Improving Posterior Estimation of COVID-19 Policy Effects With Synthetic Control Guided Compartmental Models

Marie Charpignon
MIT

Matias Cersosimo
Stanford University

Alexandr Lenk
Stanford University

Sharon Newman
Stanford University

Chetanya Rastogi
Stanford University

Jaime Gimenez
Stanford University

James Zou
Stanford University

Abstract

It is critical to accurately estimate the impact of non-pharmaceutical mitigation policies on the COVID-19 pandemic. Direct estimation of these effects using standard epidemiological models can lead to biased estimates. This is because epidemiological model estimation is not tailored to disentangle effectively naturally occurring changes in behavior unrelated to the shelter-in-place interventions from the effect of the interventions themselves. In this paper, we propose a novel framework that combines synthetic control with compartmental models to jointly estimate policy effects. Our approach detects a 14.3% decrease in R0, the expected number of secondary infections an infected individual inflicts. This is an epidemiologically significant effect given the multiplicative nature of R0.

1 Introduction

One of the main indicators that health experts and government officials have considered when assessing the effect of shelter-in-place (SIPO) policies implemented in early Spring 2020 is the evolution of R_0, the expected number of secondary infections an infected individual inflicts, pre and post lockdown. The limitation of this approach is that the drops in R_0 could be substantially driven by naturally occurring changes in people's behavior. For instance, as people observe cases and death tolls rise, they would physically distance themselves on their own without the need of imposing a lock-down. A possible way to remedy this is to model the contagion rate explicitly as decreasing after the date a lock-down has been placed and estimate that decrease. However, it is entirely possible that there are other changes happening in people's beliefs at the exact date a lock-down is implemented independent of it. In this sense, the estimated decrease in R_0 would be upward-biased for the causal effect of the lock-down. To correct for this bias, we use Synthetic Controls to construct a counterfactual series that traces the evolution of deaths in the absence of the lock-down. We use Approximate Bayesian Computation - Sequential Monte-Carlo (ABC-SMC) techniques [9, 10, 11] to estimate a compartmental model jointly for the observed and synthetic series and obtain the corresponding R_0. We display the difference in R_0s between the observed and synthetic curves that gives the causal effect of lock-downs and indeed we are able to detect a small, but statistically significant (at 10% level) causal effect on R_0 starting 5 days after lock-down for a period of 2-weeks, reducing R_0 from 1.75 to 1.5. We fitted a Bayesian model to the observed data and synthetic curve: these type of analyses are common in epidemiology, as they can better adapt

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

to model misspecification and can better handle the uncertainty in parameter estimation of these complex models [5, 3, 4]. We are able to detect such small, but epidemiologically significant effects because of the multiplicative nature of R_0, thanks to tighter parameter posteriors obtained from the joint-estimation. While the effect of the policy seems to dissipate towards the end of our sample (ie, the R_0s from both series converge), we find evidence that the lock-down managed to push the drop in R_0 to an earlier date compared to if no lock-down had occurred, which is in line with lock-down policies trying to curb the epidemic as fast as possible to avoid exponential spreads.

2 Combining Epidemiological Models and Synthetic Controls

We consider an extended Kermack-McKendrick Model [6, 7, 8] to explicitly incorporate the fatal documented cases. We will be fitting our models on cumulative death series only as we believe reported cases are noisy and imprecise as they are subject to state testing capacities, people's willingness to get tested, etc, hence not reflecting the true state of the epidemic. Our model, named SEIRHF, particularly accounts for the distinction between detected (recovered diagnosed, and therefore documented) and undetected (recovered undiagnosed, and therefore not documented) infections. The compartmental model and its dynamics are depicted in Fig. 1 of the Appendix along with the dynamical equation system. Notably, we assume a time-varying effective contagion rate, $\frac{1}{\bar{\tau}_{Con}(t)}$, where $\bar{\tau}_{Con}(t) = \rho(t)\frac{N}{N_S(t)}$ with N total population size and N_S size of susceptible population at time t. $\rho(t)$ captures how the contagion rate evolves over time and follows the equation[1] $\rho(t) = \rho_{init} + \rho_{isol}(t)\log(1+t)$, where ρ_{init} is the structural (biological) transmission probability that is not affected by behavior change whereas ρ_{isol} captures the effect of behavior change that is a combination of people becoming more cautious as the epidemic progresses and the direct effect of restrictions on reducing social contacts such as ban on gatherings or business closures. We assume that the belief adjustment process is progressive rather than sudden, with steeper adjustments in the beginning, hence the use of $\log$.[2] We are also focusing on the first few weeks following the start of the epidemic, which justifies the assumption that $\rho(t)$ is monotonically increasing.[3] As can be seen, it is hard to disentangle which part of ρ_{isol} is due to naturally changing behavior and which part is directly related to lock-downs. A partial remedy is to assume a discrete jump in $\rho_{isol}(t)$ starting at the date of lock-down implementation. In this first version of our paper we fix for simplicity $\rho_{isol}(t) = 0$ for all periods before lock-down and then we estimate it for all periods thereafter by restricting it to be constant. This may still not be satisfactory, however, as the lock-down date could be endogenous to people's behavior: it could be that lock-downs are imposed as a response to people starting to panic and distance themselves, which means people would have reduced contacts in the absence of the lock-down anyway.

To capture the pure effect of the lock-down, we need a counterfactual series that shows the evolution of deaths had no lock-down been imposed. We construct such counterfactual curve using Synthetic Controls ([1], [2]). Appendix 1.4 details this approach, but essentially we need to have a pool of N areas for T periods, with the N-th area exposed to the policy at period t^*, called the treated area, and the remaining $N-1$ areas, called control areas, never exposed to the policy. We define the time indicator t with respect to the day the first death was recorded in a given area rather than calendar dates as different areas could be at different points in their epidemic trajectory at the same calendar date. We use all pre-intervention periods for $t \in \{0, 1, ..., t^* - 1\}$ when none of the areas have been treated by the policy and find the linear combination of control areas that traces best the trajectory of the treated area. Then, we use those weights and the outcomes for the control areas post-intervention to construct a counterfactual series for the treated area that displays how the curve would have evolved had no lock-down policy been imposed. This is because the control areas have not been exposed to the policy in the post-intervention period. We use Washington-Arlington-Alexandria (WAA) metropolitan area as a case-study in this paper since it was hard to find valid control units for other areas as we explain in Section 1.4 of the Appendix. We construct our data series by pooling fatality series from the counties belonging to this metropolitan area. We use log deaths as exponential curves are not well-justified by the assumptions underlying the Synthetic Control model. Our log

[1] In theory, it could happen that the contagion rate could still be increasing even if $\rho(t)$ increases because more isolation leads to a higher share of susceptible. This effect, however, is of second-order in our estimations.

[2] We are in the process of validating this pattern using mobility data.

[3] It is perfectly possible that people might get tired of isolation in later periods so that they actually increase their interactions, driving $\rho(t)$ down. Our series, however, ends in late-April before relaxations start.

death series starts on March, 22nd (five days after the area recorded its first death to correct for initial noise in reporting) and ends 39 days later on April, 29. Applying the Synthetic Control method, we obtain the curve in Fig. 1.

