Phillipson and Richard A. Posner. *Private Choices and Public Health: the AIDS Epidemic in an Economic Perspective*. Harvard University Press, 1993.

John E Pinder III, James G Wiener, and Michael H Smith. The weibull distribution: a new method of summarizing survivorship data. *Ecology*, 59(1):175–179, 1978.

David Rubin, Jing Huang, Brian T. Fisher, Antonio Gasparrini, Vicky Tam, Lihai Song, Xi Wang, Jason Kaufman, Kate Fitzpatrick, Arushi Jain, Heather Griffis, Koby Crammer, Jeffrey Morris, and Gregory Tasian. Association of Social Distancing, Population Density, and Temperature With the Instantaneous Reproduction Number of SARS-CoV-2 in Counties Across the United States. *JAMA Network Open*, 3(7):

e2016099–e2016099, 07 2020. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2020. 16099. URL https://doi.org/10.1001/jamanetworkopen.2020.16099.

Marlena M Siwiak, Pawel Szczesny, and Marian P Siwiak. From a single host to global spread. the global mobility based modelling of the covid-19 pandemic implies higher infection and lower detection rates than current estimates. *medRxiv*, 2020. doi: 10.1101/2020.03.21.20040444. URL https://www.medrxiv.org/content/early/ 2020/05/13/2020.03.21.20040444.

Flavio Toxvaerd. Equilibrium social distancing. Working Paper 2020/8, Cambridge-INET, March 2020.

Efthymios G Tsionas. Bayesian analysis of finite mixtures of weibull distributions. *Communications in Statistics-Theory and Methods*, 31(1):37–48, 2002.

Enzo Weber. Which measures flattened the curve in germany? Covid Economics 24, CEPR, 1 June 2020.

Thomas Wieland. Change points in the spread of covid-19 question the effectiveness of nonpharmaceutical interventions in germany. Covid Economics 37, CEPR, 14 July 2020.

A Estimation Details

The procedure we use to estimate the trend in daily deaths with a seven day moving average is as follows. Let $D_i(t)$ denote the reported data on cumulative COVID deaths in region i at date t and $\Delta D_{i,t} = D_i(t) - D_i(t-1)$ denote the data on daily deaths. We estimate the level of daily deaths from these noisy data with a 7-day moving average

$$\overline{\Delta D_{i,t}} = \frac{1}{7} \sum_{s=0}^{6} \Delta D_{i,t-s},$$

starting from the day $t_{0,i}$ at which cumulative deaths in region i, $D_i(t)$, first reaches 25. We then smooth the growth rate of daily deaths with a finite difference

$$\bar{g}_i(t) = \frac{1}{7\overline{\Delta D_{i,t}}} \left[\overline{\Delta D_{i,t+3}} - \overline{\Delta D_{i,t-4}} \right].$$

The choice for computing the growth rate of daily deaths from a 7-day difference in smoothed daily deaths is to smooth estimates of the growth rate represented by $\bar{g}_i(t)$ (double-smoothing).

In Section 3, we discuss potential problems with the method of smoothing daily deaths over 7 days because such a smoothed series is still heavily influenced by large and noisy spikes of reported daily deaths despite the smoothing. Figure A.1 shows how wild the fluctuation of growth rates of 7-day smoothed daily deaths for all 50 locations can become.

By contrast, our Bayesian approach avoids the erratic movements driven by noisy daily death data. Let $\Delta D_{i,t}^{\text{Data}}$ be the daily measured object for $dD_i(t)/dt$ and denote

$$\Delta \widetilde{D}_{i,t}^{\text{Data}} = \frac{\Delta D_{i,t}^{\text{Data}}}{(1+d_i) D_{i,T}^{\text{Data}}},$$

