medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Tracking $\mathcal{R}$ of COVID-19: A New Real-Time Estimation Using the Kalman Filter*

Francisco Arroyo-Marioli[1], Francisco Bullano[1], Simas Kučinskas[†2], and Carlos Rondón-Moreno[1]

[1] *Central Bank of Chile*
[2] *Humboldt University of Berlin*

First version: April 21, 2020 Last updated: May 11, 2020

Abstract

We develop a new method for estimating the effective reproduction number of an infectious disease ($\mathcal{R}$) and apply it to track the dynamics of COVID-19. The method is based on the fact that in the SIR model, $\mathcal{R}$ is linearly related to the growth rate of the number of infected individuals. This time-varying growth rate is estimated using the Kalman filter from data on new cases. The method is very easy to apply in practice, and it performs well even when the number of infected individuals is imperfectly measured, or the infection does not follow the SIR model. Our estimates of $\mathcal{R}$ for COVID-19 for 124 countries across the world are provided in an interactive online dashboard, and they are used to assess the effectiveness of non-pharmaceutical interventions in a sample of 14 European countries.

*We would like to thank Ralph Brinks, Eric Budish, Ricardo Hausmann, Artūras Juodis, Jan Keil, Siem Jan Koopman, Andrés Neumeyer, Mathieu Pedemonte, Sergio Ocampo-Díaz, Sándor Sóvágó, Iván Werning, as well as seminar participants at the Central Bank of Chile and Harvard Growth Lab for their comments and suggestions. We also thank Chen Lin for excellent research assistance. Replication files are available online. The views and conclusions presented in this paper are exclusively those of the authors and do not necessarily reflect the position of the Central Bank of Chile or of the Board members. The present paper subsumes the previous working papers of Arroyo-Marioli, Bullano and Rondón-Moreno (2020) and Kučinskas (2020).

†Corresponding author: simas.kucinskas@hu-berlin.de.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

1 Introduction

The effective reproduction number ($\mathcal{R}$) plays a central role in the epidemiology of infectious diseases (Chowell and Brauer, 2009; Nishiura and Chowell, 2009; Allen, 2017; Stock, 2020). $\mathcal{R}$ is defined as the average number of people infected by a single infectious individual. In standard models, the number of infected individuals increases as long as $\mathcal{R} > 1$. Real-time estimates of $\mathcal{R}$ are therefore essential for public policy decisions during a pandemic (Atkeson, 2020; Leung, 2020). Such estimates can be used to study the effectiveness of non-pharmaceutical interventions (NPIs), or assess what fraction of the population needs to be vaccinated to reach herd immunity (Chinazzi et al., 2020; Kucharski et al., 2020; Wang et al., 2020). Some social scientists have argued that $\mathcal{R} < 1$ should be viewed as a fundamental constraint on public policy during the current COVID-19 pandemic (Budish, 2020).

In this paper we develop a new method to estimate $\mathcal{R}$ in real time. The method exploits the fact that in the benchmark SIR model, $\mathcal{R}$ is linearly related to the growth rate of the number of infected individuals (Kermack and McKendrick, 1927). Our estimation procedure consists of three steps. First, we use data on new cases to construct a time series of how many individuals are infected at a given point in time. Then, we estimate the growth rate of this time series with the Kalman filter. In the final step, we leverage the theoretical relationship given by the SIR model to obtain $\mathcal{R}$ from the estimated growth rate. We show theoretically that the estimates are not sensitive to potential model misspecification, and they are fairly accurate even when new cases are imperfectly measured.

We apply our methodology to estimate the $\mathcal{R}$ of COVID-19 in real-time. Our estimates for 124 countries across the world are provided in an online dashboard and can be explored interactively (link to dashboard). In empirical applications, we use these estimates to calculate the basic reproduction number ($\mathcal{R}_0$) and evaluate the effects of NPIs in reducing $\mathcal{R}$ for a sample of 14 European countries.

Under our baseline assumption that the serial interval for COVID-19 is seven days, we estimate the basic reproduction number ($\mathcal{R}_0$) to be 2.67 (95% CI: 1.96–3.44). Next, we find that lockdowns, measures of self-isolation, and social distancing all have a statistically significant effect on reducing $\mathcal{R}$. However, we also demonstrate the importance of accounting for voluntary changes in behavior. In particular, we document that most of the decline in mobility in our sample happened before the introduction of lockdowns. Failing to account for voluntary changes in behavior leads to substantially over-estimated effects of NPIs.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Related Literature

There are two broad classes of methods that can be used to estimate $\mathcal{R}$ in real time (Chowell and Brauer, 2009; Nishiura and Chowell, 2009). First, one can estimate a fully-specified epidemiological model and then construct a model-implied time series for $\mathcal{R}$. Second, one may use approaches that leverage information on the serial interval of a disease (i.e., time between onset of symptoms in a case and onset of symptoms in his/her secondary cases). For example, imagine a disease with a fixed serial interval of, say, three days. In that case, we could estimate $\mathcal{R}$ by simply dividing the number of new cases today by the number of new cases three days ago. Cori et al. (2013) exploit this idea to develop a Bayesian estimator that accounts for the randomness in the onset of infections as well as variation in the serial interval; see also Thompson et al. (2019). This method is implemented in a popular R package `EpiEstim`.

The method proposed in this paper attempts to strike a balance between the two approaches mentioned above. Although our estimator is derived from standard epidemiological theory, we use the smallest amount of theoretical structure that is necessary to obtain our estimator. In particular, the theoretical relationship used to derive our estimator is exactly valid not only in the standard SIR model with constant parameters, but also in the SIS model and a generalized SIR model with time-varying parameters and stochastic shocks. Relative to the existing literature, our estimator does not need any statistical tuning parameters, and it does not require parametric assumptions on the distribution of new cases (such as assuming that new cases are Poisson distributed). For example, the method of Cori et al. (2013) assumes that $\mathcal{R}$ is constant over fixed windows of duration τ; τ effectively becomes a tuning parameter that needs to be chosen by the user. Our approach and its mathematical derivation share some similarities with the estimator proposed by Bettencourt and Ribeiro (2008).

A key advantage of using the Kalman filter for estimating $\mathcal{R}$ is that valid confidence bounds are readily obtained. Explicitly accounting for the dynamics in $\mathcal{R}$ is also likely to produce better-behaved estimates in practice. In addition, the Kalman smoother allows the researcher to use full-sample information efficiently when estimating $\mathcal{R}$. Finally, our method can be used with both classical and Bayesian techniques, as we demonstrate in the empirical application.

2 A Real-Time Estimator

We now derive our estimator for the SIR model (Kermack and McKendrick, 1927). In the Appendix, we show that we can obtain the same estimator from an SIS model

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(Appendix A.1), and an SIR model with stochastic shocks (Appendix A.2).

2.1 Deriving the Estimator

The standard SIR model in discrete time describes the evolution of susceptible (S_t), infected (I_t), and recovered (R_t) individuals by the following equations (Allen and Van Den Driessche, 2008; Stock, 2020):

$$\begin{aligned} S_t &= S_{t-1} - \beta_t I_{t-1} \frac{S_{t-1}}{N} \\ I_t &= I_{t-1} + \beta_t I_{t-1} \frac{S_{t-1}}{N} - \gamma I_{t-1} \\ R_t &= R_{t-1} + \gamma I_{t-1} \end{aligned} \tag{1}$$

The model is stated at a daily frequency. Here, $N \equiv S_t + I_t + R_t$ is the population size, β_t is the daily transmission rate, and γ is the daily transition rate from infected to recovered. The recovered group consists of individuals who have either died or fully recovered. We allow the transmission rate β_t to vary over time. For example, individuals may choose to to reduce their social interactions voluntarily, or they could be subject to government policy restrictions.

The *basic reproduction number*, $\mathcal{R}_0^{(t)}$, is defined as $\mathcal{R}_0^{(t)} \equiv \beta_t/\gamma$, and it gives the average number of individuals infected by a single infected individual when everyone else is susceptible. Since the transmission rate β_t varies over time, the basic reproduction number is generally time varying as well. The *effective reproduction number*, $\mathcal{R}_t$, is defined as $\mathcal{R}_t = \mathcal{R}_0^{(t)} \times (S_{t-1}/N)$, and it equals the average number of individuals infected by a single infected individual when a fraction (S_{t-1}/N) of individuals is susceptible.

From Eq. (1) the daily growth rate in the number of infected individuals is

$$\text{gr}(I_t) \equiv \frac{I_t - I_{t-1}}{I_{t-1}} = \gamma(\mathcal{R}_t - 1). \tag{2}$$

Denoting the estimated growth rate of infected individuals by $\hat{\text{gr}}(I_t)$, and given a value for the transition rate γ, the plug-in estimator for the effective reproduction number is

$$\hat{\mathcal{R}}_t = 1 + \frac{1}{\gamma}\hat{\text{gr}}(I_t). \tag{3}$$

For the estimator to be feasible, we need to (i) calibrate the transition rate from infectious to recovered, γ; and (ii) estimate the growth rate of I_t. There are two potential strategies for choosing γ. First, we can use external medical evidence given that γ^{-1} is

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

the average infectious period. Second, information on the serial interval of the disease can be employed, given that the serial interval in the SIR model also equals γ^{-1} (Ma, 2020).

To estimate the growth rate of I_t empirically, we first construct a time series for I_t from data on new cases. The SIR model in Eq. (1) implies that

$$I_t = (1-\gamma)I_{t-1} + \text{new cases}_t. \tag{4}$$

We initialize I_t by $I_0 = C_0$ where C_0 is the total number of infectious cases at some initial date, and then construct subsequent values of I_t recursively.

Given the time series for I_t, we use standard Kalman-filtering tools to smooth the observed growth rate of I_t. In particular, we specify the following state-space model for the growth rate of I_t:

$$\begin{aligned} \text{gr}(I_t) &= \gamma(\mathcal{R}_t - 1) + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_\varepsilon^2) \\ \mathcal{R}_t &= \mathcal{R}_{t-1} + \eta_t, \qquad \eta_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_\eta^2) \end{aligned} \tag{5}$$

We estimate $\mathcal{R}_t$ by the Kalman smoother (see Durbin and Koopman, 2012, Chapter 2). The Kalman smoother provides optimal estimates of $\mathcal{R}_t$ (in the sense of minimizing mean-squared error) given the full-sample information on $\text{gr}(I_t)$, provided that the data are generated by the model in Eq. (5).

To estimate the unknown parameters σ_ε^2 and σ_η^2 in Eq. (5), both classical and Bayesian methods can be used. However, sample sizes are usually limited in practice, especially early on in the epidemic. Hence, incorporating prior knowledge generally leads to better-behaved estimates. The state-space model above—also known as the local-level model—can also be thought as a model-based version of exponentially-weighted moving-average smoothing (Muth, 1960).

The state-space model in Eq. (5) can be viewed as a reduced-form time-series specification. As is well known, the local-level model is sufficiently flexible to capture rich dynamic patterns in the data. In addition, in Appendix A.3, we provide a theoretical rationale for the local-level specification. In particular, Eq. (5) arises naturally in the SIR model (in the early stages of an epidemic) when the transmission rate β_t follows a random walk.

From Eq. (4), the growth rate $\text{gr}(I_t)$ is bounded below by $(-\gamma)$. Hence, for any estimator of $\text{gr}(I_t)$ that is some weighted average of the observed growth rates, the point estimate of $\mathcal{R}_t$ is automatically non-negative. To ensure that lower confidence bounds are positive as well, we estimate the q-th quantile of $\mathcal{R}_t$ by $\max\{0, 1+\gamma^{-1}\hat{g}_q\}$, where $\hat{g}_q$ is an estimate of the q-th quantile of $\text{gr}(I_t)$. In addition (see Appendix A.4), our em-

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

pirical estimates remain similar when we use a modified version of the Carter and Kohn (1994) algorithm which discards random draws violating the non-negativity constraint. Alternatively, it is possible to avoid this type of truncation by using non-linear filtering methods (Creal, 2012).

2.2 Sensitivity to Model Misspecification and Data Problems

Tracking the evolution of $\mathcal{R}_t$ is notoriously difficult. Human-contact dynamics, testing, and changes in case definitions affect the flow and quality of the available information. In this section, we test the sensitivity of our estimator to two notable issues: (i) model misspecification; and (ii) data problems (such as reporting delays or imperfect detection of infectious individuals).

For the first issue, model misspecification, a natural concern is whether the true dynamics of the disease are well captured by the benchmark SIR model. We address this issue in two ways. First, we show that our estimator remains exactly valid in the SIS model in which individuals do not obtain immunity (Appendix A.1) and a generalized SIR model with stochastic shocks (Appendix A.2). In addition, provided that the average duration of infectiousness is correctly specified, we find that our estimator yields accurate results even when the true model is SEIR rather than SIR (Appendix A.5). Second, we note that the error term ε_t in the state-space model described by Eq. (5) can be interpreted as model error. Therefore, our estimates as well as their confidence intervals explicitly account for (some amount of) potential misspecification.

The second issue relates to data reliability. For COVID-19, testing constraints and high asymptomatic prevalence (Arons et al., 2020; Nishiura et al., 2020a; Streeck et al., 2020), in particular, make it challenging to identify all infectious individuals. The simplicity of our estimator allows us to analytically characterize the effects of potential measurement error (see Appendix A.6). Furthermore, we use these results to investigate the quantitative performance of the estimator in a number of empirically relevant underdetection scenarios using Monte Carlo simulations. Overall, we conclude that our method provides accurate estimates in all cases that we analyze.