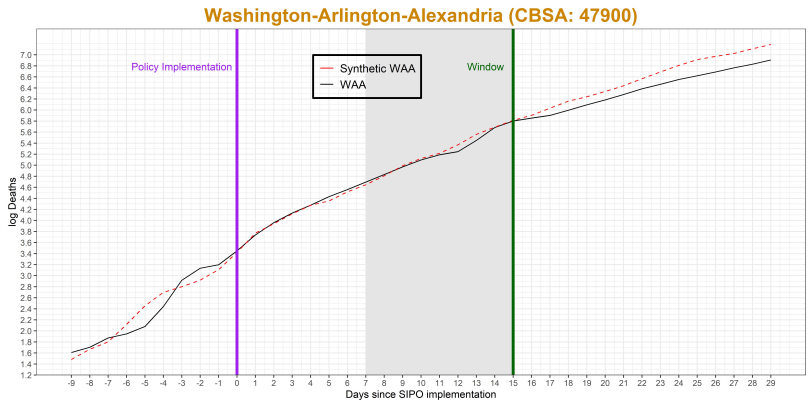

Figure 1: Evolution of Log Deaths, Observed vs Counterfactual Curves

The lock-down in the Washington-Arlington area occurs 9 days after the start of our series. Since the effect of the lock-down on deaths occurs with a two-to-three-week delay, we set the "effective policy date" to occur 16 days after the implementation of the lock-down. Hence, we have 9+16 = 25 time periods upon which we perform weight estimation and validation. We use the first 16 periods out of the 25 for estimation and remaining 9 for validation (the shaded area). We give the estimation details and the set of control units used as well as a series of robustness checks in the Appendix. Inspecting Fig. 1, it is reassuring to see that the synthetic curve tracks the observed curve almost exactly both during training period and validation period. Seeing a large disparity between the two curves would invalidate our method as we should not be seeing divergence before the expected effect of the policy kicks in. We indeed see that around two weeks after lock-down, the slope of the death series for the observed curve is lower than that of the synthetic curve which implies that lock-downs did have an incremental effect on curbing deaths independent of naturally occurring changes in people's behavior. The reduced form synthetic control treatment effect calculated as the average difference between the two curves post intervention (after the green bar) equals (after transforming the logs back to counts) -150.65 deaths. That is, on average, 151 lives have been saved due to the lock-down during our sample period. This reduced-form estimate, however, is hard to interpret for policy makers, as they are interested in how the lock-down affects the evolution of the epidemic more generally.

Hence, we use both curves as inputs to our KM-M model and derive the corresponding R_0s. We restrict all model parameters to be shared across the two series except for ρ_{isol} that we set equal to 0 for the synthetic series all throughout the period as by design, the synthetic series assumes there is no policy effect.[4] For estimation, we use the state-of-the-art variant of the ABC procedure, called Approximate Bayesian Computation - Sequential Monte-Carlo (ABC-SMC) [11], with the L1 loss function on fatal cases for training. Our error for out-of-sample validity is within the range of errors that CDC uses for their ensemble forecasts (see estimation details and model performance comparisons in Appendix 1.2).

3 Estimation Results

We now show in Fig. 2 how R_0 evolves over time both for the observed and counterfactual curves along with 90% confidence bands obtained from the Bayesian posteriors. We also present in Fig. 3 the log ratios between R_0 observed and R_0 synthetic to be able to track the effect of the policy more directly.

[4] We believe that the rest of the parameters such as hospitalizations, death rates, incubation and recovery rates are related to biology and state capacities that could not have been moved by lock-downs although we do recognize that expansion of hospital beds and medical supplies could also have affected some of those parameters, something we aim to account for in future versions of this paper.

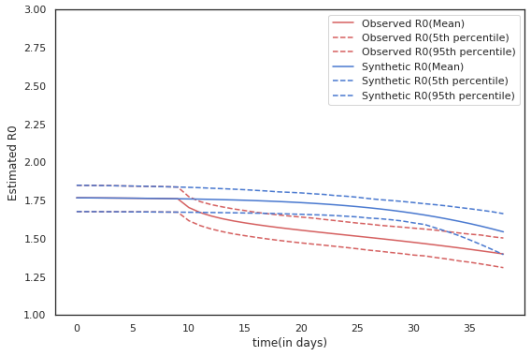

Figure 2: Evolution of R_0

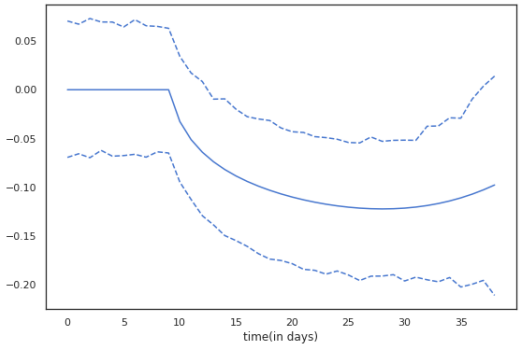

Figure 3: Evolution of $\log \frac{R0_{observed}}{R0_{synthetic}}$

First, remember that the effective contagion rate is set to be identical for both curves by design before lock-down date which occurs at Day 10 in our sample, hence the curves are identical up to that day. After Day 10, we observe that both curves trend downwards, which implies that part of the decreases in R_0 is driven by naturally occurring behavioral changes. However, we see that the R_0 for the observed series always tends to be below the R_0 for the synthetic series. This suggests that the lock-down has an incremental effect on R_0. According to Fig. 3, this effect starts to be significantly different from 0 on Day 15 (5 days after lock-down) and loses significance two weeks later. This can be due to the fact that the variance in our posteriors at the end of the sample does not allow us to detect that persistence. We need to mention, however, that our posteriors for R_0 of the observed series is much tighter when estimated using both the observed and synthetic series as common inputs to the model compared to using the observed series only (see Fig. 5 of the Appendix). This effect could be purely mechanical as we are simply adding an extra input to the model or more interestingly, it could be because we are adding extra information in the model, in particular, the fact that both curves share the same model structure, evolution dynamics and most model parameters as explained above. Had we relied on estimating the R_0 for the observed series using the observed series only rather than the joint-estimation, we would not be able to detect the small effects we observe in a statistically significant way, even though such effects could be substantial because of the exponential nature of the spread of the epidemic. [5]

While Fig. 2 shows that the magnitude of the lock-down effect seems to be small (1.75 vs. 1.5), a persisting effect of such magnitude can still generate big effects on the spread of the epidemic because of the multiplicative nature of R_0. We do see, however, from Fig. 2 that the synthetic curve starts dropping faster than the observed curve at the end of our series. So, it appears that R_0 would have eventually dropped even in the absence of lock-down, it is just that the lock-down makes the drop occur earlier. This seems to be in line with the goal of the policy that aimed at halting the spread as fast as possible, thus avoiding large exponential increases in infections and the saturation of the health systems.

4 Conclusion

In this paper, we combine the best of two worlds from epidemiological and synthetic control models to assess causal effects of non-pharmaceutical mitigation policies on the spread of the COVID-19 pandemic. While synthetic controls are good at estimating average causal effects on observed outcomes such as death counts, compartmental models are able to estimate epi-parameters that are more useful to understand the evolution of the epidemic. We focus on the effect of shelter-in-place policies on R_0 and take the Washington-Arlington-Alexandria (WAA) metropolitan area as a case-study. We construct a counterfactual death count series for WAA using synthetic controls that shows the evolution of death counts in the absence of a lock-down. We then jointly estimate the R_0s series for the observed and synthetic series via an extended Kermack-McKendrick Model using

[5] We still remain cautious about the gains in terms of reducing posterior variance, however, since we recognize that the synthetic control series is itself estimated with uncertainty, and incorporating this extra uncertainty in the estimation of the epi-parameters could adversely affect the shape of their posteriors. Unfortunately, up-to-date, there is no clear way in the Synthetic Control literature to quantify this uncertainty.

Approximate Bayesian Computation methods. We see that the lock-down starts reducing the R_0 5 days after its implementation from about 1.75 to 1.5 and the reduction is persistent at least for two more weeks. We cannot reject the hypothesis that the effect dissipates in the fourth week following the start of the lock-down (that might be due to high posterior variance), although we find evidence that the lock-down resulted in moving the drop in R_0 to an earlier date, which is in line of the goal of such restrictive measures.

References

[1] A Abadie. Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 2019.

[2] Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association*, 105(490):493–505, 2010.

[3] Igor Burstyn, Neal D Goldstein, and Paul Gustafson. Towards reduction in bias in epidemic curves due to outcome misclassification through bayesian analysis of time-series of laboratory test results: Case study of covid-19 in alberta, canada and philadelphia, usa. *BMC Medical Research Methodology*, 20:1–10, 2020.

[4] Sander Greenland. Bayesian perspectives for epidemiological research: I. foundations and basic methods. *International journal of epidemiology*, 35(3):765–775, 2006.