$$\Delta \widetilde{D}_i^{t,\,\text{Data}} \equiv \left\{ \Delta \widetilde{D}_{i,t}^{\text{Data}}, \Delta \widetilde{D}_{i,t-1}^{\text{Data}}, \Delta \widetilde{D}_{i,t-2}^{\text{Data}}, \ldots \Delta \widetilde{D}_{i,\,t_{i,0}}^{\text{Data}} \right\},$$
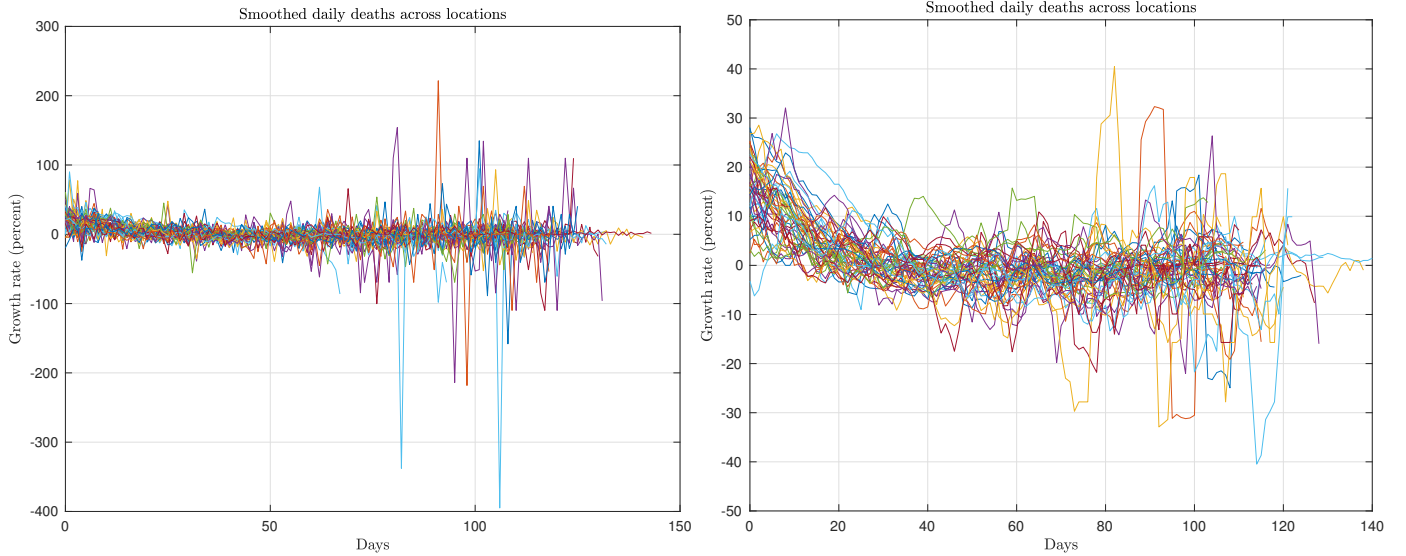
28

Figure A.1: Growth rates of 7-day smoothed daily deaths for all 50 locations. The left panel plots the day-to-day growth rates and the right panel plots the rolling smoothed day-to-day growth rate over the 7-day period. Both plots are based on a time series of 7-day average death numbers. Due to this smoothing, time 0 corresponds to the fourth day after the cumulative death reached 25 in each location.

$$\Delta \widetilde{D}_i^{\text{Data}} \equiv \left\{ \Delta \widetilde{D}_{i,\, t_{0,i}}^{\text{Data}}, \, \cdots, \, \widetilde{D}_{i,T}^{\text{Data}} \right\}.$$

Given the daily discrete-time data and the value of d_i, we estimate the following non-linear regression

$$\Delta \widetilde{D}_{i,t}^{\text{Data}} = f_i \left(t - t_{0,i}, \, \xi_i \right) + \sigma_{i,k_t} \varepsilon_{i,t}, \tag{A.1}$$

where $\varepsilon_{i,t}$ is a state-dependent standard normal random residual and independent of each other across i. The switching state $k_t \in \{1, \ldots, \mathcal{K}\}$ can accommodate both an expectedly large surge in daily deaths and a low death volatility typically associated with a low number of deaths. The transition matrix $\mathcal{Q}^k$ for k_t is unrestricted except that each column of $\mathcal{Q}^k$ sums to one. For our deterministic SIR model to offer an accurate account of health consequences of the pandemic, we set $t_{i,0}$ at the time when the number of deaths accumulates to 25 for each location i.

Let θ_i represent a collection of ξ_i, $\sigma_{i,1}, \cdots, \sigma_{i,\mathcal{K}}$, and all the free parameters in

29

the transition matrix $\mathcal{Q}^k$. From equation (A.1) one can derive the (log) conditional likelihood function for $\Delta \widetilde{D}_i^{t,\,\mathrm{Data}}$ as

$$
\log \mathcal{L}\left(\Delta \widetilde{D}_{i,t}^{\mathrm{Data}} \mid \Delta \widetilde{D}_i^{t-1,\,\mathrm{Data}}, k_t, \theta_i\right) =
$$

$$
\log \mathcal{L}\left(\Delta \widetilde{D}_{i,t}^{\mathrm{Data}} \mid k_t, \theta_i\right) = -\frac{T - t_{0,i}}{2} \log(2\pi) - \sum_{t=t_{0,i}}^{T} \log(\sigma_{i,k_t}) \tag{A.2}
$$