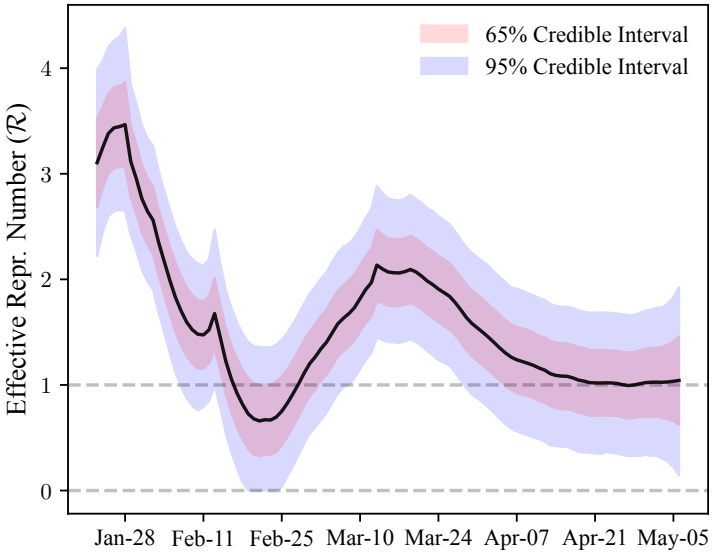
3 Estimates for COVID-19

We now use data from the John Hopkins CSSE repository to obtain real-time estimates of $\mathcal{R}_t$ for COVID-19 (Dong et al., 2020). For each country, we use data after the cumulative number of COVID-19 cases reaches 100. Appendix A.7 describes the details of the estimation procedure.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 1

$\mathcal{R}_t$ of COVID-19: Global Estimates



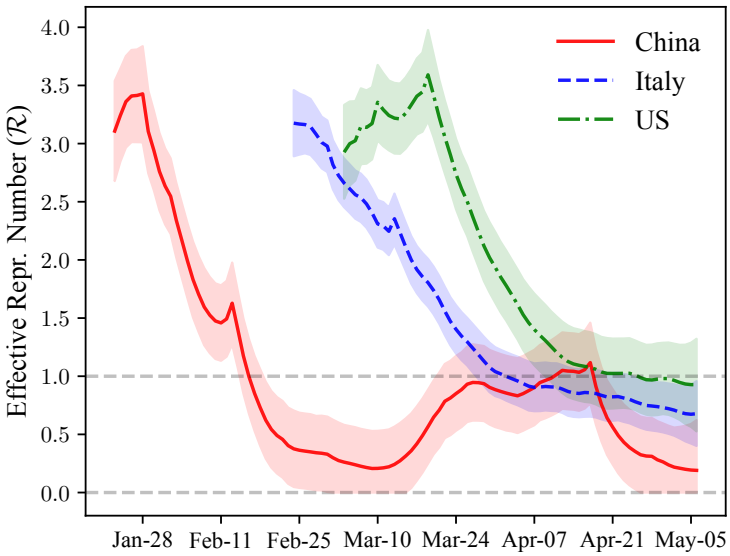
Notes: Estimates of the effective reproduction rate ($\mathcal{R}_t$) of COVID-19 for the world as whole. The sample consists of all dates after the total number of reported cases worldwide has reached 100. 65% and 95% credible bounds shown by the shaded areas.

For the baseline estimates, we assume that people are infectious for $\gamma^{-1} = 7$ days on average, similarly to Maier and Brockmann (2020) and Prem et al. (2020). This assumption is consistent with the evidence on the serial interval of COVID-19. For example, Flaxman et al. (2020) use an average serial interval of 6.5 days. Recent studies find that estimates of the serial interval for COVID-19 generally range between 4 and 9 days (Nishiura et al., 2020b; Park et al., 2020; Sanche et al., 2020). In addition, we document that $\gamma^{-1} = 7$ leads to estimates of the basic reproduction number ($\mathcal{R}_0$) that are in line with the recent estimates in the literature (Liu et al., 2020). However, we also investigate the effects of different choices for γ on our results.

In Appendix A.8, we perform two empirical validation exercises of our estimates. First, we document that our estimates of $\mathcal{R}_t$ are predictive of future deaths. Given that deaths are arguably more accurately measured, this finding alleviates concerns regarding potential data reliability issues that could contaminate our estimates. Second, we find that past mobility data is predictive of future values of $\mathcal{R}_t$. Jointly, these two exercises suggest that the estimates contain valuable information on the dynamics of COVID-19.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 2
$\mathcal{R}_t$ of COVID-19: Selected Countries



Notes: Estimates of the effective reproduction rate ($\mathcal{R}_t$) of COVID-19 for selected countries. The sample consists of all dates after the total number of reported cases in the country has reached 100. 65% credible bounds shown by the shaded areas.

3.1 Estimated Effective Reproduction Numbers

Our estimates of $\mathcal{R}_t$ for COVID-19 for the world as a whole are provided in Figure 1. The graph highlights two distinct phases of the pandemic, first in China and later in Europe and the US. At the beginning of the sample in late January, the estimates of $\mathcal{R}_t$ are above 3; the estimates fall below one by February 19, consistent with a containment of the epidemic in China. Note that there is a moderate upwards jump in the estimated $\mathcal{R}_t$ around the second week of February. This jump was caused by a temporary change in COVID-19 case definitions in the Hubei province in China; the new definition included clinically-diagnosed COVID-19 cases (Tsang et al., 2020). The estimates of $\mathcal{R}_t$ start increasing around February 19, coinciding with the spread of the pandemic to Europe and the US. Our estimates indicate that $\mathcal{R}_t$ has been trending down since mid-March globally. However, the rate of decline is substantially lower than what was observed in late January and early February. The current point estimate of $\mathcal{R}_t$ for the world as a whole (as of May 6, 2020) is very close to one.

Figure 2 plots the estimated $\mathcal{R}_t$ for China, Italy, and the US. In the Appendix (Figure A.7), we also provide a graph of the raw data that is used for estimating $\mathcal{R}_t$. For all three countries, the estimated $\mathcal{R}_t$ is initially above 3. For China, the estimated $\mathcal{R}_t$

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

falls below one around the third week of February. However, the estimated $\mathcal{R}_t$ in China drifted up towards one during late March and early April, potentially caused by a wave of imported cases; the current point estimate for China is substantially below one. In Italy, the estimated $\mathcal{R}_t$ has been steadily falling since March but at a slower rate than previously observed in China, with the point estimate for Italy falling below one in early April. In the US, a striking difference is that the point estimates of $\mathcal{R}_t$ are, in fact, increasing in the first two weeks of the epidemic, rising from around 2.9 to almost 3.6. A likely explanation is that the fraction of non-detected cases went down substantially in this period, inflating the estimates of $\mathcal{R}_t$ upward (see Appendix A.6). In particular, the number of tests conducted in the U.S went up substantially during this period (Our World in Data, 2020). The confidence bounds, however, are consistent with $\mathcal{R}_t$ in the US being stable over the first two weeks of the epidemic. The current point estimate of $\mathcal{R}_t$ in the US is somewhat below one.

In the Appendix (Figure A.8), we illustrate the difference between estimates of $\mathcal{R}_t$ for China obtained by the Kalman smoother—as in our baseline estimation—and the Kalman filter. Intuitively, the Kalman smoother uses information from the full sample when estimating $\mathcal{R}_t$, while the Kalman filter only uses information up to and including time t (Durbin and Koopman, 2012). As seen in the graph, while the two sets of estimates are fairly similar, the filtered estimates are substantially more volatile. In addition, the filtered estimates generally have wider credible bounds. As should be the case, the filtered and smoothed estimates are identical at the endpoint of the sample. From the perspective of epidemiological theory, the Kalman filter essentially produces what Fraser (2007) refers to as the instantaneous reproduction number, while the Kalman smoother yields the case reproduction number. The estimator proposed in the present paper therefore allows researchers to estimate the two types of reproduction numbers in a single unified framework.

In Figure A.8 in the Appendix, we also demonstrate the difference between our Bayesian estimates of $\mathcal{R}_t$ and classical estimates obtained via maximum likelihood. For China, the two sets of estimates are virtually indistinguishable, indicating that the chosen priors have a small effect on the estimates. Of course, for other some countries in our sample, the data are less informative, and hence the priors have a more pronounced effect.

3.2 Basic Reproduction Number

We now use our estimates of $\mathcal{R}_t$ to measure the basic reproduction number ($\mathcal{R}_0$), i.e., the average number of individuals infected by a single infectious individual when the

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 1
Estimates of the Basic Reproduction Number ($\mathcal{R}_0$)

Number of Days Infectious:	5	6	7	8	9	10
$\hat{\mathcal{R}}_0$	2.08	2.35	2.67	2.98	3.28	3.58
CI Lower Bound (95%)	1.50	1.72	1.96	2.19	2.42	2.66
CI Upper Bound (95%)	2.71	3.06	3.44	3.81	4.18	4.55

Notes: Estimates of the basic reproduction number ($\mathcal{R}_0$) for a sample of 14 European countries. The countries included in the sample are Austria, Belgium, Denmark, France, Germany, Greece, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and United Kingdom. The basic reproduction number is calculated by averaging our estimates of the effective reproduction number in the first 7 days of the epidemic, where the start of the epidemic is defined as the day when the cumulative number of cases reaches 100.

population is fully susceptible. We estimate $\mathcal{R}_0$ by the average value of $\mathcal{R}_t$ in the first week of the epidemic.

Table 1 shows the results for a sample of 14 European countries (Austria, Belgium, Denmark, France, Germany, Greece, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and United Kingdom), as in Flaxman et al. (2020). Under our baseline assumption that the individuals are infectious for 7 days on average ($\gamma = 1/7$), we obtain an estimate of $\mathcal{R}_0 = 2.67$ (95% CI: 1.96–3.44). For COVID-19, a recent meta-study has estimated a median $\mathcal{R}_0$ of 2.79 (Liu et al., 2020), suggesting that our results are consistent with the current consensus estimates.

Table 1 also provides the estimated $\mathcal{R}_0$ under different assumptions on the duration of infectiousness (or, equivalently in the SIR model, the average serial interval). As expected, the median estimate is sensitive to the choice of γ; we find an additional day of infectiousness increases $\mathcal{R}_0$ by around 0.3.

3.3 Assessing Non-Pharmaceutical Interventions

Finally, we use our estimates to assess the effects of non-pharmaceutical interventions (NPIs) in the same sample of 14 European countries as in Section 3.2. We study a total of five NPIs: (i) lockdowns; (ii) bans of public events; (iii) school closures; (iv) mandated self-isolation when exhibiting symptoms; and (e) social distancing measures. We adopt the definitions of NPIs and their introduction dates provided by Flaxman et al. (2020).

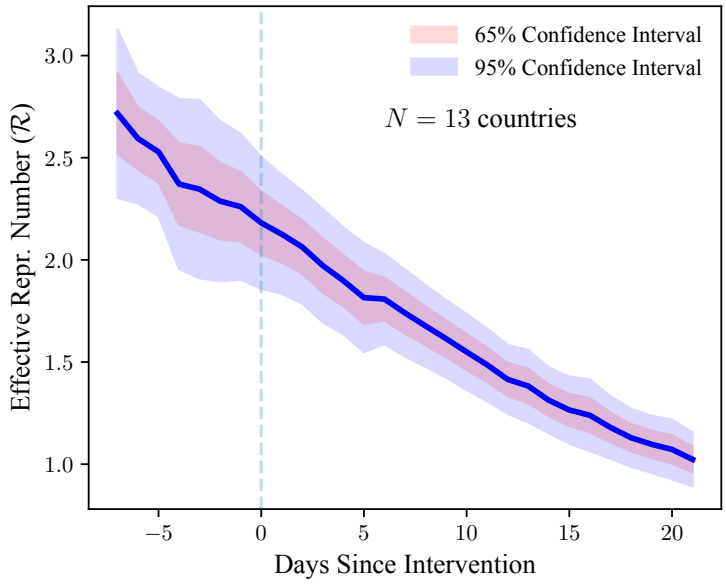
We first perform an event-study analysis. Figure 3a plots the estimated values of $\mathcal{R}_t$ around the introduction of a lockdown. $\mathcal{R}_t$ declines substantially after a lockdown is introduced, going from around 2.18 on the day of the intervention to around 1.31

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

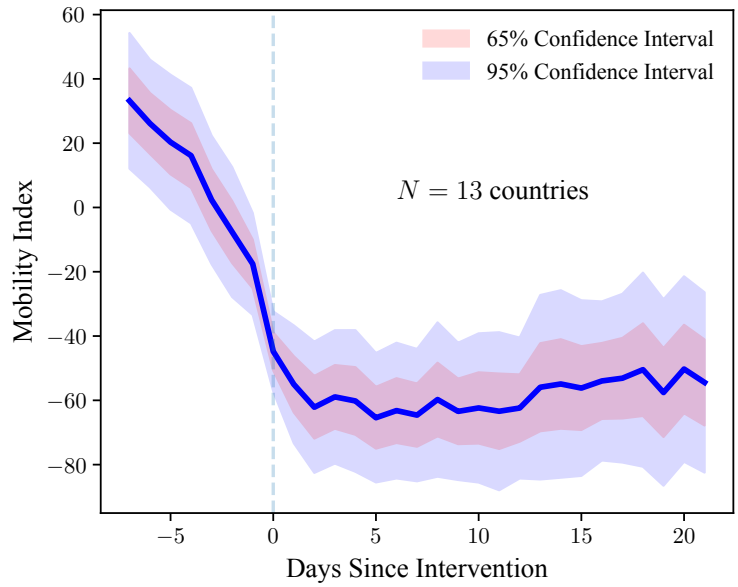
Figure 3

Event Study: $\mathcal{R}$ and Mobility Around Lockdowns

(a) Effective Reproduction Number



(b) Mobility Index



Notes: *Top panel*: estimated effective reproduction number ($\mathcal{R}_t$) one week before and three weeks after a lockdown is introduced in a country. *Bottom panel:* mobility index (constructed from "COVID-19 Community Mobility Reports" of Google (2020)) one week before and three weeks after a lockdown is introduced in a country. See Appendix A.8 for details on the construction of the mobility index. The original sample consists of 14 European countries studied by Flaxman et al. (2020). Heteroskedasticity-robust confidence bounds are shown by the shaded areas.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

two weeks later. However, $\mathcal{R}_t$ is decreasing before the lockdown as well. In particular, there is no visually detectible break in the slope of $\mathcal{R}_t$ after the lockdown (i.e., no "kink"). In the Appendix, we show that other NPIs follow a similar pattern. In particular, we document the behavior of $\mathcal{R}_t$ around the introduction of public-event bans (Figure A.10), case-based measures (such as self-isolation whenever feeling ill and experiencing fever; Figure A.11), school closures (Figure A.12), and social-distancing measures (Figure A.13). Except for school closures and public-event bans, there is no visually apparent break in the trend of $\mathcal{R}_t$ around the date of the policy intervention.