[5] Prashant K Jha, Lianghao Cao, and J Tinsley Oden. Bayesian-based predictions of covid-19 evolution in texas using multispecies mixture-theoretic continuum models. *Computational Mechanics*, pages 1–14, 2020.

[6] William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

[7] William Ogilvy Kermack and Anderson G McKendrick. Contributions to the mathematical theory of epidemics. iii.—further studies of the problem of endemicity. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 141(843):94–122, 1933.

[8] WO Kermack and AG McKendrick. Contributions to the mathematical theory of epidemics: V. analysis of experimental epidemics of mouse-typhoid; a bacterial disease conferring incomplete immunity. *Epidemiology & Infection*, 39(3):271–288, 1939.

[9] Emmanuel Klinger and Jan Hasenauer. A scheme for adaptive selection of population sizes in approximate bayesian computation-sequential monte carlo. In *International Conference on Computational Methods in Systems Biology*, pages 128–144. Springer, 2017.

[10] Tina Toni and Michael PH Stumpf. Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics*, 26(1):104–110, 2010.

[11] Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael PH Stumpf. Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6(31):187–202, 2009.

Improving Posterior Estimation of COVID-19 Policy Effects With Synthetic Control Guided Compartmental Models

Marie Charpignon
MIT

Matias Cersosimo
Stanford University

Alexandr Lenk
Stanford University

Sharon Newman
Stanford University

Chetanya Rastogi
Stanford University

Jaime Gimenez
Stanford University

James Zou
Stanford University

1 APPENDIX

1.1 Data Source Description

In this work, we use reported time series of deaths, as they are believed to be better documented than infections. In addition, the reporting of deaths should be less heterogeneous across regions than infections, in that the latter strongly depends on the timing of testing deployment and capacity, as well as on health-related behavior (e.g., toward seeking care). However, the cumulative number of deaths also has imperfections. Hence, in the future, sensitivity analysis could be conducted to adjust for the suspected under-reporting of the total number of deaths due to SARS-CoV-2 infection [14, 18, 4]. Our models leverage death data as provided by USA Today [21], which are routinely collected from the John Hopkins Coronavirus Resource Center (JHU) [6], at the county-level and state-level. Note that all COVID-19 death counts are provisional, and may also be defined differently in various reporting and surveillance systems (e.g., JHU, NYT [19], USAFacts [8], and the CDC [3]). There are mainly two reporting conventions: occurrence deaths and residential deaths. Occurrence deaths reflect the location where a person died, while residential deaths are based on a person's permanent address. As a matter of fact, the CDC [5] and NYT [19] appear to use occurrence data at the county-level, while JHU [6] appears to use residential data. People from surrounding communities may be coming to hospitals in counties with a higher prevalence of healthcare facilities or better medical centers, thereby increasing the total local death count per capita. In other words, in counties with hospitals, occurrence deaths would certainly be much higher than their residential deaths. These considerations on data quality and definition of the number of COVID-19 reported deaths are essential [5], as county-level death trajectories are the main input of our model based on Bayesian approaches.

In our analysis, we consider an intermediate geographical unit of interest, the Core-Based Statistical Area (CBSA) [11]. A CBSA is a set of one to several counties (from 1 to 40), including a urban epicenter of at least 10,000 inhabitants as well as neighboring counties that are socioeconomically tied to the urban center, mainly via commuting (e.g., residential suburbs). Considering CBSAs, which are compromised of both metropolitan (MSAs) and micropolitan (μSAs) statistical areas, is particularly relevant to pandemics, as large and densely populated urban centers are often hit first by waves of infectious diseases, and the covid19 epidemic is no exception (first Wuhan in Asia, then Milan and Madrid in Europe, and Seattle, New York City, Detroit and New Orleans in the U.S.) [9]. In the U.S. and Puerto Rico, there is a total of 388 MSAs (more than 50,000 people) and 541 μSAs (at least 10,000 and fewer than 50,000 people). So far, our modelling focus has been on metropolitan

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

areas that have been hardest hit, in terms of the total number of documented cases per 100,000 people. Note that certain counties belonging to rural areas are not part of any CBSA (1,226 out of 3,141 counties in total, i.e., 39%). Beyond the fact that CBSA-level units may be more homogeneous and therefore more comparable - at least within their respective metropolitan and micropolitan groups - because of similar levels of urbanity, the advantage of considering an intermediate geographical resolution is to partially circumvent the inherent noise of county-level time series - aggregation helping smooth out the cumulative death count at the CBSA-level.

We chose specifically to study CBSAs that represent (1) diverse communities around the US (2) featured different timelines for adoption of shelter-in-place orders and (3) suffered the largest number of cases per 100,000 people. In order to capture diversity between communities we chose CBSAs that fall into different groupings from the American Communities Project. [12] The American Communities Project divides US counties into regions that differ among dimensions such as age, race, income, religious affiliation and more. There are 15 different categories and some of these are 1) African American South, 2) Aging Farmlands, 3) Big Cities, 4) Evangelical Hubs, 5) Middle Suburbs, and more. We also split CBSAs into groups based upon when they enacted shelter-in-place orders. A CBSA is sorted into the early adopter category if its sate went into shelter-in-place before March 25th, average-speed adopter between March 25th and April 1st, and late-speed adopter otherwise. If a CBSA include counties in different states, it's category is designated based on the the state that contain the majority of its urban area.

Given these 45 groupings (15×3), we chose CBSAs that had the highest number of cases per 100,000 people. We further had to limit our study locations to areas which have adequate data available. We have only just begun our analysis of this group of chosen CBSAs, and for the purposes of this report are focusing on a case study of a single CBSA that has yielded good and stable results thus far: the Washington-Arlington-Alexandria DC-VA-MD-WV CBSA. While USA Today does not report deaths by CBSA, we construct our death dataset at CBSA level by pooling the counties belonging to that particular CBSA. The Washington-Arlington-Alexandria DC-VA-MD-WV CBSA covers several counties, and includes multiple counties in the American Communities Project groups of Exurbs, Military Posts, and Urban Suburbs, and single counties in each of the categories of Graying America, College Towns, and Big Cities. Distinct features of Exurbs are that they lie on the outskirts of major metropolitan areas and are wealthy with a median annual income of $65.5K. Military Posts are young with only 13% of the population over 62. Finally, Urban Suburbs are densely populated, diverse, and wealthy, with each having roughly 500,000 inhabitants, being 58% white, and with an average median annual household income of $67.8k.[1]

Specifically for the CBSA of Washington-Arlington-Alexandria DC-VA-MD-WV, according to USA Today[21], the first case was declared on March 6th, the first death occurred 12 days later on March 18th. Our data sample ends on April, 29th. We ended the data sample at that point since that was the time period where some regions started re-opening, mobility started increasing and people become more habituated to the new environment, factors that would require more complex modelling choices for the contagion rate in our model, which leave for a later version.

1.2 Model Details

1.2.1 Model Description

We consider an extended Kermack-McKendrick Model [15, 16, 17] to explicitly incorporate the fatal documented cases and hospitalizations (Fig. 1).

[1] For reference, the real median household income was $59k in 2016 [13]

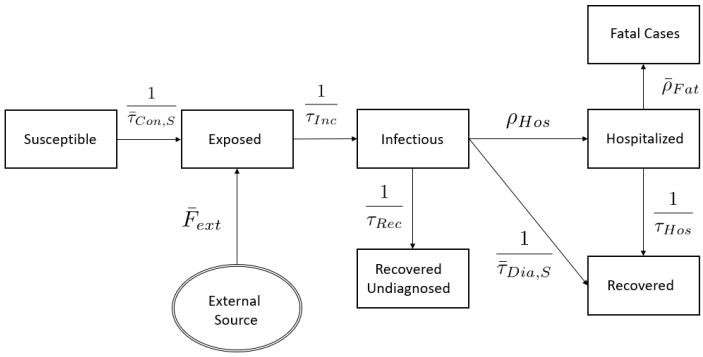

Figure 1: SEIRHF model

The latent compartments in our SEIRHF model include :

$N_S(t) :=$ Number of susceptible individuals at time t.
$N_E(t) :=$ Number of exposed individuals at time t.
$N_{IS}(t) :=$ Number of infectious symptomatic individuals at time t.
$N_{UR}(t) :=$ Number of recovered patients at time t that have not been diagnosed at any point.
$N_H(t) :=$ Number of hospitalized individuals at time t.
$N_R(t) :=$ Cumulative number up to time t of recovered patients with a positive diagnosis.