$$
- \sum_{t=t_{0,i}}^{T} \frac{\left[\Delta \widetilde{D}_{i,t}^{\mathrm{Data}} - f_i\left(t - t_{0,i},\, \xi_i\right)\right]^2}{2\sigma_{i,k_t}^2},
$$

$$
\mathcal{L}\left(\Delta \widetilde{D}_{i,t}^{\mathrm{Data}} \mid \Delta \widetilde{D}_i^{t-1,\,\mathrm{Data}}, \theta_i\right) =
$$

$$
\sum_{k_t=1}^{\mathcal{K}} \left[\mathcal{L}\left(\Delta \widetilde{D}_{i,t}^{\mathrm{Data}} \mid \Delta \widetilde{D}_i^{t-1,\,\mathrm{Data}},\, k_t,\, \theta_i\right) p\left(k_t \mid \Delta \widetilde{D}_i^{t-1,\,\mathrm{Data}}, \theta_i\right)\right]. \tag{A.3}
$$

Given the initial condition $p\left(k_{t_{0,i}-1} \mid \Delta \widetilde{D}_i^{t_{0,i}-1,\,\mathrm{Data}}, \theta_i\right) = 1/\mathcal{K}$, the predictive probability of regime, $p\left(k_t \mid \Delta \widetilde{D}_i^{t-1,\,\mathrm{Data}}, \theta_i\right)$, can be updated recursively through Hamilton (1989)'s filter as

$$
p\left(k_t \mid \Delta \widetilde{D}_i^{t-1,\,\mathrm{Data}}, \theta_i\right) = \sum_{k_{t-1}=1}^{\mathcal{K}} q_{k_t,k_{t-1}}\, p\left(k_{t-1} \mid \Delta \widetilde{D}_i^{t-1,\,\mathrm{Data}}, \theta_i\right)
$$

and

$$
p\left(k_t \mid \Delta \widetilde{D}_i^{t,\,\mathrm{Data}},\, \theta_i\right) =
$$

$$
\frac{\mathcal{L}\left(\Delta \widetilde{D}_{i,t}^{\mathrm{Data}} \mid \Delta \widetilde{D}_i^{t-1,\,\mathrm{Data}}, k_t, \theta_i\right) p\left(k_t \mid \Delta \widetilde{D}_i^{t-1,\,\mathrm{Data}}, \theta_i\right)}{\sum_{k_t=1}^{\mathcal{K}} \left[\mathcal{L}\left(\Delta \widetilde{D}_{i,t}^{\mathrm{Data}} \mid \Delta \widetilde{D}_i^{t-1,\,\mathrm{Data}}, k_t, \theta_i\right) p\left(k_t \mid \Delta \widetilde{D}_i^{t-1,\,\mathrm{Data}}, \theta_i\right)\right]}.
$$

It follows from equations (A.2) and (A.3) that the likelihood function for ΔD^{Data}

is

$$\mathcal{L}\left(\Delta\widetilde{D}_i^{\text{Data}} \mid \theta_i\right) = \prod_{t=t_{i,0}}^{T} \mathcal{L}\left(\Delta\widetilde{D}_{i,t}^{\text{Data}} \mid \Delta\widetilde{D}_i^{t-1,\text{Data}}, \theta_i\right). \tag{A.4}$$

We impose a Bayesian prior on the parameters of this model. Let $p(\theta_i)$ denote the prior pdf. We keep the prior diffuse enough to be agnostic. Specifically, a_i has the Gamma distribution with the hyperparameter values 40 and 1 and b_i with the hyperparameter values 4 and 1. The parameter c_i has the normal distribution with mean zero and variance 10. The parameter σ_{i,k_t} has the uniform distribution between 0 and 0.1.[20] The weight parametrer $w_{i,j}$ $(j = 1, \cdots, \mathcal{J})$ and the elements of $\mathcal{Q}_k$ all have the Dirichlet distribution with the hyperparameter values 2 and 2. It follows that the log posterior density function of θ_i is

$$\log p\left(\theta_i \mid \Delta\widetilde{D}_i^{\text{Data}}\right) = \log\mathcal{L}\left(\Delta\widetilde{D}_i^{\text{Data}} \mid \theta_i\right) + \log p(\theta_i) - \log p\left(\Delta\widetilde{D}_i^{\text{Data}}\right), \tag{A.5}$$

where $p\left(\Delta\widetilde{D}_i^{\text{Data}}\right)$ is the marginal likelihood. We use the Schwarz criterion (Bayesian information criterion) to select $\mathcal{J}$, the numbers of mixtures. The following results show the number of Weibulls that is needed to fit each location of the U.S. states and other countries used in our estimation:

$\mathcal{J} = 1$ Colorado, Connecticut, Georgia, Illinois, Indiana, Maryland, Massachusetts, Minnesota, Missouri, New Jersey, North Carolina, Ohio, Pennsylvania, Virginia, Washington, Spain, Brazil, Canada, Ireland, Japan, Mexico, Netherlands, Peru, Russia, Sweden, Switzerland, and India.