To investigate why $\mathcal{R}_t$ appears to be unaffected by lockdowns, we use mobility data from Google's "COVID-19 Community Mobility Reports" (Google, 2020). Figure 3b shows that most of the decline in mobility occurs *before* the imposition of the lockdown, and remains low *thereafter*. This finding shows a clear change in people's behavior in the early days of the pandemic. Shifting habits before the introduction of NPIs is consistent with the existence of private motives that can induce a reduction in mobility as people avoid becoming infected (Farboodi et al., 2020; Guerrieri et al., 2020; Krüger et al., 2020). Our results are also consistent with empirical evidence for the U.S and anecdotal reports from Sweden (Farboodi et al., 2020; The New York Times, 2020). The documented relationship between $\mathcal{R}_t$ and mobility does not necessarily constitute evidence against the effectiveness of lockdowns. On the contrary, it is possible that lockdowns reinforce attitudes towards disease-awareness and self-isolation, helping to ensure lower values of $\mathcal{R}_t$ in the long run.

A potential concern with the evidence in Figure 3a is that our estimates of $\mathcal{R}_t$ use information from the full sample. Hence, estimates of $\mathcal{R}_t$ *before* the lockdown implicitly depend on the estimates of $\mathcal{R}_t$ *after* the lockdown. This feature of the estimation procedure may result in low statistical power to detect any effects of NPIs. To investigate this possibility, we conduct a power analysis (Appendix A.9). Given our empirical estimates of signal-to-noise ratios, we find that the statistical procedure appears sufficiently powerful to detect moderate changes in $\mathcal{R}_t$.

To assess the effects of NPIs more formally, we employ the following fixed-effect regressions (Table 2). Specifically, we regress $\mathcal{R}_t$ on a set of indicator variables capturing interventions and different types of fixed effects:

$$\log(R_{i,t}) = (\text{fixed effects}) + \sum_{j=1}^{5} \beta_i \text{NPI}_{i,t}^{(j)} + u_{i,t}.$$

Here, $\text{NPI}_{i,t}^{(j)}$ is an indicator variable that equals 1 after the j-th NPI is introduced, and zero before its introduction. The index i denotes countries, and t stands for the number

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 2

Effective Reproduction Number After Introduction of NPIs

	(1)	(2)	(3)	(4)
Lockdown	-0.63***	-0.09*	-0.06	-0.04
	(0.05)	(0.05)	(0.05)	(0.06)
Public Events	-0.09**	0.21***	0.2***	0.38***
	(0.04)	(0.04)	(0.04)	(0.07)
School Closure	0.01	-0.17***	-0.15***	-0.09
	(0.06)	(0.05)	(0.05)	(0.08)
Self Isolation	-0.14**	-0.11**	-0.06	-0.01
	(0.06)	(0.05)	(0.04)	(0.07)
Social Distancing	-0.18***	-0.1*	-0.11**	-0.13*
	(0.06)	(0.06)	(0.06)	(0.07)
N	840	840	768	445
R^2	0.54	0.88	0.88	0.91
Country FE	✓	✓	✓	✓
Days-Since-Outbreak FE		✓	✓	✓
Mobility Controls			✓	✓
Testing Controls				✓

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Notes: Results of panel-data regressions of the (log of) effective reproduction number ($\mathcal{R}_t$) on indicator variables that are equal to 1 after the introduction of a non-pharmaceutical intervention (NPI) and 0 before the introduction. The sample consists of 14 European countries studied by Flaxman et al. (2020). Regressions always include country fixed effects; regressions in columns (2)–(4) also include days-since-outbreak fixed effects. Outbreak is defined as the date on which 100 cases of COVID-19 are reached. The regression with mobility controls in (3) includes the one- and two-week lags of the mobility index, constructed from Google (2020); see Appendix A.8 for details. The regression with testing controls in (4) controls for the change in the number of daily tests per capita conducted in the country. To allow for reasonably precise estimation of days-since-outbreak fixed effects, we only consider days after the outbreak for which we have data for at least 5 countries. Heteroskedasticity-robust standard errors in parentheses.

of days since the outbreak of the epidemic.

Column (1) of Table 2 provides estimated effects of NPIs when only country fixed effects are included. We observe a strong negative effect of lockdowns, social distancing, and measures of self isolation. Taken at face value, the estimates suggest that lockdowns reduce $\mathcal{R}_t$ by almost 60%. Bans of public events and school closures are not statistically

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

significant in this specification. These regressions as well the point estimates are similar to the statistical analysis performed by Flaxman et al. (2020).

The regression with country fixed effects only, however, is likely misspecified. Implicitly, such a specification assumes that the only reason why $\mathcal{R}_t$ can fall is because of introduction of NPIs. However, $\mathcal{R}_t$ would likely trend downwards even in the absence of any public policy interventions. First, $\mathcal{R}_t$ tends to fall during an epidemic as the number of susceptibles is depleted. Second, people may adjust their behavior even in the absence of any policy measures. Failing to control for the dynamics of $\mathcal{R}_t$ in the absence of NPIs therefore likely leads to an over-estimation of the effects of NPIs.

We acknowledge that obtaining credible counterfactuals in the present empirical context is extremely challenging. However, we can exploit the panel structure of the dataset to reduce the potential endogeneity problems in the previous specification. We do so by including days-since-outbreak fixed effects. Intuitively, with such fixed effects we are comparing $\mathcal{R}_t$'s in two countries (e.g., country A and country B) that are both five days from the outbreak (say), with a school closure in country A but not in country B.

The results from the regression with days-since-outbreak fixed effects are shown in column (2). The coefficient for lockdowns becomes substantially smaller in absolute value and only marginally statistically significant. The coefficients for self-isolation and social-distancing measures are also reduced and lose some of their statistical significance. The coefficient for public events is highly statistically significant but positive rather than negative. A naïve interpretation would suggest that banning public events has a positive effect on $\mathcal{R}_t$. More likely, however, is that the positive coefficient is due to countries where $\mathcal{R}_t$ is declining more slowly being faster to ban public events. In the Appendix (Table A.3), we show that the results remain similar when the NPIs are included separately, reducing concerns about potential multicollinearity problems between the different NPI variables.

In column (3), we also include lagged mobility variables as additional controls. With mobility controls, the coefficient on lockdowns loses statistical significance altogether. School closures, self isolation, and social-distancing measures are estimated to have a statistically-significant negative effect on $\mathcal{R}_t$, with each of these NPIs estimated to reduce $\mathcal{R}_t$ by around 10%.

A potential concern is that countries may introduce NPIs and simultaneously increase the number of tests for COVID-19 that they perform. To help alleviate this concern, in column (4) we add the change in the daily number of tests per capita as an additional explanatory variable. The data on daily tests per capita comes from *Our World in Data* (Our World in Data, 2020). While the sample size is reduced significantly as we

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

do not have testing data for all countries in the sample, the results are largely unchanged.

We caution readers against over-interpreting the results of this section. Obtaining unbiased estimates of the true causal impact of NPIs is exceptionally challenging. As a result, even our best estimates still suffer from important potential endogeneity issues. In particular, the timing of NPIs is not random. Countries that introduced NPIs earlier likely did so because they had previously observed a stubbornly high $\mathcal{R}_t$. Finally, we find that people adjusted their mobility patterns *before* the introduction of lockdowns. We believe that these findings bolster the importance of taking changes in human behavior into account when evaluating the effects of NPIs.

4 Conclusions and Limitations

In this paper we develop a new way to estimate the effective reproduction number of an infectious disease ($\mathcal{R}$). The new methodology is straightforward to apply in practice, and according to our simulation checks, it yields accurate estimates. We use the new method to track $\mathcal{R}$ of COVID-19 around the world, and assess the effectiveness of public policy interventions in a sample of European countries.

The current paper faces several limitations. First, a local-level specification for the growth rate implicitly assumes that the growth rate of the number of infected individuals remains forever in flux. However, in the long-run, this growth rate must converge to zero. Since our model does not capture this feature, it seems likely that our estimated confidence bounds are overly conservative in the late stages of an epidemic. Second, when applying the model to cross-country data, one may achieve important gains in statistical efficiency if the model is estimated jointly for all countries (for example, by estimating a multivariate local-level model). Finally, for assessing the effects of NPIs more accurately, it would be desirable to collect data for a larger sample of countries.

In our empirical application, we find that lockdowns, measures of self-isolation, and social distancing all have statistically significant effects on reducing $\mathcal{R}$. Self-isolation, in particular, appears to have a strong effect across all statistical specifications that we consider. However, we also demonstrate the importance of accounting for voluntary changes in behavior. In particular, most of the decline in mobility in our sample took place before lockdowns were introduced. This finding suggests that people respond to the risk of contracting the virus by changing their mobility patterns and reducing social interactions. Failing to account for such voluntary changes in behavior yields estimated effects of NPIs that are arguably too large.

Given that even our best estimates still suffer from potential endogeneity problems, it is important to interpret these results cautiously. However, from an economic perspec-

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

tive, these findings point to large private incentives to avoid infection. These incentives can induce a contraction in economic activity as people voluntarily choose to self-isolate (Farboodi et al., 2020; Guerrieri et al., 2020; Krüger et al., 2020). As a result, even if countries lift the NPIs that are currently in place, it is not clear whether people would voluntarily return to their pre-pandemic mobility and consumption patterns. Our real-time estimator may be used to track the dynamics of COVID-19 as the current restrictions are relaxed.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References

Allen, L.J., 2017. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. Infectious Disease Modelling 2, 128–142.

Allen, L.J., Van Den Driessche, P., 2008. The basic reproduction number in some discrete-time epidemic models. Journal of Difference Equations and Applications 14, 1127–1147.

Arons, M.M., Hatfield, K.M., Reddy, S.C., Kimball, A., James, A., Jacobs, J.R., Taylor, J., Spicer, K., Bardossy, A.C., Oakley, L.P., Tanwar, S., Dyal, J.W., Harney, J., Chisty, Z., Bell, J.M., Methner, M., Paul, P., Carlson, C.M., McLaughlin, H.P., Thornburg, N., Tong, S., Tamin, A., Tao, Y., Uehara, A., Harcourt, J., Clark, S., Brostrom-Smith, C., Page, L.C., Kay, M., Lewis, J., Montgomery, P., Stone, N.D., Clark, T.A., Honein, M.A., Duchin, J.S., Jernigan, J.A., 2020. Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. New England Journal of Medicine , 1–10.

Arroyo-Marioli, F., Bullano, F., Rondón-Moreno, C., 2020. Dynamics of transmission and control of COVID-19: A real-time estimation using the Kalman filter. Unpublished manuscript .

Atkeson, A., 2020. What will be the economic impact of COVID-19 in the US? Rough estimates of disease scenarios. NBER Working Paper 26867.

Bettencourt, L.M., Ribeiro, R.M., 2008. Real time Bayesian estimation of the epidemic potential of emerging infectious diseases. PLoS ONE 3, 1–9.

Budish, E., 2020. R < 1 as an economic constraint: Can we "expand the frontier" in the fight against Covid-19? http://dx.doi.org/10.2139/ssrn.3567068 .

Carter, C.K., Kohn, R., 1994. On Gibbs sampling for state space models. Biometrika 81, 541–553.

Chinazzi, M., Davis, J.T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Pastore Y Piontti, A., Mu, K., Rossi, L., Sun, K., Viboud, C., Xiong, X., Yu, H., Halloran, M.E., Longini, I.M., Vespignani, A., 2020. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 9757, 1–12.

Chowell, G., Brauer, F., 2009. The basic reproduction number of infectious diseases: Computation and estimation using compartmental epidemic models, in: Chowell, G., Hyman, J.M., Bettencourt, L.M.A., Castillo-Chavez, C. (Eds.), Mathematical and Statistical Estimation Approaches in Epidemiology. Springer. chapter 1, pp. 1–30.

Cori, A., Ferguson, N.M., Fraser, C., Cauchemez, S., 2013. A new framework and software to estimate time-varying reproduction numbers during epidemics. American Journal of Epidemiology 178, 1505–1512.

Creal, D., 2012. A survey of sequential Monte Carlo methods for economics and finance. Econometric Reviews 31, 245–296.

Dong, E., Du, H., Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. The Lancet 3099, 19–20.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Durbin, J., Koopman, S.J., 2012. Time Series Analysis by State Space Methods. Oxford University Press, Oxford.

Farboodi, M., Jarosch, G., Shimer, R., 2020. Internal and external effects of social distancing in a pandemic. NBER Working Paper 27059.

Ferguson, N.M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá, Z., Cuomo-Dannenburg, G., Dighe, A., Fu, H., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Okell, L.C., Van, S., Thompson, H., Verity, R., Volz, E., Wang, H., Wang, Y., Walker, P.G.T., Walters, C., Winskill, P., Whittaker, C., Donnelly, C.A., Riley, S., Ghani, A.C., 2020. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College London .

Flaxman, S., Mishra, S., Gandy, A., Unwin, J.T., Coupland, H., Mellan, T.A., Zhu, H., Berah, T., Eaton, J.W., Guzman, P.N.P., Schmit, N., Cilloni, L., Ainslie, K.E.C., Baguelin, M., Blake, I., Boonyasiri, A., Boyd, O., Cattarino, L., Ciavarella, C., Cooper, L., Cucunubá, Z., Cuomo-Dannenburg, G., Dighe, A., Djaafara, B., Dorigatti, I., Van Elsland, S., Fitzjohn, R., Fu, H., Gaythorpe, K., Geidelberg, L., Grassly, N., Green, W., Hallett, T., Hamlet, A., Hinsley, W., Jeffrey, B., Jorgensen, D., Knock, E., Laydon, D., Nedjati-Gilani, G., Nouvellet, P., Parag, K., Siveroni, I., Thompson, H., Verity, R., Volz, E., Gt Walker, P., Walters, C., Wang, H., Wang, Y., Watson, O., Xi, X., Winskill, P., Whittaker, C., Ghani, A., Donnelly, C.A., Riley, S., Okell, L.C., Vollmer, M.A.C., Ferguson, N.M., Bhatt, S., 2020. Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. https://doi.org/10.25561/77731.