We include number of hospitalized as a latent compartment as in the specific case of Washington, DC, the website only publishes current and not cumulative hospitalizations. In future versions, we would like to incorporate information on hospitalized as well as recovered patients. Hence, the only observed compartment of the model includes:

$$N_F(t) := \text{Cumulative number of fatal cases up to time } t.$$

We posit that the dynamics of these compartments can by modeled by a linear system of first order ODEs:

$$\frac{d}{dt}N(t) = A(t, N(t)) * N(t) + b(t)$$

with the initial conditions $N_S(0) = N_0$ representing the initial total population, where we define the (time-location varying) matrix A of coefficients, vector N, and the vector b as follows.

$$N(t) = \begin{pmatrix} N_S(t) & N_E(t) & N_{IS}(t) & N_{UR}(t) & N_H(t) & N_F(t) & N_R(t) \end{pmatrix}^T$$

$$A(t,N) = \begin{pmatrix} 0 & 0 & -\bar{\tau}_{Con,S}^{-1} & 0 & 0 & 0 & 0 \\ 0 & -\tau_{Inc}^{-1} & \bar{\tau}_{Con,S}^{-1} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\tau_{Inc}} & -(\tau_{Rec}^{-1} + \bar{\tau}_{Dia,S}^{-1} + \rho_{Hos}) & 0 & 0 & 0 & 0 \\ 0 & 0 & \tau_{Rec}^{-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_{Hos} & 0 & -(\tau_{Hos}^{-1} + \bar{\rho}_{Fat}) & 0 & 0 \\ 0 & 0 & 0 & 0 & \bar{\rho}_{Fat} & 0 & 0 \\ 0 & 0 & \bar{\tau}_{Dia,S}^{-1} & 0 & \tau_{Hos}^{-1} & 0 & 0 \end{pmatrix}$$

$$b(t,N) = \begin{pmatrix} 0 & \bar{F}_{ext} & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^T$$

The parameters defining this equation are characterized as follows, with superscript bars indicating the corresponding terms are functions of (t, N).:

$\bar{F}_{ext}$:= Incoming individuals exposed to the virus from an external system.

$\bar{F}_{ext,date}$:= The date at which the incoming individuals exposed to the virus from an external system start affecting the suscep

$\bar{\tau}_{Con,S}$:= Typical time it takes for a symptomatic patient to infect a susceptible person.

$\bar{\tau}_{Dia,S}$:= Typical time until diagnosis of a symptomatic patient.

τ_{Inc} := Typical time representing the incubation period.

τ_{Rec} := Typical time representing the infectious period until recovery.

ρ_{Hos} := Rate of diagnosed symptomatic patients requiring hospitalization.

τ_{Hos} := Typical time to recovery for hospitalized patients.

$\bar{\rho}_{Fat}$:= Rate of fatal cases among the hospitalized patients.

We assume that $\bar{\tau}_{Con,S}(t)$ is time-varying and posit the following equation:

$$\bar{\tau}_{Con,S}(t) = \rho(t)\frac{N}{N_S(t)}$$

with $\rho(t) = \rho_{init} + \rho_{isol}(t)\log(1+t)$ The first term, $\rho(t)$ reflects the structural probability of infection independent of people's behavior whereas the second term, $\rho_{isol}(t)$ depends on people's behaviour and implemented policies. More cautious behavior or stricter isolation policies limit interactions between individuals, which reduces the effective contagion rate. We multiply $\rho_{isol}(t)$ by $\log$ as we believe such behaviour adjustment process is progressive and increases at a concave rate, which is well reflected by the functional form implied by the $\log$. In future versions, we plan to more rigorously micro-found the functional form of $\bar{\tau}_{Con,S}(t)$ from an optimization problem where an agent maximizes her utility function equalizing marginal costs and benefits at the optimum.

The prior distributions of the parameters are set as follows:

Parameter	Distribution name	Arguments	Source
$F_{ext,date}$	uniform	0 to 14	[23]
F_{ext}	uniform	0 to 15	uninformative prior covering the true posterior [10]
ρ_{init}	uniform	0.2 to 0.6	uninformative prior covering the true posterior [10]
ρ_{isol}	uniform	0 to 5	uninformative prior covering the true posterior [10]
$\bar{\tau}_{Dia,S}$	uniform	1 to 4	uninformative prior covering the true posterior [10]
$\bar{\rho}_{Fat}$	uniform	0.05 to 0.33	[20]
ρ_{Hos}	uniform	0.02 to 0.25	[20]
τ_{Hos}	uniform	28 to 35	[23]
τ_{Inc}	fixed	5.06	[23]
τ_{Rec}	fixed	14	[23]

Table 1: Prior distribution of the parameters

To estimate the model parameters based on observed data series, we use the state-of-the-art variant of the ABC procedure, called Approximate Bayesian Computation - Sequential Monte-Carlo (ABC-SMC) [22], with the l1 loss function on fatal cases for training. ABC allows us to recover a set of parameters that we hope best explains the data. Having defined a set of prior distributions $p(\theta)$ over the set of parameters, one tries to recover the posterior distribution $p(\theta|X)$ over these parameters having observed the data. The usual Bayes formula $p(\theta|X) \propto p(X|\theta)p(\theta)$ cannot be used here as we do not have a tractable likelihood. Thus, ABC approximates the posterior distribution through a rejection sampling scheme. For a given random sample of a set of parameters, the procedure evaluates the mismatch between the corresponding generated sample and the observed data. An acceptance-rejection criterion is then used to either keep the set of parameters or reject them. Doing this successively while making the acceptance criterion more stringent, leads to an algorithm that outputs refined sets of parameters that increasingly match the observed data. We describe the priors we use as well as the parameters we fix based on prior medical literature that has shown those to be consistent across geographical location in the Appendix.

1.2.2 Model Performance

We compare our model's out-of-sample validity to other models upon which the CDC makes ensemble forecasts on evolution of death. We obtain the other model errors from the Covid19 Forecast Hub Evaluation GitHub Repository (https://github.com/youyanggu/covid19-forecast-hub-evaluation#evaluation). We follow the same rules to calculate the 2-week ahead forecast error ϵ for Washington-Arlington-Alexandria metropolitan area[2]:

$$\epsilon = \frac{(\hat{D}_{evaluation\ date} - D_{projection\ date-1}) - (D_{evaluation\ date} - D_{projection\ date-1})}{(D_{evaluation\ date} - D_{projection\ date-1})} * 100$$

where D is ground true death, $\hat{D}$ is predicted death, evaluation date is the last date of the forecast and projection date is the first date of the forecast. The Covid19 Hub gives forecast errors on a weekly basis starting April, 20. Since our data sample ends April, 29th, we evaluate our model by comparing the 2-week ahead forecast error calculated on April, 27 for the time period April, 27-May, 9. Accordingly, we set our evaluation date to May,9 and projection date to April, 27 (so that day before projection date corresponds to April, 26). We compare the forecast errors in Table 2. We see

Model	**Error**
Baseline	19.3
COVIDhub-baseline	2.0
CovidAnalytics-DELPHI	-8.0
IowaStateLW-STEM	25.9
IHME-CurveFit	-63.3
LANL-GrowthRate	-18.5
UT-Mobility	44.4
UMass-MechBayes	45.0
YYG-ParamSearch	-25.9
CurrentModel	39.5

Table 2: Model Comparisons across Various Models.

that our model is within the error range of other models, and performs reasonably well as compared to others. Our goal for future versions is to drive that error down so that we get higher quality estimates and hence results.