$\mathcal{J} = 2$ Alabama, Arizona, California, Florida, Louisiana, Michigan, Mississippi, New York, South Carolina, Texas, the rest of the U.S., Argentina, Belgium, France, Germany, Iran, Italy, Panama, Portugal, Turkey, Denmark, United Kingdom,

---

[20]The linear regression literature only uses the inverse Gamma distribution for σ_{i,k_t} as conjugate prior. For our nonlinear regression here, the inverse Gamma prior is no longer conjugate. Moreover, since the death data implies that the values of σ_{i,k_t} are far less than one, the inverse Gamma prior only with no moments can cover such small values. This makes the inverse Gamma prior not only undesirable but also impractical.

and Chile.

The above estimation procedure is conditional on the value of d_i, which does not affect the dynamics of the model's variables except scaling the cumulative deaths $D_i(t)$. We propose the following algorithm to find the scaling parameter d_i.

Algorithm 1. *Let $d_i^{(0)} > 0$ be the initial value of d_i and $\theta_i^{(0)}$ be the estimate conditional on $d_i^{(0)}$, and denote $\Delta \widehat{D}_{i,t}^{(0)} = f_i \left(t - t_{0,i}, \xi_i^{(0)} \right)$. For $\ell = 1, 2, \cdots$, the algorithm proceeds as follows.*

1. *Find d_i that minimizes*

$$\log \sum_{t=t_{0,i}}^{T} \left[D_{i,t}^{Data} - \sum_{s=t_{0,i}}^{t} \Delta \widehat{D}_{i,s}^{(\ell-1)} (1 + d_i) D_{i,T}^{Data} \right]^2$$

   and denote this value by $d_i^{(\ell)} > 0$.

2. *Estimate θ_i conditional on $d_i^{(\ell)} > 0$ as described in this section and denote the estimate by θ_i^{ℓ}.*

3. *Repeat the last two steps until $d_i^{(\ell)}$ converges.*

The left column of Figure 1 in the main text reports estimation results for four selected locations. For all other locations, the fit is as good as Figure 1.[21] The computation is expensive: it takes about 24 hours to finish 100,000 posterior draws for each location. There are a total of 50 locations. We use a cluster of computers to accomplish this task with a total of 5 million posterior draws. For each posterior draw of a_i, b_i, and c_i, we invert the four SIR-related structural models to obtain the effective reproduction number and transmission rate. The computation is also expensive if the number of estimated Weibull functions is greater than one. For a mixture of two Weibull functions, for instance, solving each structural model with

---

[21]See also Atkeson et al. (2020) for detailed results from the estimation with earlier death data.