Fraser, C., 2007. Estimating individual and household reproduction numbers in an emerging epidemic. PLoS ONE 2.

Gelman, A., Lee, D., Guo, J., 2015. Stan: A probabilistic programming language for bayesian inference and optimization. Journal of Educational and Behavioral Statistics 40, 530–543.

Google, 2020. Google COVID-19 Community Mobility Reports. https://www.google.com/covid19/mobility/.

Guerrieri, V., Lorenzoni, G., Straub, L., Werning, I., 2020. Macroeconomic implications of COVID-19: Can negative supply shocks cause demand shortages? NBER Working Paper 26918. arXiv:arXiv:1011.1669v3.

Hortaçsu, A., Liu, J., Schwieg, T., 2020. Estimating the fraction of unreported infections in epidemics with a known epicenter: An application to COVID-19. Working paper .

Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society A 115, 700–721.

Krüger, D., Uhlig, H., Xie, T., 2020. Macroeconomic dynamics and reallocation in an epidemic. Becker Friedman Institute Working Paper 2020-43.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Kucharski, A.J., Russell, T.W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R.M., Sun, F., Jit, M., Munday, J.D., Davies, N., Gimma, A., van Zandvoort, K., Gibbs, H., Hellewell, J., Jarvis, C.I., Clifford, S., Quilty, B.J., Bosse, N.I., Abbott, S., Klepac, P., Flasche, S., 2020. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. The Lancet Infectious Diseases 3099, 1–7.

Kučinskas, S., 2020. Tracking *R* of COVID-19. Unpublished manuscript .

Lauer, S.A., Grantz, K.H., Bi, Q., Jones, F.K., Zheng, Q., Meredith, H.R., Azman, A.S., Reich, N.G., Lessler, J., 2020. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. Annals of Internal Medicine .

Leung, G., 2020. Lockdown can't last forever. Here's how to lift it. The New York Times (https://nyti.ms/3dWXHZR).

Liu, Y., Gayle, A.A., Wilder-Smith, A., Rocklöv, J., 2020. The reproductive number of COVID-19 is higher compared to SARS coronavirus. Journal of Travel Medicine 27, 1–4.

Ma, J., 2020. Estimating epidemic exponential growth rate and basic reproduction number. Infectious Disease Modelling 5, 129–141.

Maier, B.F., Brockmann, D., 2020. Effective containment explains sub-exponential growth in confirmed cases of recent COVID-19 cases in China. Science 4557, 1–8. arXiv:2002.07572.

Muth, J.F., 1960. Optimal properties of exponentially weighted forecasts. Journal of the American Statistical Association 55, 299–306.

Nishiura, H., Chowell, G., 2009. The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends, in: Chowell, G., Hyman, J.M., Bettencourt, L.M.A., Castillo-Chavez, C. (Eds.), Mathematical and Statistical Estimation Approaches in Epidemiology. Springer. chapter 5, pp. 103–122.

Nishiura, H., Kobayashi, T., Miyama, T., Suzuki, A., mok Jung, S., Hayashi, K., Kinoshita, R., Yang, Y., Yuan, B., Akhmetzhanov, A.R., Linton, N.M., 2020a. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). International Journal of Infectious Diseases 94, 154–155.

Nishiura, H., Linton, N.M., Akhmetzhanov, A.R., 2020b. Serial interval of novel coronavirus (COVID-19) infections. International Journal of Infectious Diseases 93, 284–286.

Our World in Data, 2020. Coronavirus disease (COVID-19) – statistics and research. https://ourworldindata.org/coronavirus.

Park, M., Cook, A.R., Lim, J.T., Sun, Y., Dickens, B.L., 2020. A systematic review of COVID-19 epidemiology based on current evidence. Journal of Clinical Medicine 9, 967.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Prem, K., Liu, Y., Russell, T.W., Kucharski, A.J., Eggo, R.M., Davies, N., Jit, M., Klepac, P., Flasche, S., Clifford, S., Pearson, C.A.B., Munday, J.D., Abbott, S., Gibbs, H., Rosello, A., Quilty, B.J., Jombart, T., Sun, F., Diamond, C., Gimma, A., van Zandvoort, K., Funk, S., Jarvis, C.I., Edmunds, W.J., Bosse, N.I., Hellewell, J., 2020. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. The Lancet Public Health 2667, 1–10.

Sanche, S., Lin, Y.T., Xu, C., Romero-Severson, E., Hengartner, N., Ke, R., 2020. High contagiousness and rapid spread of Severe Acute Respiratory Syndrome Coronavirus 2. Emerging Infectious Diseases 26.

Shephard, N., 2015. Martingale unobserved component models, in: Koopman, S.J., Shephard, N. (Eds.), Unobserved Components and Time Series Econometrics. Oxford University Press, Oxford, pp. 218–249.

Stock, J.H., 2020. Data gaps and the policy response to the novel coronavirus. NBER Working Paper 26902. arXiv:arXiv:1011.1669v3.

Streeck, H., Schulte, B., Kümmerer, B.M., Richter, E., Höller, T., Fuhrmann, C., Bartok, E., Dolscheid, R., Berger, M., Wessendorf, L., Eschbach-bludau, M., Kellings, A., Schwaiger, A., Coenen, M., Schmithausen, R.M., Schmid, M., Hartmann, G., 2020. Infection fatality rate of SARS-CoV-2 infection in a German community with a superspreading event. https://doi.org/10.1101/2020.05.04.20090076 .

The New York Times, 2020. "Life has to go on": How Sweden has faced the virus without a lockdown. https://www.nytimes.com/2020/04/28/world/europe/sweden-coronavirus-herd-immunity.html.

Thompson, R.N., Stockwin, J.E., van Gaalen, R.D., Polonsky, J.A., Kamvar, Z.N., Demarsh, P.A., Dahlqwist, E., Li, S., Miguel, E., Jombart, T., Lessler, J., Cauchemez, S., Cori, A., 2019. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. Epidemics 29.

Tsang, T.K., Wu, P., Lin, Y., Lau, E.H.Y., Leung, G.M., Benjamin, J., 2020. Impact of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China. https://doi.org/10.1101/2020.03.23.20041319.

Verity, R., Okell, L.C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-dannenburg, G., Thompson, H., Walker, P.G.T., Fu, H., Dighe, A., Griffin, J.T., Baguelin, M., Bhatia, S., Boonyasiri, A., Cori, A., Cucunubá, Z., Fitzjohn, R., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Laydon, D., Nedjati-gilani, G., Riley, S., Elsland, S.V., Volz, E., Wang, H., Wang, Y., Xi, X., Donnelly, C.A., Ghani, A.C., Ferguson, N.M., 2020. Estimates of the severity of coronavirus disease 2019: A model-based analysis. Lancet Infectious Diseases 3099, 1–9.

Wang, H., Wang, Z., Dong, Y., Chang, R., Xu, C., Yu, X., Zhang, S., Tsamlag, L., Shang, M., Huang, J., Wang, Y., Xu, G., Shen, T., Zhang, X., Cai, Y., 2020. Phase-adjusted estimation of the number of coronavirus disease 2019 cases in Wuhan, China. Cell Discovery 6, 4–11.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Appendix A Supplementary Methods and Materials

A.1 SIS Model

We now show that the estimator in Eq. (3) also obtains when the dynamics of the disease follow the SIS model. The SIS model, again in discrete time, is given by

$$S_t = S_{t-1} - \beta_t I_{t-1} \frac{S_{t-1}}{N} + \gamma I_{t-1}$$
$$I_t = I_{t-1} + \beta_t I_{t-1} \frac{S_{t-1}}{N} - \gamma I_{t-1}$$

The only difference from the SIR model in Eq. (1) is that formerly infected individuals do not obtain immunity after recovery and instead again join the pool of susceptibles. As is well known, the basic reproduction number in the SIS model is the same as in the SIR model (e.g., Chowell and Brauer, 2009) and given by $\mathcal{R}_0^{(t)} = \beta_t/\gamma$. Since the law of motion for I_t in the SIS model is the same as in the SIR model, we can repeat the same steps as in the benchmark analysis to arrive at Eq. (3).

A.2 Generalized SIR Model

In this section, we show that the estimator in Eq. (3) also obtains in a generalized version of the SIR model with stochastic shocks. Specifically, we consider the following generalized SIR model:

$$S_t = S_{t-1} - \beta_t I_{t-1} \frac{S_{t-1}}{N} - v_{1,t}$$
$$I_t = I_{t-1} + \beta_t I_{t-1} \frac{S_{t-1}}{N} - \gamma I_{t-1} + v_{1,t} - v_{2,t}$$
$$R_t = R_{t-1} + \gamma I_{t-1} + v_{2,t}$$

Differently from the baseline model, we introduce random shocks $v_{1,t}$ and $v_{2,t}$. The shocks are i.i.d., and the time-varying support of $v_{1,t}$ is $[0, S_{t-1} - \beta_t I_{t-1}/N]$, while the support of $v_{2,t}$ is $[0, I_{t-1} + \beta_t I_{t-1} S_{t-1}/N]$. We also assume that $\mathbb{E}_{t-1}[v_{1,t} - v_{2,t}] = 0$, so that the conditional expectation $\mathbb{E}_{t-1}[I_t]$ coincides with the value for I_t given by the noiseless SIR model. With these modifications, the model can capture rich patterns of infectious disease dynamics. For example, "super spreader events" can be modeled either as $v_{1,t}$ shocks or as a spike in β_t. The model can also capture richer forms of population structures than the baseline SIR model. For example, if individuals who are more infectious (e.g., those with more connections in a network model) are more likely to become infected first, that can be captured by assuming that β_t becomes lower over

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

time.

Defining the time-varying basic reproduction number as $\mathcal{R}_0^{(t)} = \beta_t/\gamma$, and $\mathcal{R}_t \equiv \mathcal{R}_0^{(t)} S_{t-1}/N$, we obtain that

$$\text{gr}(I_t) = \gamma(\mathcal{R}_t - 1) + v_t,$$

where $v_t \equiv (v_{1,t} - v_{2,t})/I_{t-1}$. Taking expectations on both sides of the equation, we arrive at

$$\mathbb{E}[\mathcal{R}_t] = 1 + \frac{1}{\gamma}\,\mathbb{E}[\text{gr}(I_t)].$$

Hence, the generalized SIR model of the present section leads to the same estimator as the baseline SIR model in Eq. (1).

Finally, we note that if γ varies deterministically over time, the equation above remains essentially unchanged, the only difference being that γ is replaced by γ_t. If γ_t follows a non-degenerate stochastic process, then the estimator for $\mathbb{E}[\mathcal{R}_t]$ would need to correct for the covariance between γ_t and $\mathcal{R}_t$.

A.3 Foundation for the Local-Level Model

When estimating $\mathcal{R}_t$, we use a local-level specification for the growth rate of the number of infected individuals. In this section, we show that the local-level model arises naturally in the SIR model in the early stages of an epidemic when the transmission rate follows a random walk.

Specifically, consider the generalized SIR model in Appendix A.2. We now specialize the process for the transmission rate β_t to be a random walk:

$$\beta_t = \beta_{t-1} + \eta_t, \eta_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_\eta^2),$$

with a given initial value $\beta_0 > 0$. We calculate that

$$\text{gr}(I_t) = \left(\frac{S_{t-1}}{N}\right)\beta_t - \gamma + v_t \approx \beta_t - \gamma + v_t$$

in the early stages of the epidemic when $S_t \approx N$. Defining the effective reproduction number early on in the epidemic as $\mathcal{R}_t = \beta_t/\gamma$, we therefore have directly that the growth rate of I_t follows a local-level model with

$$\begin{aligned}\text{gr}(I_t) &= \gamma(\mathcal{R}_t - 1) + v_t \\ R_t &= R_{t-1} + \tilde{\eta}_t\end{aligned}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

where $\tilde{\eta}_t \equiv \eta_t/\gamma$. Provided that the distribution of v_t can be approximated with a normal distribution, we directly obtain the specification in Eq. (5). Alternatively, to obtain an exact normal local-level model, we could assume that $v_{1,t} = v_{2,t} = 0$ (no shocks in the original model, just as in Eq. (1)) but that instead of observing the true growth rate $\text{gr}(I_t)$, we only observe $\text{gr}(I_t) + \varepsilon_t$ where ε_t is i.i.d. normally distributed mean-zero measurement error.

A.4 Gibbs-Sampling Algorithm

In this section we discuss how the parameters of the state-space model in Eq. (5) can be estimated with a Gibbs-sampling algorithm *à la* Carter and Kohn (1994). Besides being a natural robustness check to our methodology, this algorithm uses a somewhat different approach to ensure the non-negativity of $\mathcal{R}_t$.

To use the Kalman filter, we need to estimate σ_ϵ^2 and σ_η^2. For the Gibbs sampler, we break the model down into conditional densities from which we can sample iteratively. The algorithm is the following:

1. Conditional on σ_ε^2 and σ_η^2, use the Kalman filter to infer the state vector $\mathcal{R}_t$;
2. Conditional on the sequence of $\mathcal{R}_t$ computed in the previous step, take samples of σ_ε^2 and σ_η^2 from their prior distributions;
3. Conditional on the new draws of σ_ε^2 and σ_η^2, estimate $\mathcal{R}_t$;
4. Verify that each element of $\mathcal{R}_t$ is positive. If yes, store the draws of σ_ε^2 and σ_η^2. If not, discard the draws and repeat step 2;
5. Compute the Kalman smoother;
6. Iterate forward for as many replications as needed.

Finally, we contrast the estimates of $\mathcal{R}_t$ obtained with the Gibbs sampler to our baseline estimates. We obtain a correlation of 0.85 between the two sets of estimates. Credible intervals are highly correlated as well. Overall, we conclude that our estimates are similar across different statistical estimation approaches.