1.3 Derivation of $\mathcal{R}_0$

We follow closely Chapter 6 of Mathematical Epidemiology. We omit the reference to time for simplicity. For our model, we take the Exposed and Infected compartment, E and I as disease compartments. Note that this leaves out the Hospitalized compartment from the set of disease compartments, and while we recognize that the evolution in that compartment is highly relevant for the dynamics of the reproduction rate (since hospitalized individuals can expose susceptible ones), we are not considering the case for this version of the paper.

We define the next generation matrix for the system at the disease-free equilibrium K as FV^{-1}, where F is the matrix representing the dynamics between the non-disease and disease compartments, and V represents the dynamics between the disease compartments. It is important to note that for our model, once individuals leave the infected compartment, they do not affect the dynamics of the disease anymore (cannot become Susceptible/Exposed/Infected again), so there is no interaction between Hospitalized and subsequent compartments with the disease compartments. Therefore, in our case (recall that $\tau_{Con,S} = \rho\frac{N}{N_S}$. That is, it accounts for the size of the susceptible compartment S).

$$F = \begin{bmatrix} 0 & \frac{1}{\tau_{Con,S}} \\ 0 & 0 \end{bmatrix}, \quad V = \begin{bmatrix} \frac{1}{\tau_{Inc}} & 0 \\ -\frac{1}{\tau_{Inc}} & \frac{1}{\tau_{Rec}} + \frac{1}{\tau_{Dia,S}} + \rho_{Hos} \end{bmatrix}$$

[2]These models only calculate state-wide errors rather than metropolitan area errors, so for comparison we look at the errors for District of Columbia, which we believe is a good approximation as DC is the major component of the Washington-Arlington metropolitan area

F_{ij} is the rate at which the incoming individuals affect compartment i through compartment j. There is only one non-zero element, since the only way individuals enter the dynamics of the disease is by the exposed compartment through the infected compartment (as it interacts with the susceptible compartment).

In turn, V_{ij} is the rate at which exiting individuals affect compartment i through compartment j. Individuals in the exposed compartment exit and enter the infected compartment at a rate $\frac{1}{\tau_{Inc}}$, and there is no reentering of individuals from the infected compartment to the exposed compartment. The lower right corner represents is the rate at which individuals in the infected compartment exit the dynamics of the disease.

Then, we have that:

$$K = FV^{-1} = \frac{1}{\tau_{Con,S}} \times \begin{bmatrix} (\frac{1}{\tau_{Rec}} + \frac{1}{\tau_{Dia,S}} + \rho_{Hos})^{-1} & (\frac{1}{\tau_{Rec}} + \frac{1}{\tau_{Dia,S}} + \rho_{Hos})^{-1} \\ 0 & 0 \end{bmatrix}$$

Which yields:

$$K = \frac{1}{\tau_{Con,S}} \left(\frac{1}{\tau_{Rec}} + \frac{1}{\tau_{Dia,S}} + \rho_{Hos} \right)^{-1} \times \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

$$eig(K) = \frac{1}{\tau_{Con,S}} \left(\frac{1}{\tau_{Rec}} + \frac{1}{\tau_{Dia,S}} + \rho_{Hos} \right)^{-1} \times eig\left(\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \right)$$

It directly follows from the expression above that the matrix part of K has two eigenvalues, 1 and 0. Since 1 is the eigenvalue with the largest modulus, then, the reproductive factor will be equal to:

$$\mathcal{R}_0 = \frac{1}{\tau_{Con,S}} \left(\frac{1}{\tau_{Rec}} + \frac{1}{\tau_{Dia,S}} + \rho_{Hos} \right)^{-1}$$

.

1.4 Synthetic Control Details

1.4.1 General Approach to Synthetic Controls

Synthetic Controls have become "arguably the most important innovation in the policy evaluation literature in the last 15 years" according to Athey & Imbens (2017) [2]. In basic causal inference language terms, the treatment effect τ of a policy on individual i is the difference between the individual's potential outcome under the policy $Y_i(1)$ and her potential outcome with no policy $Y_i(0)$, i.e., $\tau = Y_i(1) - Y_i(0)$. We know from the fundamental problem of causal inference that we can only observe individual i under a single policy regime given some point of time. Hence, the role of the researcher is to come up with the best **counterfactual** possible that approximates the potential outcome of individual i under the alternative policy regime that is unobserved. Synthetic Controls constitute one such method for data with panel structure. We now present the basic set-up we are facing in the context of COVID-19. We observe N areas for T periods of time. The N^{th} area is treated at time T^* with a lock-down policy while the remaining $N-1$ areas are never treated during these T periods. Let $\mathbf{Y_0}$ be an $N \times (T^* - 1)$ matrix which includes the observed outcomes of all units in the pre-intervention period. We also have $\mathbf{Y_{01}}$, an $N - 1 \times (T - T^*)$ matrix containing the outcomes for the control units in the post-intervention period and $\mathbf{Y_{11}}$, a $1 \times (T - T^*)$ matrix containing the outcomes for the treated unit in the post-intervention period. Notice that $\mathbf{Y_{11}}$ contains the potential outcomes for the treated unit under treatment. Our goal is to use the data in $\mathbf{Y_0}$ such that we can re-construct $\mathbf{Y_{10}}$, the potential outcomes of the treated unit post-treatment had the unit not been treated, from $\mathbf{Y_{01}}$. To achieve this, we first need to make an assumption that $\mathbf{Y_0}$ can be approximated by a lower dimensional matrix $\mathbf{L}$ and some noise ϵ:

$$\mathbf{Y_0} = \mathbf{L} + \epsilon$$

Under this assumption, there exists a representation of $\mathbf{L}$ as a linear factor model:

$$\mathbf{L} = M\Lambda$$

where M is an $N \times R$ matrix and Λ is a $R \times (T^* - 1)$ matrix. Both M and Λ are latent unobserved matrices with M containing the vector of unit-specific effects fixed over time and Λ the vector of time effects that are common to all units but change across time. Remember, pre-intervention, none of the units have been treated such that $Y_0 = Y(0)$, i.e., all pre-intervention observed outcomes are equal to the potential outcome of the units without treatment. So, for each area i and $t < T^*$, the model becomes[3]:

$$Y_{it} = L_{it} + \epsilon_{it} = \sum_r \mu_{i,r} \lambda_{r,t} + \epsilon_{it}$$

The goal of synthetic controls is to find area-specific weights such that:

$$L_{Nt} = \sum_{i=1}^{N-1} \omega_i^* L_{it}$$

Clearly, we can never obtain $\{\omega_i^*\}_{i=1}^{N-1}$ since L is unobserved. Rather we observe L and noise. Hence, the best we could do is to obtain proxy weights that satisfy:

$$Y_{Nt} = \sum_{i=1}^{N-1} Y_{it} = \sum_{i=1}^{N-1} \omega_i L_{it} + \sum_{i=1}^{N-1} \omega_i \epsilon_{it}$$

It becomes clear from the above formula that the proxy weights $\{\omega_i\}_{i=1}^{N-1}$ are not ideal since they will necessarily fit to noise as well. Hence, we need to feel confident that we work with series that have less noise (variance of ϵ is small), we have a large enough number of pre-treatment periods relative to the number of control units (this is both to average out noise over time as well as to have enough time variation coming from L in order to be able to estimate the weights precisely), and we also have access to good controls (we do not want the controls to be too different from the treated unit, i.e., if control outcomes are too different, it could be that the heterogeneity comes from the idiosyncratic shock ϵ rather than L which is what we are matching on). For details, check the tutorial by Abadie (2019) [1]. To estimate the weights in practice, we need to solve the following constrained linear optimization program:

$$\min_{\{\omega_i\}_{i=1}^{N-1},\, \omega_0} \sum_{t=1}^{T^*-1} \left(Y_{N,t} - \omega_0 - \sum_{i=1}^{N-1} \omega_i Y_{i,t} \right)^2$$

Now, we could impose various types of constraints depending on what we think makes sense:

1. The original SC paper imposes:

$$\sum_{i=1}^{N-1} \omega_i = 1, \omega_i \geq 0$$

   This makes intuitive sense since these constraints restrict the set of predictions to the convex hull of the donor states (modulo the systematic additive difference thanks to the unit intercept, ω_0) which reduces the danger of extrapolation and give interpretability to the weights. Furthermore, the standard SC weights act as a natural regularizer thanks to introducing sparsity: if the treated outcome is outside of the convex hull of the control units, the weights ensure that the estimated counterfactual becomes a projection of Y_{Nt} onto one side of the convex hull of controls, which automatically excludes some controls. The obvious limitation is that if the treated unit already is inside the convex hull, then there might be an infinite number of solutions and a general regularization technique might be needed to select the best one.