1,000 posterior draws takes at least two hours for each location. Figure A.2 report the results from the four structural models based on the median estimates of Weibull parameters.
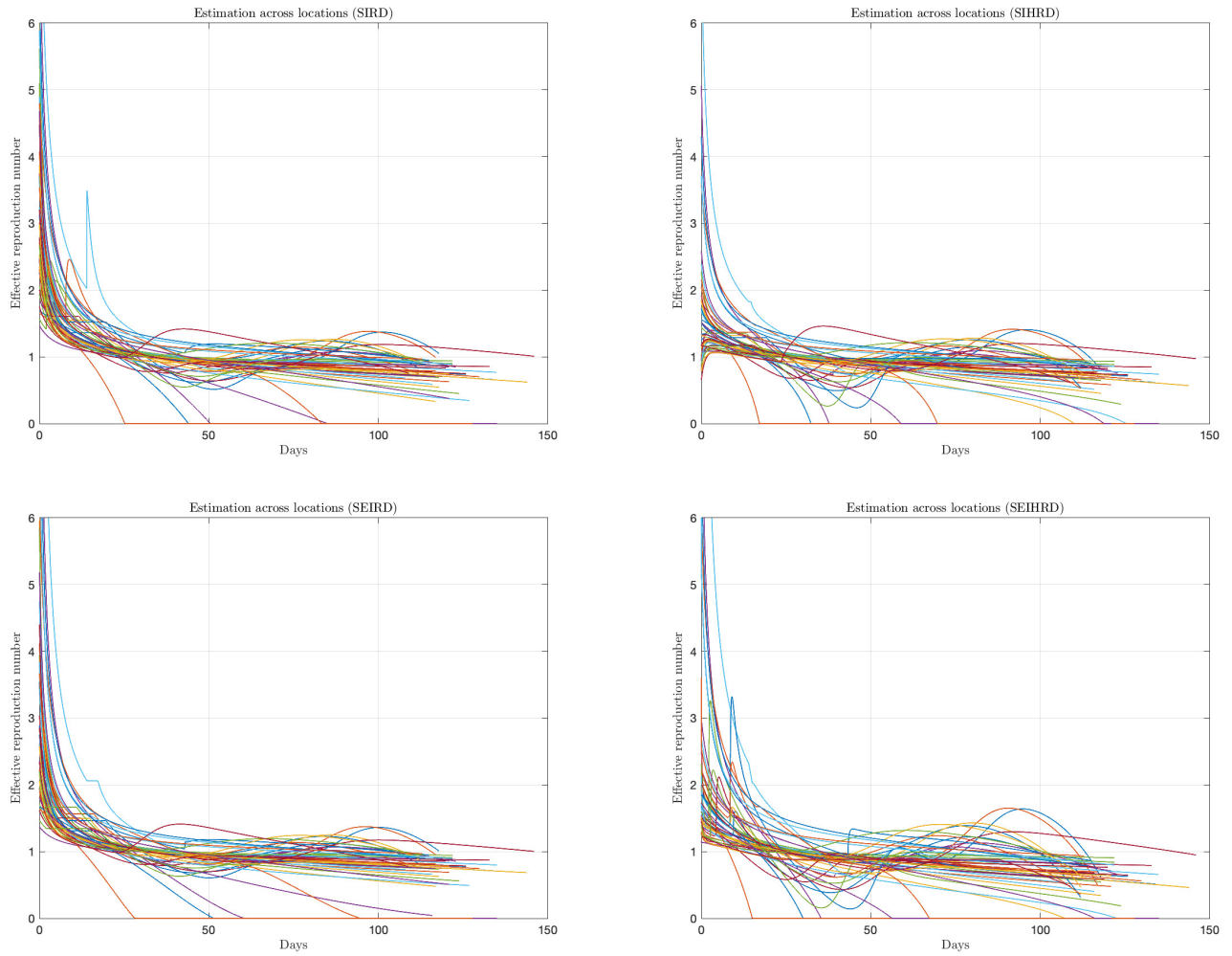

Figure A.2: Effective reproduction numbers across all 50 locations, estimated from the four extended models. The day 0 is the earliest date when the cumulative death toll reached 25 in each location.

33

B Extensions of the SIR model

In this section we describe the three extensions of the standard SIR model that we consider.

B.1 SEIR model

The SEIR model extends the SIR model by assuming agents first become exposed to the disease before becoming infected (infectious). At each moment in time, the population N is divided into five states: susceptible S, exposed E, infected I, recovered R, and dead D. Susceptible agents are at risk of becoming exposed to the disease. Agents in the exposed state are not infectious but transition to the infectious state at rate σ. Note that $1/\sigma$ is the average number of days that an agent has been exposed to the disease but is not yet infectious.

The equations of the model are given by

$$\frac{dS(t)}{dt} = -\mathcal{R}(t)\gamma I(t), \tag{B.1}$$

$$\frac{dE(t)}{dt} = \mathcal{R}(t)\gamma I(t) - \sigma E(t), \tag{B.2}$$

$$\frac{dI(t)}{dt} = \sigma E(t) - \gamma I(t), \tag{B.3}$$

$$\frac{dR(t)}{dt} = (1-\nu)\gamma I(t), \tag{B.4}$$

$$\frac{dD(t)}{dt} = \nu\gamma I(t). \tag{B.5}$$

As in the SIR model, the effective reproduction number of the disease is

$$\mathcal{R}(t) \equiv \frac{\beta(t)}{\gamma} \frac{S(t)}{1 - D(t)}.$$

We assume that $E(0) = R(0) = D(0) = 0$, $S(0)$ is slightly below N and $I(0)$ is slightly above 0.