A.5 SEIR Model: Monte Carlo Simulation

Our estimation method uses a structural mapping between $\mathcal{R}_t$ and $\text{gr}(I_t)$ derived from the basic SIR model. While we can generalize the baseline SIR model to include stochastic shocks (Appendix A.2), and the estimator remains valid when the disease follows an SIS model (Appendix A.1), the model is nevertheless restrictive. In particular, it ignores incubation periods as well as transmission during the incubation period. These features are likely especially important when modeling COVID-19.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

We now perform a simulation exercise to see how our estimator of $\mathcal{R}_t$ performs in a richer model that accounts for these additional features. Specifically, we consider an SEIR model in which the exposed are infectious:

$$\begin{aligned} S_t &= S_{t-1} - \beta I_{t-1}\frac{S_{t-1}}{N} - \beta\epsilon E_{t-1}\frac{S_{t-1}}{N} \\ E_t &= E_{t-1} + \beta I_{t-1}\frac{S_{t-1}}{N} + \beta\epsilon E_{t-1}\frac{S_{t-1}}{N} - \kappa E_{t-1} \\ I_t &= I_{t-1} + \kappa E_{t-1} - \gamma I_{t-1} \\ R_t &= R_{t-1} + \gamma I_{t-1} \end{aligned} \tag{6}$$

Here, E_t denotes the number of individuals that are exposed at day t, κ is the daily transition rate from exposed to infected, and $\epsilon \in [0, 1]$ measures the degree to which the exposed are less infectious than the infected. If $\epsilon = 0$, the exposed are not infectious at all, and we obtain the benchmark SEIR model. If $\epsilon = 1$, the exposed are as infectious as the infected, and the model is isomorphic to the standard SIR model.

We calibrate the parameters following Wang et al. (2020) who apply the benchmark SEIR model (with $\epsilon = 0$) to study the dynamics of COVID-19 in Wuhan. In particular, we use $\kappa = 1/5.2$ and $\gamma = 1/18$ as in Wang et al. (2020). Then, we set $\epsilon = 2/3$, following Ferguson et al. (2020) who assume that symptomatic individuals are 50% more infectious than the asymptomatic (that is, $\epsilon^{-1} = 1.5$). Finally, we choose β by targeting a basic reproduction number of $\mathcal{R}_0 = 2.6$, again as in Wang et al. (2020). In the model above, $\mathcal{R}_0$ is given by $\mathcal{R}_0 = \beta/\gamma + \beta\epsilon/\kappa$, implying $\beta = \mathcal{R}_0\gamma\kappa/(\gamma\epsilon + \kappa)$. The formula yields $\beta \approx 0.12$. Finally, we set $S_0 = 11 \times 10^6$ (approximating the population size of Wuhan), $E_0 = R_0 = 0$, and $I_0 = 1$.

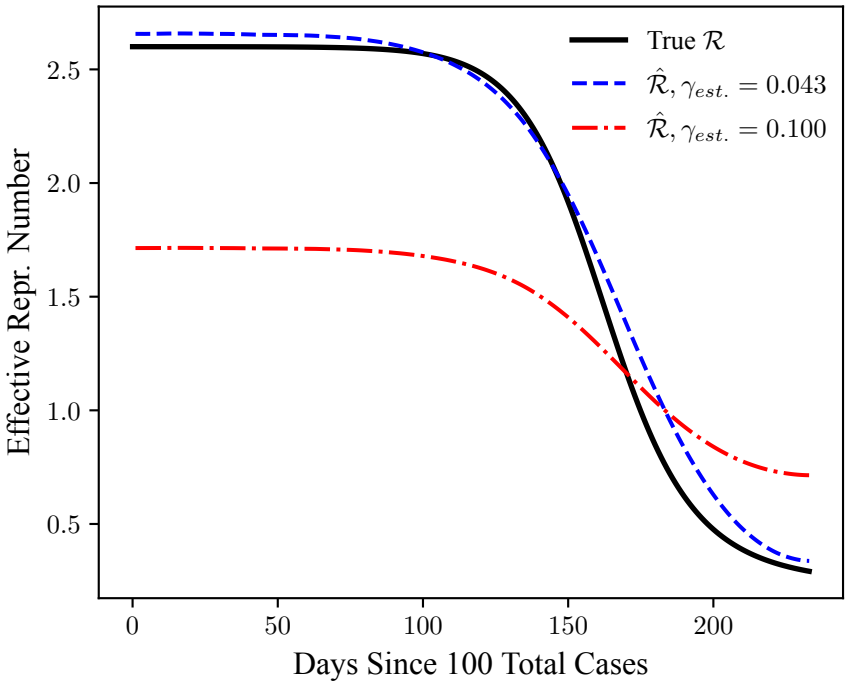
The Monte Carlo design is as follows. First, we simulate the deterministic system in Eq. (6) using the parameters above. Then, we calculate the growth rate in the true number of infected individuals, i.e., $\text{gr}(I_t) = I_t/I_{t-1} - 1$. However, instead of knowing the true growth rate, the statistician is assumed to observe a noisy version of it given by $\widetilde{\text{gr}}(I_t) = \text{gr}(I_t) + \varepsilon_t$. Here, ε is an i.i.d. normal disturbance with mean zero and standard deviation of 0.10. The standard deviation of the disturbances is roughly equal to the range of the true growth rates. Hence, the amount of noise used in the simulation is fairly large. For each realization of the disturbances, we estimate $\mathcal{R}_t$ using our method. As in our empirical application, only data after 100 total cases have been reached is used.

We investigate two values for $\gamma_{\text{est.}}$ that are used when estimating $\mathcal{R}_t$ via Eq. (3). First, we consider a situation in which the statistician uses the correct time that individuals are infected, given by $\gamma_{\text{est.}} = (\gamma^{-1} + \kappa^{-1})^{-1}$ where γ and κ are the true parameter values

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure A.1

Monte Carlo Simulation: Effects of Misspecification



Notes: Estimates of the effective reproduction rate ($\mathcal{R}_t$) when the true dynamics of the disease follow an SEIR model. We investigate two values for $\gamma_{\text{est.}}$, the transition rate from infected to recovered, that are used when estimating $\mathcal{R}_t$. First, we use the correct value of $\gamma_{\text{est.}} = (\gamma^{-1} + \kappa^{-1})^{-1} \approx 0.043$. Second, we use a misspecified values of $\gamma_{\text{est.}} = 1/10$. Average values from 10,000 Monte Carlo replications are shown. See text for more details.

of the SEIR model. Second, we investigate a case in which the statistician incorrectly thinks that individuals are infectious only for ten days ($\gamma_{\text{est.}} = 1/10$). We repeat the process for 10,000 Monte Carlo replications.

The results of the Monte Carlo simulation are shown in Figure A.1. When the statistician uses the correct number of days that an individual is infectious (that is, taking into account the incubation time), the estimates of $\mathcal{R}_t$ from our method are very close to their true theoretical values. That is in spite of the fact that our estimator for $\mathcal{R}_t$ is derived assuming that the dynamics of the disease are described by an SIR model. However, we also show that if the statistician misspecifies the number of days than an individual is infectious (assuming 10 days instead of the true number of 23.2 days), the estimates of $\mathcal{R}_t$ are substantially biased, especially in the early stages of the epidemic. As is to be expected from Eq. (3), underestimating the number of days that an individual is infectious leads to a downwards bias in the estimates of $\mathcal{R}_t$ early on in the epidemic (when $\mathcal{R}_t > 1$), and upwards bias when the true $\mathcal{R}_t$ falls below one. Overall, the results imply

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

that the new method performs well when estimating $\mathcal{R}_t$ even when the true dynamics of the disease do not follow the SIR model, provided that the duration of infectiousness used in the estimation is sufficiently accurate.

A.6 Effects of Potential Data Issues

We now discuss the effects of various data issues on the performance of our estimator.

Reporting delays. In practice, data may be subject to significant reporting delays. For example, suppose that due to testing constraints there is a lag of ℓ days between the date that an individual becomes infected and the date on which the case is registered. In this case, the estimates of $\mathcal{R}_t$ would also be subject to delay of ℓ days. If there are significant reporting delays, one may first obtain, say, one-week-ahead forecasts of new cases, and then use these forecasts to construct a time series for I_t.

Imperfect detection. A natural worry with any estimator of $\mathcal{R}_t$ is that it may be substantially biased if not all of infected individuals are detected. Given the simplicity of our estimator, we can analytically assess the effects of imperfect detection.

Suppose that the true numbers of susceptible, infected, and recovered individuals are given by S_t^*, I_t^*, and $\mathcal{R}_t$, respectively. Their evolution is the same as in Eq. (1). However, we only observe $I_t = \alpha_t I_t^*$, where $\alpha_t \equiv I_t/I_t^*$ is the *detection rate*. In practice, α is typically less than one, although the mathematical calculation below does not require this.

With this notation, we have that

$$\text{gr}(I_t) = \text{gr}(\alpha_t)[1 + \text{gr}(I_t^*)] + \text{gr}(I_t^*) \approx \text{gr}(\alpha_t) + \text{gr}(I_t^*),$$

since $\text{gr}(\alpha_t) \times \text{gr}(I_t^*) \approx 0$ at a daily frequency; the approximation is exact in continuous time. Using the approximation above and Eq. (2), we therefore obtain that the bias of the estimator under imperfect detection is given by

$$\hat{\mathcal{R}}_t - \mathcal{R}_t \approx \frac{1}{\gamma}\text{gr}(\alpha_t).$$

We now discuss several cases of practical importance:

- Constant detection rate ($\alpha_t = \alpha$). If the detection rate is constant over time, then our estimator is unbiased, and $\hat{\mathcal{R}}_t = \mathcal{R}_t$. Hence, for example, even if we only detect 10% of the infectives (but the fraction detected remains constant over time), the estimator remains unbiased. Note that if the number of tests increases over time, that is *not* inconsistent with $\alpha_t = \alpha$ given that the number of infected

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

individuals is likely to be growing at the same time.

- Constant growth in the detection rate ($\text{gr}(\alpha_t) = g_\alpha$). If the growth rate of α_t is constant over time, then our estimate of $\mathcal{R}_t$ is biased upwards if $g_\alpha > 0$ and downwards if $g_\alpha < 0$. Note, however, that we are often mostly interested in the *trend* of $\mathcal{R}_t$ over time and whether the trend is affected by various policy interventions. The trend in $\mathcal{R}_t$ is estimated accurately even if $g_\alpha \neq 0$. Intuitively, constant growth in the detection rate leads to a level bias, but the slope is still estimated correctly.
- Detection rate converges over time ($\alpha_t \to \alpha$). The final case of interest occurs when the detection rate converges to a constant over time. For example, if everyone is detected towards the end of the epidemic, we would have $\alpha_t \to 1$. Since our method uses Kalman-filtering techniques to estimate the growth rate of I_t, transient fluctuations in α_t would have a limited effect on the estimates of $\mathcal{R}_t$ later on in the sample. Given that we are often precisely interested in the behavior of $\mathcal{R}_t$ in the later stages of the epidemic (when the detection rate is likely fairly constant), our method would still yield reliable estimates.

To provide a quantitative illustration, we perform a small-scale Monte Carlo study. We choose the parameters of the simulation to match our empirical estimates. First, we use our estimates of $\mathcal{R}$ for the world as a whole (see Figure 1) for the first 50 days of the sample as the true values of $\mathcal{R}$. From these values of $\mathcal{R}$, we calculate the true (but unobserved) values of the growth rate of the number of infected individuals as $\mu_t \equiv \text{gr}(I_t^*) = \gamma(\mathcal{R}_t - 1)$, using $\gamma = 1/7$. Finally, the observed growth rate—as seen by the statistician—is generated as $\text{gr}(I_t) = \text{gr}(\alpha_t)(1 + \mu_t) + \mu_t + \varepsilon_t$. We use the empirical estimate of σ_ε^2 to simulate ε_t shocks (i.i.d. draws from a normal distribution with mean zero). We then apply our estimator to the generated data on $\text{gr}(I_t)$, using the same value of $\gamma = 1/7$ to back out $\mathcal{R}$.

In the simulation we consider three underdetection scenarios:

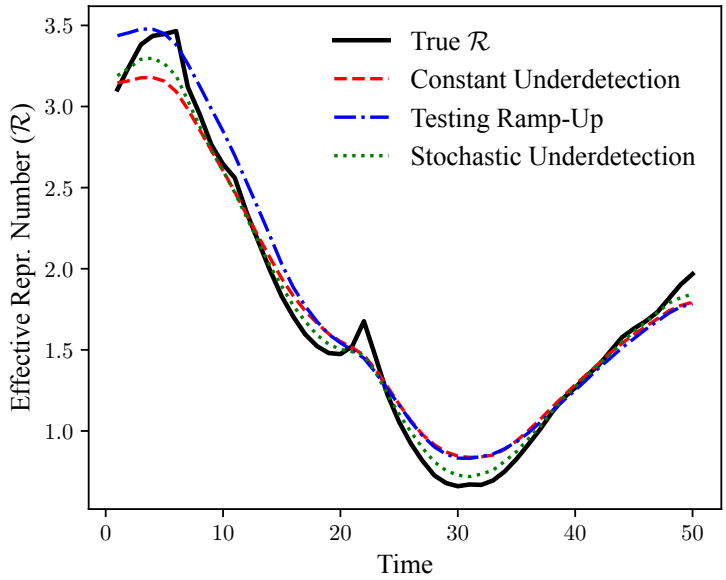
- *Constant underdetection*. Given our analytical results, it is immaterial what percentage of cases is detected, as long as that percentage is constant over time.
- *Ramp-up in testing*. Next, we consider a situation in which—due to an increase in the number of tests performed—the fraction of detected infected individuals goes up from $\alpha_0 = 0.10$ to $\alpha_{14} = 0.15$ (an increase of 50%) over a period of two weeks. After the two weeks, the detection probability remains constant at 0.15. Again, the precise values of α_0 and α_{14} are irrelevant, and what matters is the growth rate in the detection rate.
- *Stochastic underdetection*. Finally, we suppose that the detection rate satisfies