[3] To exemplify, μ_i could be a vector of size 2 that contains residents' propensity to hug each other in a given state and the proportion of inhabitants with a genetically low immunity system. These characteristics clearly differ across states, but they are all constant across time. Now, let us assume $\lambda_{1,t}$, is positive and increasing over time whereas $\lambda_{2,t}$ is positive and decreasing overtime. The interpretation is that the more socially proximate residents are in a given state, the more easily the disease will be transmitted. This effect will be even stronger over time, since more and more people become infected overtime such that hugging becomes more dangerous over time. On the other hand, the more people there are with low immunity, the more vulnerable the population is to COVID-19; however, as it spreads more and more, residents achieve herd immunity, such that having genetically low immunity no longer has such a strong effect on spreading the disease.

2. A general approach that is closer to standard training algorithms in Machine Learning is to not impose particular constraints on the weights, but rather impose a flexible regularizer such as Elastic Net. The researcher then tunes the regularization parameters on a train set and validates the fits on a validation set. This is obviously more demanding on the size of the pre-intervention period as we need to ensure we keep enough pre-treatment periods for training.

In this report, we experiment with both approaches. We now explain how we approached the Synthetic Control analysis in our particular context.

1.4.2 Application of Synthetic Controls

Remember again the linear factor model for the potential outcome under no treatment:

$$Y_{it}(0) = \sum_r \mu_{i,r}\lambda_{t,r} + \epsilon_{it}$$

The model imposes that λ_t be uniform across units (common time shocks) so that the difference in observed outcomes across units for a given t depends on how these time-shocks are differentially amplified at each period t which depends on the μ_is which are unit-specific. μ_i could for instance include the share of vulnerable population in area i or cultural practices (such as hugging, kissing, etc.), all are plausible determinants for the spread of the virus. The role of SC weights is indeed to construct a copy of the treated unit which will be equivalent in terms of the amplification effect which depends on μ_i. Hence, it is critical to ensure that the assumption on the uniformity of λ_t across units is satisfied and that is why we need to pay particular attention to how we define our time dimension. It makes sense, then, to work with the series in logs, as this assumption is more likely to apply in this context of exponential raw outcomes. Ultimately, it boils down to pick the correct starting date of the epidemic, T_0. Picking T_0 correctly becomes all the more important as the sensitivity analysis of the epidemiological model above has shown that along with the isolation effect, the start date of the epidemic contributes most towards the loss function. Ideally, we would set T_0 as the day at which patient 0 gets infected and starts spreading the virus. Hence, it becomes clear that using calendar dates is not a great choice as this would amount to assuming that patient 0 occurred at the same calendar date for all units. It might be tempting to believe this holds approximately since by looking at case time series across states, it seems like most states report their first documented cases around the same time in early March. However, the uniform documented case reporting could just be due to uniform testing capacities across states at the initial stages and might not be revealing of the actual spread of the epidemic.

Rather than looking at documented cases, we propose to look at time series of reported deaths to infer T_0, as death reporting is more accurate and less dependent on reporting or testing capacity. We believe the day of first reported death is informative about T_0. For illustration, assume a mortality rate of 0.01. Then, at their respective T_0, all units will have experienced around 100 documented cases three weeks earlier relative to T_0 (as according to epidemiologists the average time between contracting the virus and dying is three weeks [7, 24]). We should also keep in mind, however, that areas which have reported more deaths in the first few days since the first death probably had many more infections three weeks ago, as compared to areas that reported only a few deaths in the first week after the first recorded death. Hence, we will ensure to select control areas that had a similar number of deaths in the first week since the first death as our treated unit. We give specific details of our approach by conducting a case study for the Washington-Arlington-Alexandria metropolitan area.

The shelter in place order (SIPO) for the Washington-Arlington metropolitan area was implemented on April 1st, or 14 days after its first registered death. In addition, we know from epidemiological evidence that death rates should not be affected earlier than 20 days since the implementation of the policy [7, 24], hence we could use $14 + 20 = 34$ periods for estimating the weights. However, we choose to be conservative and select a placebo window of only 15 periods such that we have 29 periods for estimation. In addition, we decide to cut the first five days of the series because of noise in reporting during the first days. We also decide to set aside the last 8 periods of the estimation period for validation (where we would be testing our out-of-sample fit). Overall, we are left with

16 time periods for the estimation of weights. With so few periods for estimation, we cannot fully rely on letting the training process choose which areas matter most for the counterfactual and we should use our judgment to further restrict the set of potential controls by design. First, we restrict the sample to metropolitan areas only. Metropolitan areas are, by their definition, similar in terms of density and urbanity levels so as they constitute a credible control pool. It would not make sense to consider micropolitan areas as those are potentially very different in terms of urbanity compared to our treated unit, which would make it hard to achieve a low training error. As outlined above, we also restrict the donor pool to areas that had a similar number of deaths reported in the first week since the first death occurred. The Washington DC area had 5 deaths on the 5th day since its first death, so we keep controls whose number of deaths lies within a similar window. There is obviously a trade-off in terms of setting the width of the window. If the window is too wide, we get a larger donor pool, but then the algorithm also becomes more prone to overfitting as we need to keep the number of controls small relative to the number of estimation periods available. On the other hand, making the window too small leaves us with few controls which makes it hard to achieve a low training error. Hence, we experiment with a width of 1 and 2 (i.e, the control units have a cumulative number of deaths on the fifth day between 4 and 6 (respectively, 3 and 7)).

We experiment with various types of weights and pick the best which has both a low training and validation error. We always include a weight intercept as given preliminary analysis this leads to lower errors both on training and validation sets. We try:

1. Traditional SC weights (which satisfy non-negativity and sum up to 1).
2. Non-restricted SC weights.
3. Non-restricted SC weights with L2 regularization.
4. Non-restricted SC weights with elastic net.

We also try various rules for smoothing the series to remove noise due to reporting and weekend effects which is present all throughout. We obtain our best fit (both on training and validation set) when setting a window of 2, using non-restricted SC weights with L2 penalty term equal to 0.2 and using a smoothing the series by using a moving average of order 2.We get a Training RMSE of 0.5797 and a Validation RMSE of 0.1791. Our selected controls with weights in parentheses: Controls: Cedar Rapids, IA (0.5562), Des Moines, IA (0.8297), Provo-Orem, UT (0.3588) and Salt Lake City, UT (0.4327).