The model can be inverted as follows. As before, we have from equations B.4 and B.5 that

$$\frac{dR(t)}{dt} = \frac{1 - \nu}{\nu} \frac{dD(t)}{dt}.$$

Integrating this expression over time and using the initial conditions $D(0) = R(0) = 0$ gives

$$R(t) = \frac{1 - \nu}{\nu} D(t).$$

Equation B.5 gives

$$I(t) = \frac{1}{\nu\gamma} \frac{dD(t)}{dt},$$

and differentiating this equation we have

$$\frac{dI(t)}{dt} = \frac{1}{\nu\gamma} \frac{d^2 D(t)}{dt^2}.$$

Thus, from equation B.3, we have

$$E(t) = \frac{1}{\sigma} \left[\frac{1}{\nu} \frac{dD(t)}{dt} + \frac{1}{\nu\gamma} \frac{d^2 D(t)}{dt^2} \right].$$

Using the constraint that categories sum to one gives

$$S(t) = 1 - E(t) - I(t) - R(t) - D(t).$$

Note that the model implications for the numbers infected and recovered are the same as the SIR model. The number still susceptible, however, is adjusted to account for the stock of exposed agents.

Now consider the effective reproduction number. Note that if we sum equations B.2 and B.3, we get

$$\frac{dE(t)}{dt} + \frac{dI(t)}{dt} = [\mathcal{R}(t) - 1] \, \gamma I(t).$$

Differentiating the equation for $E(t)$ and plugging the other expressions in to the equation above gives

$$\frac{1}{\sigma} \left[\frac{d^2 D(t)}{dt^2} + \frac{1}{\gamma} \frac{d^3 D(t)}{dt^3} \right] + \frac{1}{\gamma} \frac{d^2 D(t)}{dt^2} = [\mathcal{R}(t) - 1] \frac{dD(t)}{dt}.$$

Thus, we get the following expression for the effective reproduction number

$$\mathcal{R}(t) = 1 + \left(\frac{1}{\gamma} + \frac{1}{\sigma} \right) \frac{\frac{d^2 D(t)}{dt}}{\frac{dD(t)}{dt}} + \frac{1}{\sigma\gamma} \frac{\frac{d^3 D(t)}{dt^3}}{\frac{dD(t)}{dt}}.$$

Notice that the effective reproduction number is now a function of both the first and second derivatives of daily deaths.

We choose the two parameters γ and σ so that the model is consistent with typical observed doubling times of daily deaths early on in the epidemic and a basic reproduction number of around 2.5. If we have daily deaths growing exponentially in the early phase of the epidemic with growth rate δ, then

$$\frac{dD(t)}{dt} = \exp(\bar{d} + \delta t).$$

From our formula for the effective reproduction number above

$$\mathcal{R}(0) = 1 + \left(\frac{1}{\gamma} + \frac{1}{\sigma} \right) \delta + \frac{\delta^2}{\sigma\gamma}.$$

We set σ to 0.5 which implies that the average number of days between exposure and infection is 2. The recovery rate γ is set at 0.4. At this value, a 30% growth rate of daily deaths (doubling time of 2.3 days) corresponds to a basic reproduction number of 2.8.

B.2   SIHR model

We now extend the SIR model by adding a state H corresponding to hospitalized. This additional state allows for a longer period from infection to death than in the simpler SIR model. Infected agents flow from state I to either hospitalized, H, or recovered, R. Agents in state H can flow to death, D, or recovery, R. At each moment in time, the total population N is divided between the five states: S, I, H, R, D.

The dynamics of the model are given by

$$\frac{dS(t)}{dt} = -\mathcal{R}(t)\gamma I(t), \tag{B.6}$$

$$\frac{dI(t)}{dt} = [\mathcal{R}(t) - 1]\gamma I(t), \tag{B.7}$$

$$\frac{dH(t)}{dt} = \eta\gamma I(t) - \zeta H(t), \tag{B.8}$$

$$\frac{dR(t)}{dt} = (1 - \nu)\zeta H(t) + (1 - \eta)\gamma I(t), \tag{B.9}$$

$$\frac{dD(t)}{dt} = \nu\zeta H(t), \tag{B.10}$$

where the definition of the effective reproduction number, $\mathcal{R}(t)$, is the same as in the SIR and SEIR versions of the model. Note that the parameter ν is now the fatality rate conditional on hospitalization and η is the fraction of the infected population that ends up hospitalized. The parameter ζ determines the duration of hospital stays. For initial conditions, we assume that $D(0) = R(0) = 0$, $S(0)$ is slightly below N, $I(0)$ is slightly above 0, and $H(0)$ may be greater than 0.