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

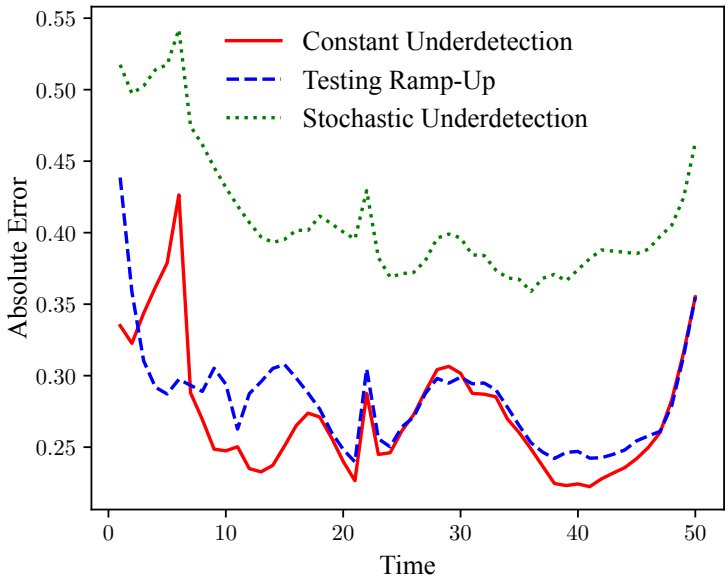
Figure A.2

Monte Carlo Simulation: Effects of Underdetection

(a) Average Estimates



(b) Estimation Accuracy



Notes: Estimates of the effective reproduction rate ($\mathcal{R}_t$) when the true number of infected individuals is measured with error. The true values for $\mathcal{R}_t$ are given by our empirical estimates of $\mathcal{R}_t$ for the world as a whole in the first 50 days of the sample, and the standard deviation of the irregular component is chosen to match our empirical estimates. *Constant Underdetection*: A constant fraction of infected individuals is detected. *Testing Ramp-Up*: The fraction of detected individuals increases by 50% in the first two weeks of the sample. *Stochastic Underdetection*: The fraction of detected individuals is stochastic, and its growth rate follows an AR(1) process. Average values from 1,000 Monte Carlo replications are shown. *Top panel*: Average estimates. *Bottom panel*: average absolute error of the estimates.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

$\text{gr}(\alpha_t) = \phi\,\text{gr}(\alpha_{t-1}) + \nu_t$ where ν_t is an i.i.d. normally distributed shock. Intuitively, the detection rate is assumed to be unconditionally constant, but its growth rate is stochastic and follows an AR(1) process. We set the persistence parameter to $\phi = 0.75$ in order to allow for fairly long-lasting deviations from the average detection rate. To ensure that the variance of $\text{gr}(I_t)$ remains constant across simulations (to ensure an apples-to-apples comparison), we suppose that 50% of the noise in the observed growth rate comes from variation in α_t, with the other 50% coming from the ε_t shocks. (Here, by "noise" we refer to the variation in $\text{gr}(I_t)$ that is not solely due to the variation in μ_t, namely, $\text{gr}(\alpha_t)(1+\mu_t)+\varepsilon_t$.) Denoting our estimates of the variance of the growth rate by $\hat{\sigma}^2_\mu$ and the variance of the irregular component by $\hat{\sigma}^2_\varepsilon$, we therefore set $\mathbb{V}\text{ar}[\nu_t] = (1-\phi^2)\hat{\sigma}^2_\varepsilon/[2(1+\hat{\sigma}^2_\mu)]$ and $\mathbb{V}\text{ar}[\varepsilon_t] = \hat{\sigma}^2_\varepsilon/2$. We draw the initial value for $\text{gr}(\alpha_0)$ from its unconditional distribution.

The results of the Monte Carlo simulation are summarized in Figure A.2. As seen in the top panel, the estimated average effective reproduction numbers are fairly close to their theoretical values in all three scenarios. In the *Testing Ramp-Up* scenario, the estimates are biased upwards at the beginning of the sample, but the amount of bias is quantitatively relatively small. In addition, the estimates converge to those in the other two scenarios quite quickly. In all three scenarios, the estimates are able to pick up the reversal in the trend of $\mathcal{R}$ that happens at around time 30. The bottom panel of the figure plots the average absolute error of the estimates. As expected, the estimates are least accurate in the *Stochastic Undertesting* case, and mostly within 0.25–0.30 of the true $\mathcal{R}$ in the *Constant Underdetection* and *Testing Ramp-Up* scenarios.

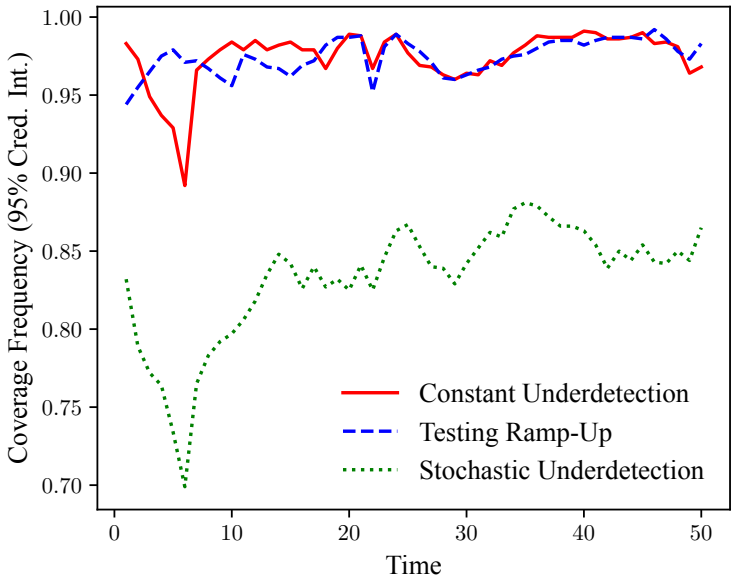
Finally, in Figure A.3 we provide Monte Carlo estimates of the coverage frequency of credible intervals obtained by our estimation procedure. As seen in the graph, the credible intervals in the *Constant Undertesting* and *Testing Ramp-Up* scenarios have good coverage properties, with confidence bounds that are slighly overly conservative. The credible bounds are somewhat too narrow in the *Stochastic Underdetection* scenario, resulting in lower than nominal coverage frequency. However, the size distortion appears not too severe, especially considering the small sample size.

Imported cases. Our estimates may be biased if the fraction of cases that is imported changes over time (the previous results on imperfect detection apply to misclassification because of imported cases, too). If the source of infections is known, it is possible to correct for the issue by simply not including imported cases when constructing the time series for I_t.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure A.3

Monte Carlo Simulation: Coverage Frequency



Notes: Coverage frequency of 95% credible intervals: Monte Carlo results. We investigate three scenarios for measurement error. *Constant Underdetection*: A constant fraction of infected individuals is detected. *Testing Ramp-Up*: The fraction of detected individuals increases by 50% in the first two weeks of the sample. *Stochastic Underdetection*: The fraction of detected individuals is stochastic, and its growth rate follows an AR(1) process. Average values from 1,000 Monte Carlo replications are shown. See text and Figure A.2 for more details.

Table A.1

Priors

Parameter	Prior
Precision of irregular component ($1/\sigma_\varepsilon^2$)	Gamma$(0.15511, 0.00094)$
Signal-to-noise ratio ($q \equiv \sigma_\eta^2/\sigma_\varepsilon^2$)	Gamma$(0.01196, 0.05018)$
Initial value (μ_0)	$\mathcal{N}(0.35, 0.5^2)$

Notes: Priors used in the Bayesian estimation of $\mathcal{R}_t$. See text for description on how the priors for the precision of the irregular component ($1/\sigma_\varepsilon^2$) and the signal-to-noise ratio ($q \equiv \sigma_\eta^2/\sigma_\varepsilon^2$) are calibrated based on cross-country frequentist estimates.

A.7 Estimation Details

To estimate $\mathcal{R}_t$ of COVID-19, we use Bayesian filtering methods. We employ the following strategy to calibrate the prior distributions. First, we estimate a local-level model for gr(I_t) using a frequentist Kalman filter with diffuse initial conditions. In particular,

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

we estimate the following model:

$$\begin{aligned} \text{gr}(I_t) &= \mu_t + \varepsilon_t, \qquad \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_\varepsilon^2) \\ \mu_t &= \mu_{t-1} + \eta_t, \quad \eta_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_\eta^2) \end{aligned} \tag{7}$$

The model is the same as in Eq. (5), except for a slight simplification in notation.

The procedure yields maximum likelihood estimates of σ_ε^2 (variance of the irregular component) and the signal-to-noise ratio $q \equiv \sigma_\eta^2/\sigma_\varepsilon^2$ for each country in the sample (with σ_η^2 denoting the variance of the level component). We then use the distribution of $\hat{\sigma}_\varepsilon^2$ and $\hat{q}$ across countries to calibrate the priors for the precision of the irregular component ($1/\sigma_\varepsilon^2$) and the signal-to-noise ratio (q). To ensure that the priors are not too "dogmatic," we inflate the variance of the estimates by a factor of 3 when calibrating the prior distributions. We use a gamma prior for both the signal-to-noise ratio and the precision of the irregular component, and we calibrate the parameters of the gamma distribution by matching the expected value and variance of the gamma-distributed random variables to their sample counterparts. Finally, we use a fairly uninformative normal prior for the initial value of the smoothed growth rate. The resulting priors are given in Table A.1.

Intuitively, these priors shrink the estimates of the precision and signal-to-noise ratio for each country towards their grand mean (average across countries). Such Bayesian shrinkage ensures that the parameter estimates are well behaved even though the sample size for many countries is fairly small, and the data are often noisy. We use the Stan programming language (Gelman, Lee and Guo, 2015) to specify and estimate the Bayesian model. In particular, we use the `pystan` interface to call Stan from Python.

A.8 Empirical Validation

In this section, we perform two empirical validation exercises to check the performance of our estimates in practice.

Since our estimates are based on data on new cases, they may be misleading if new cases are subject to significant measurement problems. To help assuage this concern, we now perform the following exercise. We ask whether *current* values of $\mathcal{R}_t$ help predict *future* growth in deaths. Since deaths are likely to be measured more accurately, this exercise provides a test of whether our estimates contain meaningful information and are not contaminated by data problems.

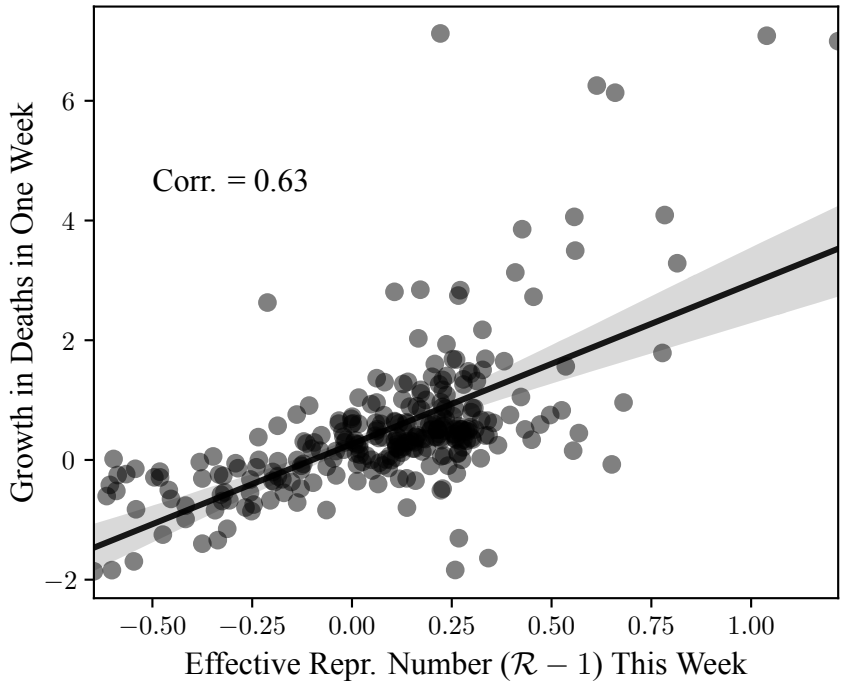
Formally, we consider the following regression:

$$\text{gr}(d_{i,t+1}) = \alpha_i + \beta(\hat{\mathcal{R}}_{i,t} - 1) + u_{i,t},$$

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure A.4

$\mathcal{R}_t$ and Future Deaths



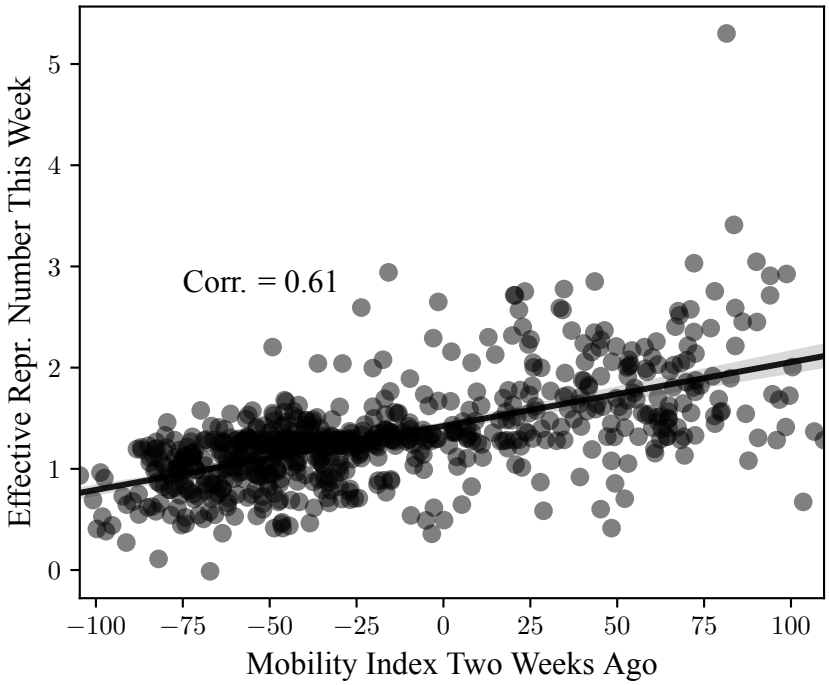
Notes: Relationship between current estimates of the effective reproduction number ($\mathcal{R}_t$) and the growth rate of the number of new deaths in one week. The data is aggregated to a weekly frequency. Both variables are residualized to subtract country fixed effects by performing the within transformation. Only data after the cumulative number of deaths reaches 50 is included in the scatter plot. We include all countries in the John Hopkins database for which we have at least 20 observations after the outbreak. We remove data for the week of 2020-04-13–2020-04-19 in China that contain a large number of deaths that were previously unrecognized.

where i denotes a particular country, and t indexes calendar weeks. Although our original data is daily, we aggregate to a weekly frequency; otherwise, measures of the growth rate of new deaths are too noisy. In addition, we only include weeks after the cumulative number of COVID-19 deaths has reached 50. Given that we have panel data, we can include country fixed effects α_i to account for time-invariant unobserved heterogeneity (such as differences in average age—a key correlate of COVID-19 mortality (Verity et al., 2020)—or family structures). The relationship given above is predicted by the baseline SIR model. Specifically, consider Eq. (2). Letting $\text{CFR} = d_t/I_{t-\ell}$ denote the case fatality rate (assumed to be constant over time), with ℓ standing for the average time between becoming infected and death, we have that $\text{gr}(d_t) = \text{gr}(I_{t-\ell})$, yielding the regression equation above.