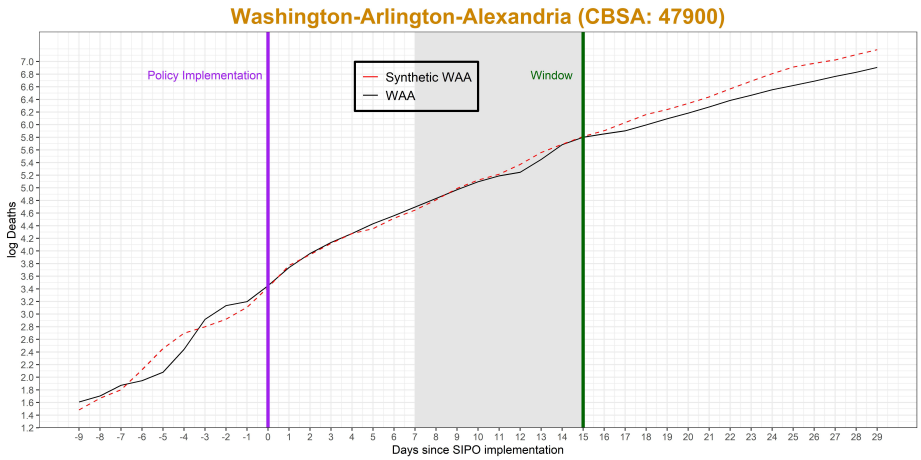

Figure 2: Days after 1st death: 5; Cluster range: 2; Smoothing Order: 2; ℓ_2-pen: 0.2; Standard SC weights: No

What happens if we add more controls? We now allow for a window of 4 so that our donor pool now contains control areas with number of cumulative deaths on the 5th day since the 1st day of reported death between 1 and 9. Now, we get many more controls to match on, but we end up basically with as many controls as estimation periods. Hence, we need sparsity. Standard synthetic control weights indeed act as a natural regularizer as long as the treated unit lies outside of the convex hull of the

control units. The synthetic unit is then the projection of the treated unit onto one of the outer sides of the $N-1$-dimensional convex hull, which implies that some of the controls will be assigned a zero weight. Applying the standard SC weights results in a trend that can be seen in Fig. 3.

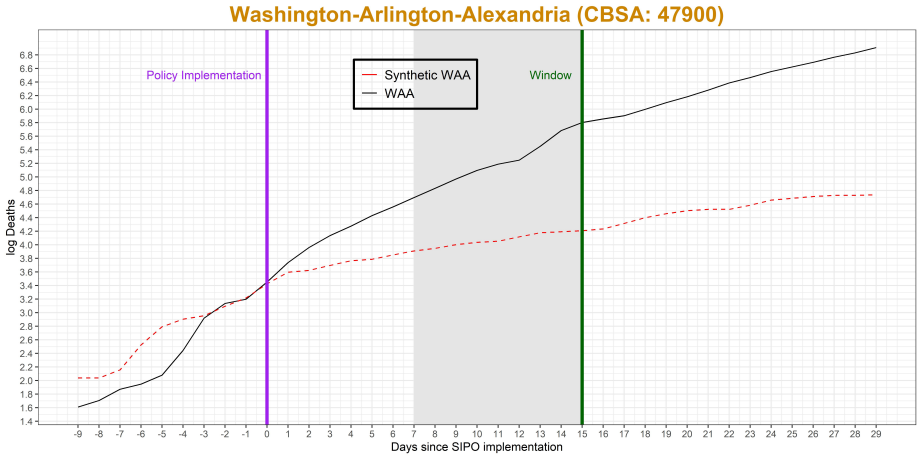

Figure 3: Days after 1st death: 5; Cluster range: 1; Smoothing order: 2; ℓ_2-pen: 0; Standard SC weights: Yes

We see that the procedure does not produce good results. This allows us to conclude that while SC weights do indeed impose sparsity (indeed we check that all the weight is concentrated on a single area!), they represent only one particular form of regularizer and there is no reason to assume that this is the best one. Obviously, the advantage of SC weights is interpretability but if the procedure itself is generating too high training and validation errors, interpretability loses its purpose as we are getting the wrong counterfactual anyway. This motivates us to look for a more general way of regularizing: elastic net. Regularization from elastic net is a convex combination of L_1 and L_2 regularization. L_1 buys us sparsity whereas L_2 reduces variance and hence buys a better out-of-sample fit. Using an elastic net with $\alpha = 0.05$ and a regularization parameter of 1.4 indeed gets us much closer to the first best although does not reach the first best. We also obtained close to first-best results with further restricting the set of controls by design, ie, by reducing the cluster window to 1. This resulted in two remaining controls, and perhaps not surprisingly, no regularization was needed. In that case, we obtain the Fig. 4.

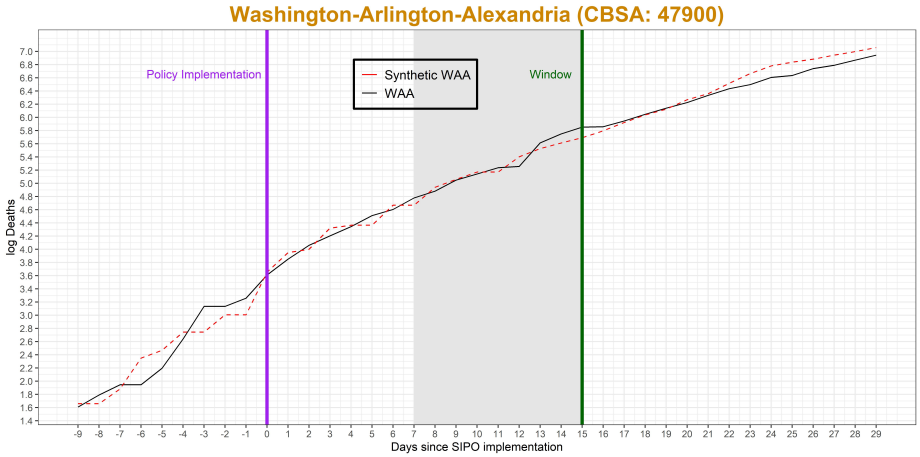

Figure 4: Days after 1st death: 5; Cluster range: 1; Smoothing Order: 0; ℓ_2-pen: 0; Standard SC weights: No

We get a Training RMSE of 0.74 and a Validation RMSE of 0.24. Our selected controls with weights in parentheses: Cedar Rapids, IA (0.99), Des Moines, IA (0.93). Additional analyses available upon request show that our epi-parameter estimation results are stable to using the above synthetic control

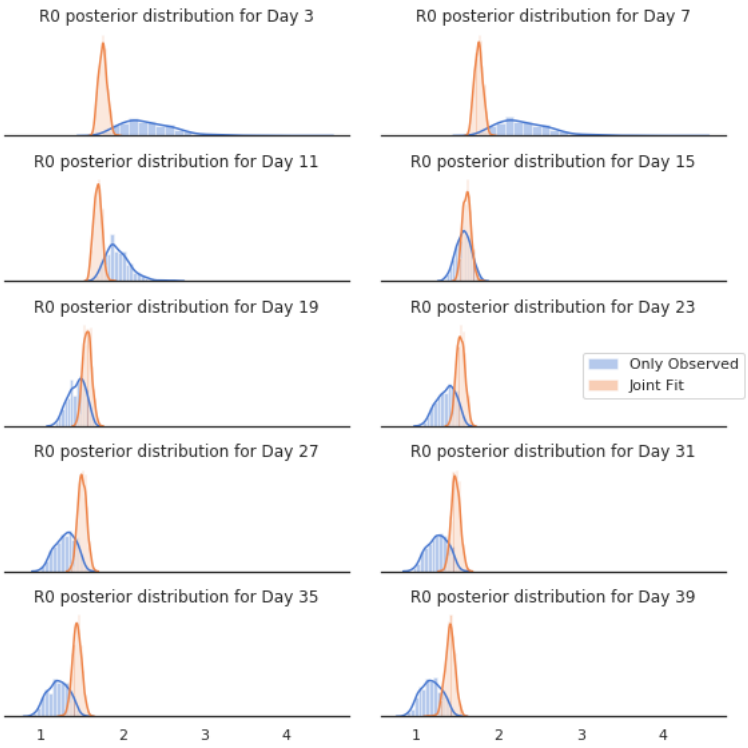

Figure 5: R_0

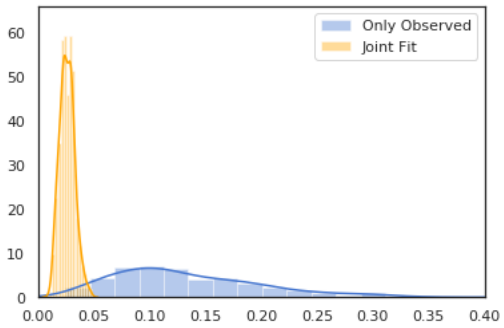

Figure 6: ρ_{isol} for the Observed Series post-intervention

curve, which is slightly worse that the first best. This shows that slight variations in the estimation of the synthetic curve should not affect our epi-model estimation results significantly.