We now show how to invert this version of the model to express the effective reproduction number in terms of total deaths and its time derivatives. From equation

we have

$$H(t) = \frac{1}{\nu\zeta}\frac{dD(t)}{dt}.$$

Differentiating this equation gives

$$\frac{dH(t)}{dt} = \frac{1}{\nu\zeta}\frac{d^2D(t)}{dt^2}.$$

These results together with equation give

$$I(t) = \frac{1}{\eta\gamma}\left[\frac{1}{\nu}\frac{dD(t)}{dt} + \frac{1}{\nu\zeta}\frac{d^2D(t)}{dt^2}\right].$$

Integrating and the initial conditions imply

$$R(t) = (1-\nu)\zeta\int_{s=0}^{t} H(s)ds + (1-\nu)\zeta H(0) + (1-\eta)\gamma\int_{s=0}^{t} I(s)ds + (1-\eta)\gamma I(0),$$

or

$$R(t) = \frac{1-\nu}{\nu}D(t) + \frac{1-\eta}{\eta}\left[\frac{1}{\nu}D(t) + \frac{1}{\nu\zeta}\frac{dD(t)}{dt}\right] + (1-\nu)\zeta H(0) + (1-\eta)\gamma I(0).$$

where

$$H(0) = \frac{1}{\nu\zeta}\frac{dD(0)}{dt},$$

and

$$I(0) = \frac{1}{\eta\gamma}\left[\frac{1}{\nu}\frac{dD(0)}{dt} + \frac{1}{\nu\zeta}\frac{d^2D(0)}{dt^2}\right].$$

Using the constraint that categories sum to one gives

$$S(t) = 1 - I(t) - H(t) - R(t) - D(t).$$

Having inverted the model, we can now turn to the effective reproduction number.

Differentiating our expression for $I(t)$ gives

$$\frac{dI(t)}{dt} = \frac{1}{\eta\gamma}\left[\frac{1}{\nu}\frac{d^2D(t)}{dt^2} + \frac{1}{\nu\zeta}\frac{d^3D(t)}{dt^3}\right],$$

and combining these two expressions with equation B.7 yields

$$\frac{1}{\gamma}\left[\frac{d^2D(t)}{dt^2} + \frac{1}{\zeta}\frac{d^3D(t)}{dt^3}\right] = [\mathcal{R}(t) - 1]\left[\frac{dD(t)}{dt} + \frac{1}{\zeta}\frac{d^2D(t)}{dt^2}\right].$$

Thus,

$$\mathcal{R}(t) = 1 + \frac{1}{\gamma}\frac{\left[\frac{d^2D(t)}{dt^2} + \frac{1}{\zeta}\frac{d^3D(t)}{dt^3}\right]}{\left[\frac{dD(t)}{dt} + \frac{1}{\zeta}\frac{d^2D(t)}{dt^2}\right]}.$$

As in the SEIR model, the reproductive ratio depends on daily deaths and both its first and second derivatives. It also depends on the rate at which agents transition out of hospitalization, ζ. We set this rate to $1/7$ such that the average duration of hospital stays is one week consistent with values reported on the CDC website. We choose γ so that our model is consistent with observed doubling times of daily deaths in the early phase of the epidemic when the basic reproduction number in the model is in line with CDC estimates of $\mathcal{R}(0) = 2.5$. If we have daily deaths growing exponentially in the early phase of the epidemic, then

$$\frac{dD(t)}{dt} = \exp(\bar{d} + \delta t),$$

so

$$\mathcal{R}(0) = 1 + \frac{1}{\gamma}\frac{\left[\delta + \frac{1}{\zeta}\delta^2\right]}{\left[1 + \frac{1}{\zeta}\delta\right]} = 1 + \frac{\delta}{\gamma},$$

which is the same expression as for the simple SIR model. Thus, we set $\gamma = 0.2$, the same value we used for the SIR version. With γ set to this value, a 30% daily growth rate of new deaths, $\delta = 0.3$, corresponds to a basic reproduction number of 2.5.

B.3 SEIHR Model

The SEIHR model extends the SIR model by adding both the exposed state E and the hospitalized state H. In this version of the model the total population N is given by the sum of susceptible agents in state S, exposed in state E, infected in I, hospitalized in H, recovered in R, and dead in D. The dynamics of the model are given by

$$\frac{dS(t)}{dt} = -\mathcal{R}(t)\gamma I(t), \tag{B.11}$$

$$\frac{dE(t)}{dt} = \mathcal{R}(t)\gamma I(t) - \sigma E(t), \tag{B.12}$$

$$\frac{dI(t)}{dt} = \sigma E(t) - \gamma I(t), \tag{B.13}$$

$$\frac{dH(t)}{dt} = \eta\gamma I(t) - \zeta H(t), \tag{B.14}$$

$$\frac{dR(t)}{dt} = (1 - \nu)\zeta H(t) + (1 - \eta)\gamma I(t), \tag{B.15}$$

$$\frac{dD(t)}{dt} = \nu\zeta H(t), \tag{B.16}$$

where the effective reproduction number is as defined for the other versions of the model. Initial conditions are $E(0) = R(0) = D(0) = 0$, $S(0)$ slightly below N, $I(0)$ slightly above 0, and $H(0) \geq 0$.