The relationship is shown in Figure A.4. In the scatter plot, both variables are resid-

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure A.5
$\mathcal{R}_t$ and Past Mobility



Notes: Relationship between current estimates of the effective reproduction number ($\mathcal{R}_t$) and value of the movement index two weeks ago (first principal component of the six movement categories in Google (2020)). The data is aggregated to a weekly frequency. Both variables are residualized to subtract country fixed effects by performing the within transformation. We include all countries in the John Hopkins database for which we have at least 20 observations after the outbreak.

ualized to remove country fixed effects. We observe a strong positive relationship between the value of $\mathcal{R}_t$ this week and the growth in deaths one week later (corr. = 0.63). In Supplementary Figure (Figure A.9), we demonstrate that there is also positive correlation (corr. = 0.44) between $\mathcal{R}_t$ and deaths two weeks later. We note that while the average medical duration from the onset of symptoms to death for COVID-19 is longer than two weeks (around 18 days, see Verity et al., 2020), the duration from *reported* cases to deaths is likely to substantially shorter because of reporting delays. For example, Hortaçsu, Liu and Schwieg (2020) assume that new cases of COVID-19 are reported with a lag of 8 days in their baseline calculations (5 days for symptoms to appear, consistent with the evidence from Lauer et al. (2020) and Park et al. (2020), as people are unlikely to be tested without exhibiting symptoms, and an additional 3 days to capture delays in obtaining test results, based on andecdotal reports from the US). Since deaths are likely reported in a timely manner, if new cases are reported with a lag of 8 days, we would expect an average duration of around 10 days (≈1.43 weeks) between reported

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

cases and reported deaths.

As a second validation check, we ask whether our estimates of $\mathcal{R}_t$ are correlated with past movement data, as it should be if the estimates are meaningful. For information on movement, we use aggregated smartphone location data collected by Google and published in their "COVID-19 Community Mobility Reports" (Google, 2020). Google provides data on percentage changes in movement for six types of places: (i) groceries and pharmacies; (ii) parks; (iii) transit stations; (iv) retail and recreation; (v) residential; and (vi) workplaces. Since the six categories are strongly correlated, we take the first principal component of the six categories (the first principal component explains a little less than 85% of the total variance in the data). We refer to the first principal component as the "Mobility Index." As shown in Figure A.5, current estimates of $\mathcal{R}_t$ are strongly correlated with the value of the mobility index two weeks ago (corr. = 0.61).

For both validation exercises performed in the present section, we include all countries for which we have at least 20 observations after the onset of the epidemic (100 cumulative cases of COVID-19 reached). If we narrow the sample down to countries with more and higher-quality data—such as the sample of European countries analyzed in Section 3.3—the correlations generally become substantially stronger. Hence, we consider the tests of the present section to be conservative.

A.9 Power Analysis

In this section, we study the statistical power of the empirical analysis in Section 3.3 using a Monte Carlo simulation.

We now describe the design of the power study. Intuitively, we simulate data using a stochastic process that is calibrated to match the properties of the observed data. We then simulate a sharp drop in the effective reproduction number—say, because of a lockdown. We apply our estimator to the simulated data and ask how often this abrupt change is detected by the estimation procedure.

To simplify the notation, we use the parametrization in Eq. (7). Optimal nowcasts from the local-level model in the steady state can be written as (Muth, 1960; Shephard, 2015, Section 3.4):

$$m_t = \omega y_t + (1-\omega) m_{t-1}, \omega = \frac{q + \sqrt{q^2 + 4q}}{2 + q + \sqrt{q^2 + 4q}}, \tag{8}$$

where we denote $y_t \equiv \text{gr}(I_t)$, $\mu_t \equiv \gamma(\mathcal{R}_t - 1)$, $m_t \equiv \mathbb{E}_t[\mu_t]$, $q \equiv \sigma_\eta^2/\sigma_\varepsilon^2$ is the signal-to-noise ratio, and ω is the steady-state Kalman gain. Hence, nowcast errors, $m_t - \mu_t$,

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table A.2

Power Analysis: Parameter Values

γ	q	σ_ε^2	ω
0.143	0.048	0.020	0.197

Notes: Parameter values used in the power analysis. The parameters values for q (signal-to-noise ratio) and σ_ε^2 (variance of the irregular component) are given by the median estimates from the 14 countries considered in the empirical analysis of Section 3.3. The mean duration of infectiousness is assumed to be $\gamma^{-1} = 7$, and the Kalman gain ω is calculated from Eq. (8).

follow an AR(1) process with

$$m_t - \mu_t = (1-\omega)(m_{t-1} - \mu_{t-1}) + \{\omega\varepsilon_t - (1-\omega)\eta_t\}.$$

Given that the shocks ε_t and η_t are uncorrelated, the variance of nowcast errors is

$$\mathbb{V}\text{ar}(m_t - \mu_t) = \frac{\omega^2\sigma_\varepsilon^2 + (1-\omega)^2 q\sigma_\varepsilon^2}{1-(1-\omega)^2}. \tag{9}$$

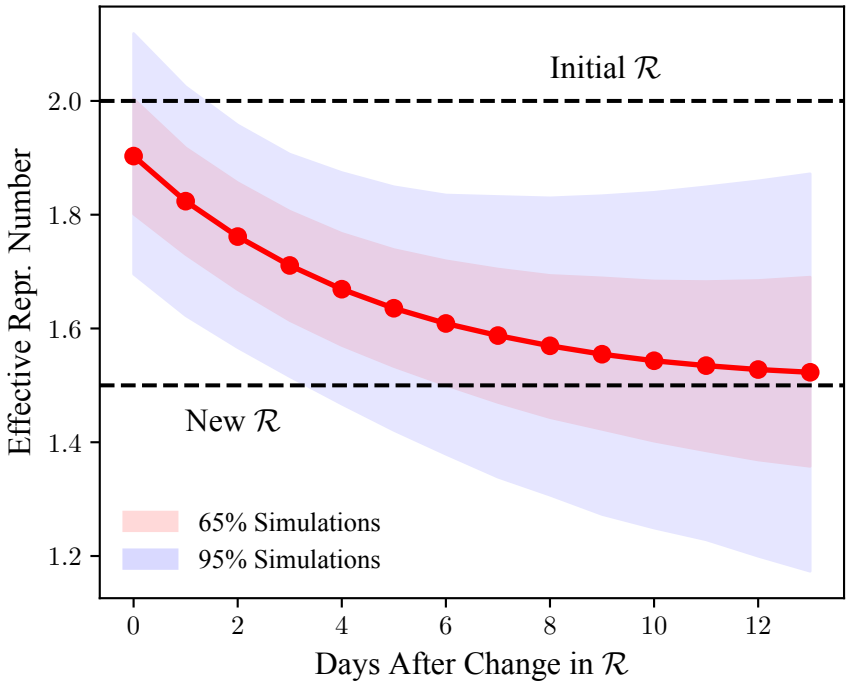
The design of the power analysis is as follows:

1. We set $\gamma = 1/7$, and calibrate the remaining parameters of the data-generating process (q and σ_ε^2) using the median values of the empirical estimates from Section 3.3. The resulting parameter values are given in Table A.2.
2. We simulate $\mu_t = \gamma(\mathcal{R}_t - 1)$. We initially set $\mu_0 = 1/7$, implying an effective reproduction number of 2. At time 1, we simulate an abrupt decline in $\mathcal{R}_t$ by setting $\mu_1 = 1/14$, yielding a new effective reproduction number of 1.5, or a decline of 25%. For $2 \leq t \leq 14$, we simulate μ_t as a random walk, as in Eq. (7).
3. We simulate the observed growth rate of the number of infected individuals, $y_t = \text{gr}(I_t)$, as $y_t = \mu_t + \varepsilon_t$ where ε_t is an i.i.d. normal random variable with mean zero and variance σ_ε^2.
4. We simulate the nowcasts m_t. We draw the initial nowcast m_0 from a normal distribution with mean $1/7$ and variance given in Eq. (9), and simulate further values of m_t by the recursion in Eq. (8). The estimated value of $\mathcal{R}_t$ is then given by the estimator in Eq. (3).
5. We repeat steps 2–4 for 14 times, to simulate data for 14 "countries", as in the empirical application, and obtain estimates of the effective reproduction number by averaging across the 14 "countries".
6. We repeat steps 2–5 for 10,000 Monte Carlo replications.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure A.6

Power Analysis: Monte Carlo Results



Notes: Power study of the statistical analysis of Section 3.3 (effects of non-pharmaceutical interventions on $\mathcal{R}_t$, the effective reproduction rate). We simulate an abrupt change in $\mathcal{R}_t$ from 2.0 to 1.5 using a data-generating process that is calibrated to match our empirical estimates in Section 3.3. We then apply the estimator of Eq. (3) to the simulated data and ask how often the change is detected by the estimation procedure. The solid line gives the average estimate of $\mathcal{R}_t$, while the shaded lines denote 65% and and 95% of simulations (in particular, the shaded area for 65% of simulations is given by the 17.5 and 82.5 percentiles of the estimated $\mathcal{R}_t$ across simulations, and the shaded area for 95% of simulations is given by the 2.5 and 97.5 percentiles of the estimated $\mathcal{R}_t$'s). 10,000 Monte Carlo replications used.

The results of the power analysis are shown in Figure A.6. We observe that in 95% of the simulations, the change in $\mathcal{R}_t$ is detected as soon as two days after the drop in $\mathcal{R}_t$. Hence, the analysis in Section 3.3 appears sufficiently powerful to detect moderate changes in $\mathcal{R}_t$. The key reason why the analysis has high statistical power, even though the signal-to-noise ratio is quite low (see Table A.2) is that data from multiple countries are used to obtain cross-country averages. This feature of the estimation procedure reduces estimation error substantially. While the signal-to-noise ratio is fairly low, we also note that the weight placed on data is that are more than one-week old is only $(1-\omega)^7 \approx 21.5\%$. Hence, one week after the change in $\mathcal{R}_t$, the estimates of $\mathcal{R}_t$ are based primarily on data received after the change in $\mathcal{R}_t$.

The power analysis in the current section is arguably somewhat conservative. Specifically, we assume that after the abrupt decline, $\mathcal{R}_t$ follows a random walk rather than

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

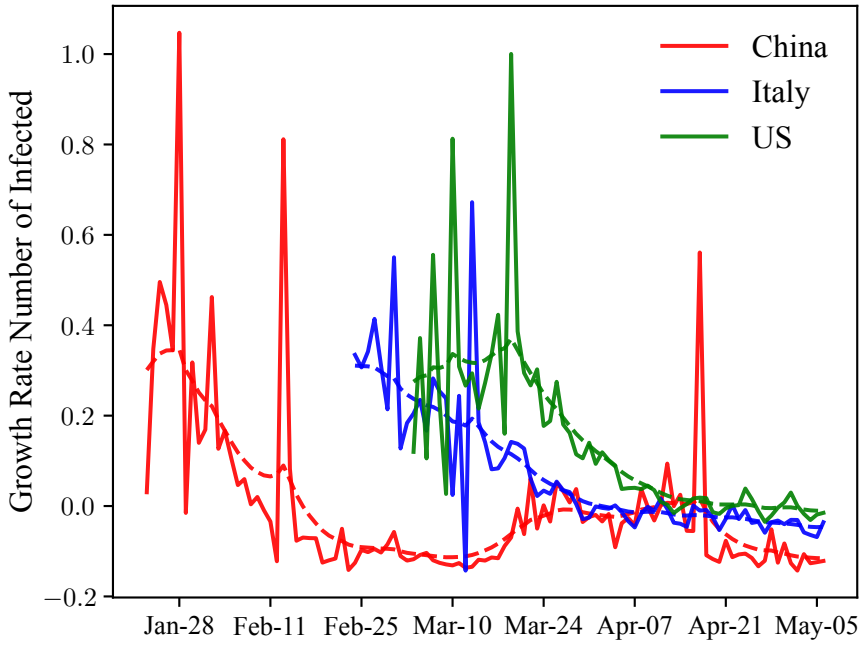
staying fixed at the new level. As a result, as time goes on, the estimates of $\mathcal{R}_t$ become more "spread out" across simulations, as is visible towards the end of Figure A.6.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Appendix B Supplementary Figures and Tables