Overall, this section suggests that with few pre-intervention data, we need to balance out a fully data-driven approach to selecting the weights with a more judgement-based approach based on selecting good controls ex-ante. We believe the approach we adopt illustrates this point well.

1.5 Full Results

Here, we present the full estimation results for the parameters of our compartmental model. Posteriors are give both when fitting the observed curve only and when fitting the model using both the observed and synthetic curves. We can see that posteriors using the joint fits are generally tighter. Part of this could be simply mechanical because of using more input and part of it could be that we are actually bringing more information to our model via the synthetic curve.

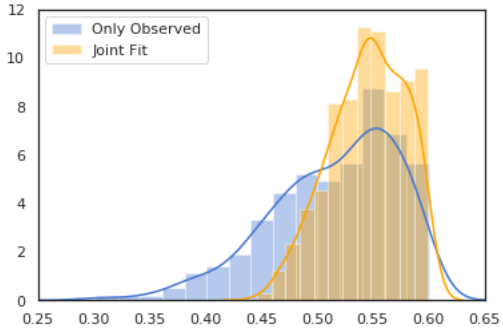

Figure 7: ρ_{isol} for the Observed Series post-intervention

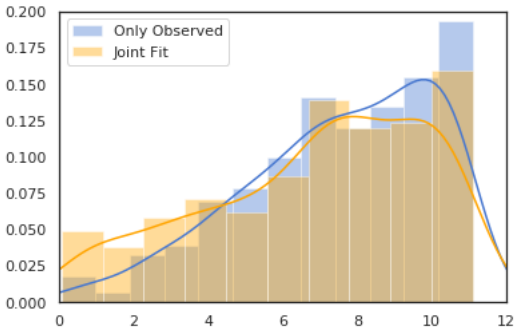

Figure 8: F_{ext}

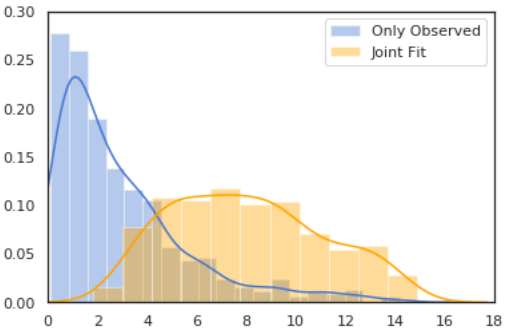

Figure 9: $F_{ext,rate}$

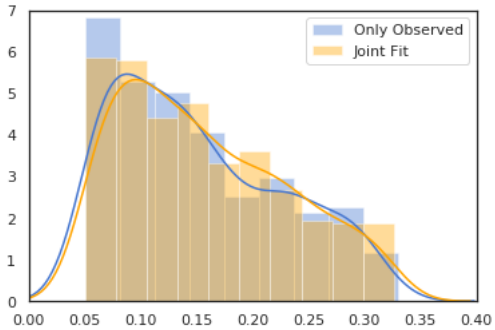

Figure 10: ρ_{fat}

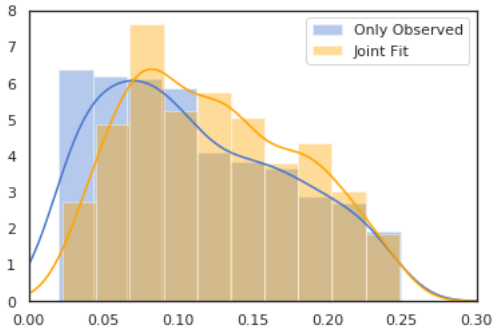

Figure 11: ρ_{hos}

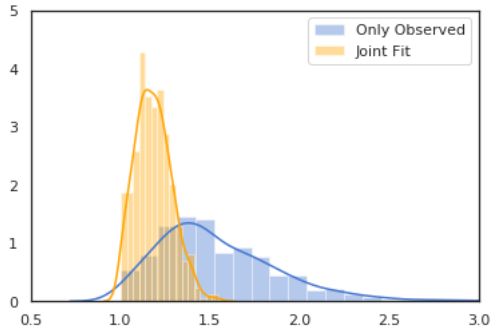

Figure 12: $\tau_{Dia,S}$

References

[1] A Abadie. Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 2019.

[2] Susan Athey and Guido W Imbens. The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32, 2017.

[3] CDC. COVID-19 Data from NCHS, 2020. https://www.cdc.gov/nchs/nvss/vsrr/covid19/index.htm.

[4] CDC. Excess Deaths Associated with COVID-19, 2020. https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm.

[5] CDC. Provisional Death Counts for Coronavirus Disease (COVID-19) - Technical Notes, 2020. https://www.cdc.gov/nchs/nvss/vsrr/covid19/tech_notes.htm.

[6] Center for Systems Science and Engineering at Johns Hopkins University. Coronavirus COVID-19 Global Cases, 2020. https://coronavirus.jhu.edu/map.html.

[7] World Health Organization China. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19), 2020. https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf.

[8] USA Facts. Coronavirus Locations: COVID-19 Map by County and State, 2020. https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/.

[9] Richard Florida. The geography of coronavirus. *Citylab*, 2020. https://www.citylab.com/equity/2020/04/coronavirus-spread-map-city-urban-density-suburbs-rural-data/609394/.

[10] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. CRC press, 2013.

[11] ArcGIS Hub. Metropolitan statistical areas for US counties - Sept 2018, 2020. https://hub.arcgis.com/datasets/47c5474c19cb4f4d9f102a84b4b8e462_2?orderBy=METDIVFP&orderByAsc=false.

[12] Iris Hui, Dante Chinni, and Ari Pinkus. The American Communities Project. https://www.americancommunities.org/.

[13] The Internet. Household Income in the United States. https://en.wikipedia.org/wiki/Household_income_in_the_United_States#:~:text=The%20U.S.%20Census%20Bureau%20reported,2016%2C%20exceeding%20any%20previous%20year.

[14] Josh Katz, Denise Lu, and Margot Sanger-Katz. What is the real coronavirus death toll in each state? *The New York Times*, 2020. https://www.nytimes.com/interactive/2020/05/05/us/coronavirus-death-toll-us.html.

[15] William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

[16] William Ogilvy Kermack and Anderson G McKendrick. Contributions to the mathematical theory of epidemics. iii.—further studies of the problem of endemicity. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 141(843):94–122, 1933.

[17] WO Kermack and AG McKendrick. Contributions to the mathematical theory of epidemics: V. analysis of experimental epidemics of mouse-typhoid; a bacterial disease conferring incomplete immunity. *Epidemiology & Infection*, 39(3):271–288, 1939.

[18] Chirag Modi, Vanessa Boehm, Simone Ferraro, George Stein, and Uros Seljak. Total covid-19 mortality in italy: Excess mortality and age dependence through time-series analysis. *medRxiv*, 2020.

[19] NYT. Coronavirus (covid-19) data in the united states. https://github.com/nytimes/covid-19-data.

[20] Government of the District of Columbia. Covid-19 surveillance, 2020. https://coronavirus.dc.gov/data.

[21] USA Today. US coronavirus map: Tracking the outbreak, 2020. https://www.usatoday.com/in-depth/graphics/2020/03/10/us-coronavirus-map-tracking-united-states-outbreak/4945223002/.

[22] Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael PH Stumpf. Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6(31):187–202, 2009.

[23] Bryan Wilder, Marie Charpignon, Jackson A Killian, Han-Ching Ou, Aditya Mate, Shahin Jabbari, Andrew Perrault, Angel N Desai, Milind Tambe, and Maimuna S Majumder. Modeling between-population variation in covid-19 dynamics in hubei, lombardy, and new york city. *Proceedings of the National Academy of Sciences*, 2020.

[24] Fei Zhou, Ting Yu, Ronghui Du, Guohui Fan, Ying Liu, Zhibo Liu, Jie Xiang, Yeming Wang, Bin Song, Xiaoying Gu, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, 2020.