We proceed as before with inverting the model. From equation B.16 we have

$$H(t) = \frac{1}{\nu\zeta}\frac{dD(t)}{dt}.$$

Differentiating this equation gives

$$\frac{dH(t)}{dt} = \frac{1}{\nu\zeta}\frac{d^2 D(t)}{dt^2}.$$

These results together with equation B.14 give

$$I(t) = \frac{1}{\eta\gamma}\left[\frac{1}{\nu}\frac{dD(t)}{dt} + \frac{1}{\nu\zeta}\frac{d^2 D(t)}{dt^2}\right].$$

Integrating B.15 and the initial conditions imply

$$R(t) = (1-\nu)\zeta\int_{s=0}^{t} H(s)ds + (1-\nu)\zeta H(0) + (1-\eta)\gamma\int_{s=0}^{t} I(s)ds + (1-\eta)\gamma I(0),$$

or

$$R(t) = \frac{1-\nu}{\nu}D(t) + \frac{1-\eta}{\eta}\left[\frac{1}{\nu}D(t) + \frac{1}{\nu\zeta}\frac{dD(t)}{dt}\right] + (1-\nu)\zeta H(0) + (1-\eta)\gamma I(0),$$

where

$$H(0) = \frac{1}{\nu\zeta}\frac{dD(0)}{dt},$$

and

$$I(0) = \frac{1}{\eta\gamma}\left[\frac{1}{\nu}\frac{dD(0)}{dt} + \frac{1}{\nu\zeta}\frac{d^2 D(0)}{dt^2}\right].$$

Note that differentiating our expression above for $I(t)$ gives

$$\frac{dI(t)}{dt} = \frac{1}{\eta\gamma}\left[\frac{1}{\nu}\frac{d^2 D(t)}{dt^2} + \frac{1}{\nu\zeta}\frac{d^3 D(t)}{dt^3}\right].$$

Equation B.13 implies that

$$E(t) = \frac{1}{\sigma}\left[\frac{dI(t)}{dt} + \gamma I(t)\right],$$

and combining with our expressions for $I(t0$ and $dI(t)/dt$ we have

$$E(t) = \frac{1}{\sigma\eta\nu}\left[\frac{1}{\gamma}\left[\frac{d^2D(t)}{dt^2} + \frac{1}{\zeta}\frac{d^3D(t)}{dt^3}\right] + \left[\frac{dD(t)}{dt} + \frac{1}{\zeta}\frac{d^2D(t)}{dt^2}\right]\right].$$

Finally, we have

$$S(t) = 1 - E(t) - I(t) - H(t) - R(t) - D(t).$$

In terms of measuring the effective reproduction number, this model shares with the SEIR model that the growth of exposed and infected individuals is determined by

$$\frac{dE(t)}{dt} + \frac{dI(t)}{dt} = [\mathcal{R}(t) - 1]\,\gamma I(t).$$

Differentiating the expression for $E(t)$, plugging it and the expressions for $I(t)$ and $dI(t)/dt$ into the above equation, and rearranging terms gives

$$\mathcal{R}(t) = 1 + \frac{\left(\frac{1}{\sigma} + \frac{1}{\gamma}\right)\frac{d^2D(t)}{dt^2} + \left(\frac{1}{\sigma\gamma} + \frac{1}{\sigma\zeta} + \frac{1}{\gamma\zeta}\right)\frac{d^3D(t)}{dt^3} + \frac{1}{\sigma\gamma\zeta}\frac{d^4D(t)}{dt^4}}{\frac{dD(t)}{dt} + \frac{1}{\zeta}\frac{d^2D(t)}{dt^2}}.$$

Notice that the reproductive ratio now depends not only on the first two derivatives of daily deaths but also the third derivative.

To calibrate the parameters γ, σ, and ζ we proceed as before. The parameter ζ is set to $1/7$ so that the average duration of a hospital stays is 7 days in line with CDC reports. The parameters γ and σ are set to 0.4 and 0.5. These are the same values used in the SEIR version of the model. This combination of parameter values implies that a 30% growth rate of new daily deaths corresponds with a basic reproduction number, $\mathcal{R}(0)$, of 2.8.