Figure A.7

Growth Rate of the Number of Infected Individuals

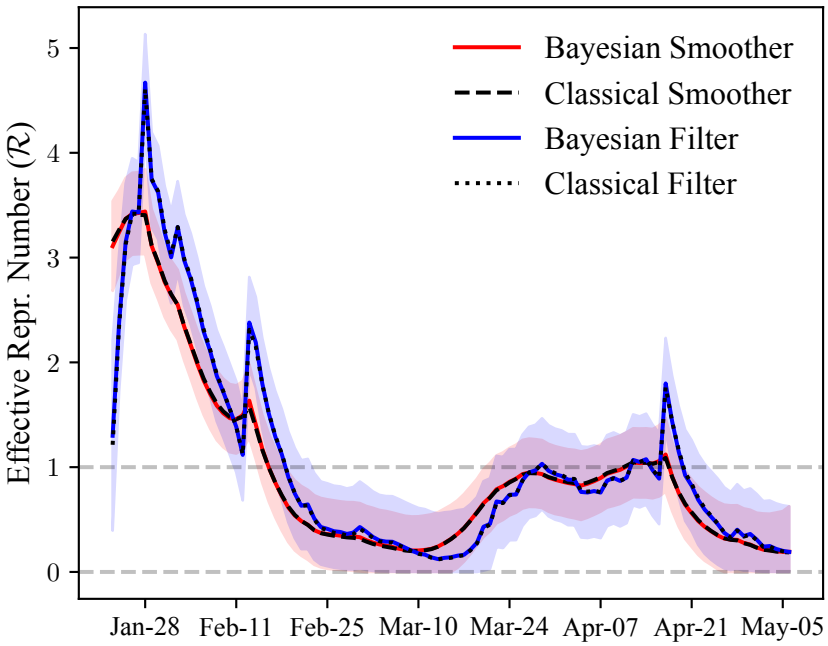


Notes: Raw data for the growth rate of the number of infected individuals (solid lines) and our estimate of its time-varying average (dashed lines) for China, Italy, and the US.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure A.8

Filtered and Smoothed Estimates: China

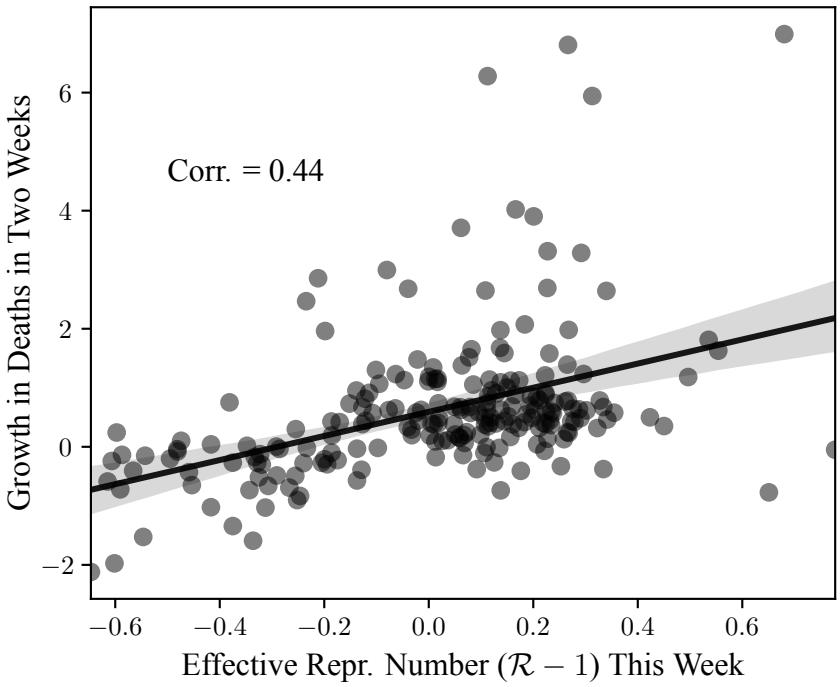


Notes: Estimated effective reproduction number ($\mathcal{R}$) for China: filtered and smoothed estimates, using both Bayesian and classical estimation procedures. The Bayesian estimates are given by our baseline estimation procedure, as explained in the text. The classical estimates are obtained by maximum likelihood estimation (with diffuse initial conditions). The smoothed estimates use information from the full sample, while the filtered estimates at time t only use information up to time t. 65% credible bounds shown by the shaded areas.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure A.9

$\mathcal{R}_t$ and Deaths in Two Weeks

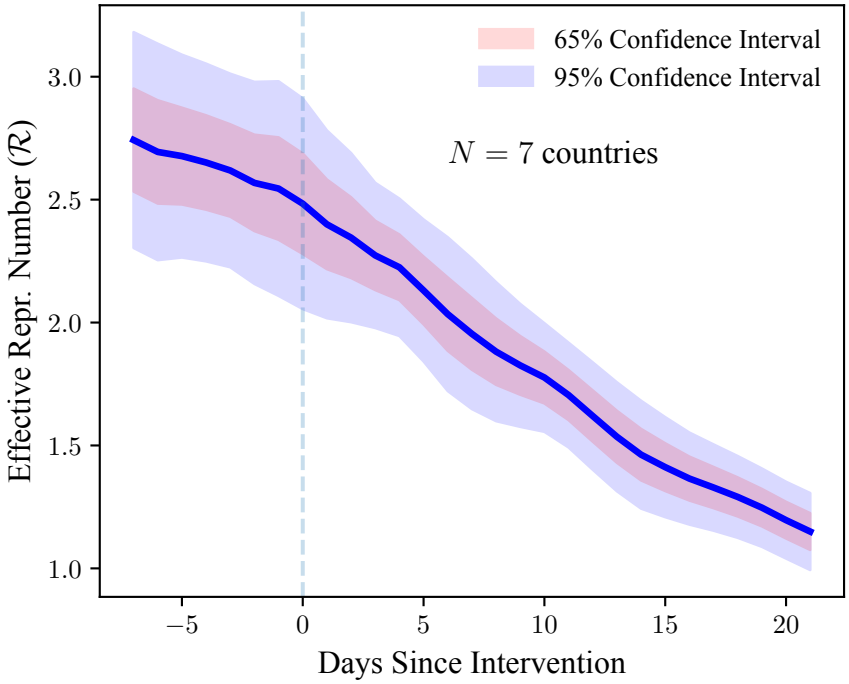


Notes: Relationship between current estimates of the effective reproduction number ($\mathcal{R}_t$) and the growth rate of the number of new deaths in two weeks. The data is aggregated to a weekly frequency. Both variables are residualized to subtract country fixed effects by performing the within transformation. See Figure A.4 for more details.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure A.10

$\mathcal{R}_t$ and Policy Interventions: Bans of Public Events

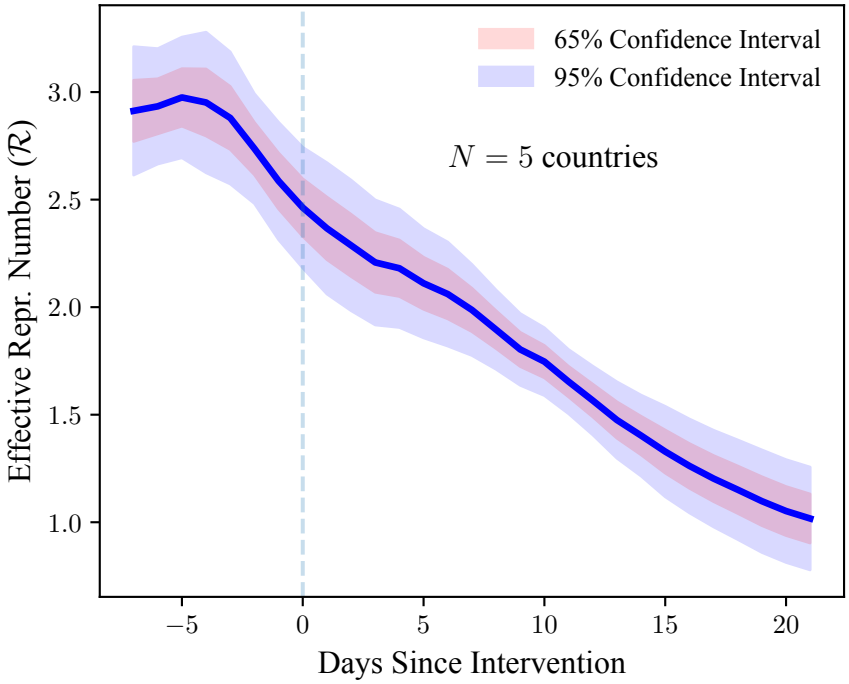


Notes: The graph plots the estimated effective reproduction number ($\mathcal{R}_t$) one week before and three weeks after public events are banned in a country. The original sample consists of 14 European countries studied by Flaxman et al. (2020). For the event-study graph, we restrict the sample to countries for which data on $\mathcal{R}_t$ is available for the whole event window. Heteroskedasticity-robust confidence bounds are shown by the shaded areas.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure A.11

$\mathcal{R}_t$ and Policy Interventions: Case-Based Measures

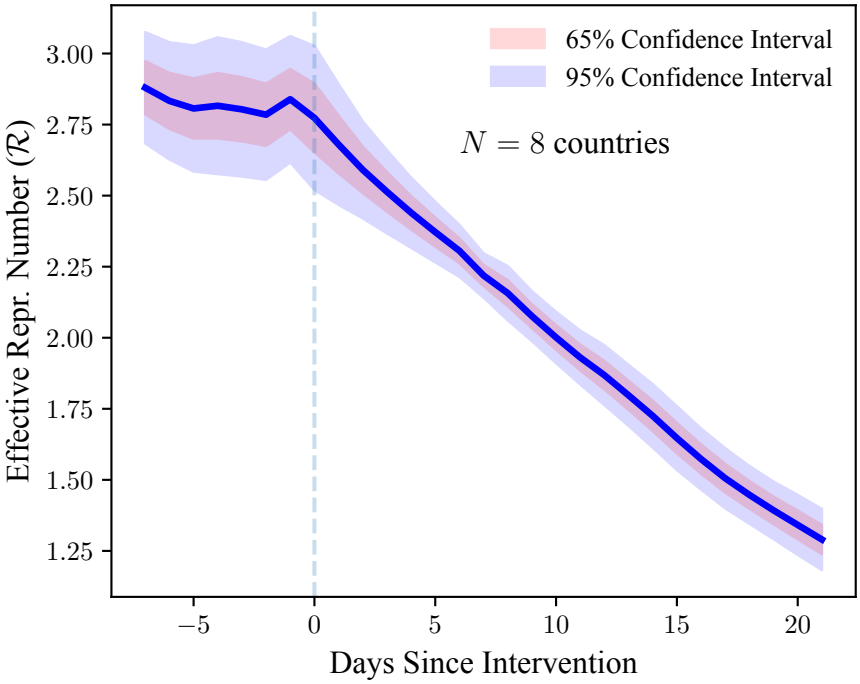


Notes: The graph plots the estimated effective reproduction number ($\mathcal{R}_t$) one week before and three weeks after case-based measures are introduced in a country. The original sample consists of 14 European countries studied by Flaxman et al. (2020). For the event-study graph, we restrict the sample to countries for which data on $\mathcal{R}_t$ is available for the whole event window. Heteroskedasticity-robust confidence bounds are shown by the shaded areas.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure A.12

$\mathcal{R}_t$ and Policy Interventions: School Closures

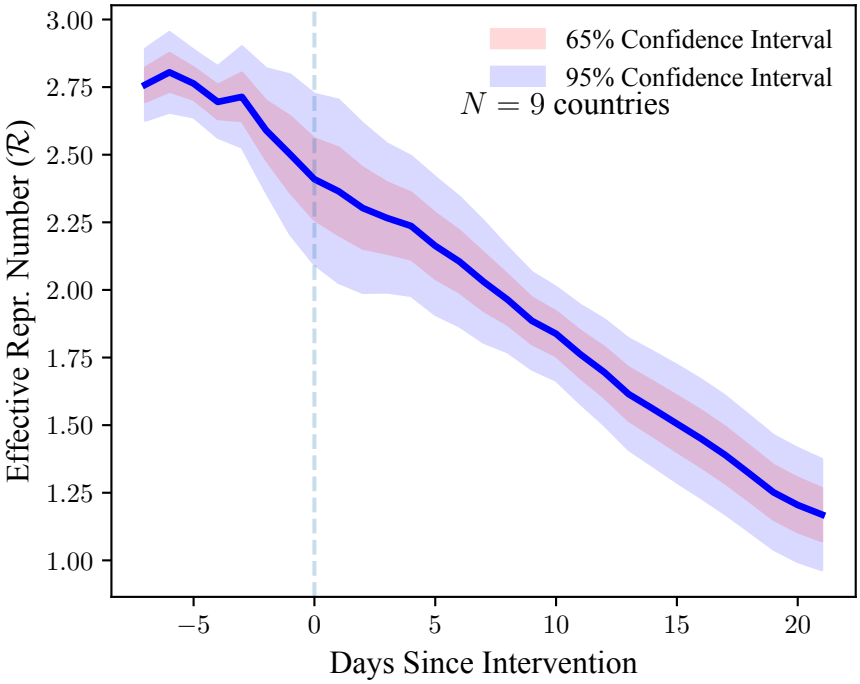


Notes: The graph plots the estimated effective reproduction number ($\mathcal{R}_t$) one week before and three weeks after school closures are ordered in a country. The original sample consists of 14 European countries studied by Flaxman et al. (2020). For the event-study graph, we restrict the sample to countries for which data on $\mathcal{R}_t$ is available for the whole event window. Heteroskedasticity-robust confidence bounds are shown by the shaded areas.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure A.13

$\mathcal{R}_t$ and Policy Interventions: Social Distancing



Notes: The graph plots the estimated effective reproduction number ($\mathcal{R}_t$) one week before and three weeks after social distancing is encouraged in a country. The original sample consists of 14 European countries studied by Flaxman et al. (2020). For the event-study graph, we restrict the sample to countries for which data on $\mathcal{R}_t$ is available for the whole event window. Heteroskedasticity-robust confidence bounds are shown by the shaded areas.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.19.20071886; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table A.3

Effective Reproduction Number and NPIs

	(1)	(2)	(3)	(4)	(5)
Lockdown	-0.11***				
	(0.04)				
Public Events		0.07*			
		(0.04)			
School Closure			-0.15***		
			(0.05)		
Self Isolation				-0.17***	
				(0.05)	
Social Distancing					-0.14**
					(0.06)
N	840	840	840	840	840
R^2	0.87	0.87	0.87	0.87	0.87
Country FE	✓	✓	✓	✓	✓
Days-Since-Outbreak FE	✓	✓	✓	✓	✓
Mobility Controls					
Testing Controls					

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Notes: Results of panel-data regressions of the (log of) effective reproduction number ($\mathcal{R}_t$) on indicator variables that are equal to 1 after the introduction of a non-pharmaceutical intervention (NPI) and 0 before the introduction. The regressions are similar to those in Table 2 except that the intervention variables are included separately (one-at-a-time). See Table 2 for more details.