1   Introduction

1.1   Related Works

Other modeling approaches. SIR-F, DCM, Agent, Hybrid - not post fact A/B testing tools.

1.2   Tools

As stated above, our objective is to study the effects of government response at an aggregate level in terms of life saved, and limiting the number of cases that requires hospitalization. Such interventions can effectively be studied at a comparative level. In other words, if we have data for the evolution of aggregate outcomes, e.g. number of confirmed cases and deaths, when policy is applied in a group under study versus when the same policy is not applied in a control group. However, government policies were applied at different level across a geographic region. We do not have a mechanism to conduct a randomized trial. Hence, we consider constructing synthetic control methodAbadie et al. (2010); Amjad et al. (2018, 2019). In a synthetic control set up, where observational data is available for different groups, we can construct a synthetic or virtual control group by combining measurements from alternatives (or donors). In following we provide a brief overview of Multi-dimensional Robust Synthetic Control followingAmjad et al. (2019).

Suppose that observations from N different geographically distinct groups or units are indexed by $i \in [N]$ in T time periods (days) indexed by $j \in [T]$. Let $k \in [K]$ be the metrics of interest (e.g. number of confirmed cases, number of deaths, number of tests conducted, etc.). By M_{ijk} we denote the ground-truth measurement of interest, and by X_{ijk}, an observation of this measurement with some noise. Let $1 \leq T_0 \leq T$ be the time instance in which our group of interest experiences an intervention, namely a government response to control the spread (e.g. stay home order, school or business closure, or mass vaccination). Without loss of generality we consider unit $i = 1$ (say, New York) and metric $k = 1$ (say, number of deaths) as our unit and metric of interest respectively.

Our objective now is to estimate the trajectory of metric of interest $k = 1$ for unit $i = 1$ if no government response to control the spread had occurred. In order to do that we will use the trajectory associated with the donor units $(2 \leq i \leq N)$, and metrics $k \in [K]$. In following we make two assumptions:

(1) for all $2 \leq i \leq N$, $k \in [K]$ and $j \in [T]$, we have $X_{ijk} = M_{ijk} + \epsilon_{ijk}$ where, ϵ_{ijk} is the observational noise, and (2) Same model is obeyed by $i = 1$ in pre–intervention period, i.e. for all $j \in [T_0]$ and $k \in [K]$ we have $X_{1jk} = M_{1jk} + \epsilon_{1jk}$. As described by authors in Amjad et al. (2019), in following we also assume that for unit $i = 1$, we only observe the measurement X_{1jk} for pre-intervention period, i.e. for all $j \in [T_0]$ and $k \in [K]$. Our objective is to compute a counterfactual sequence of observation M_{1jk} for the time period $j \in [T]$, and $k \in [K]$, and in specific for $T_0 \leq j \leq T$, and $k = 1$, using synthetic version of unit $i = 1$.

Define $\mathcal{M} = [M_{ijk}] \in \mathbb{R}^{N \times T \times K}$. $\mathcal{M}$ is assumed to have a few well behaved properties as required by the algorithm, namely, it must be approximately low-rank and boundedness of $|M_{ijk}|$ (for details see Amjad et al. (2019)). To check whether our model assumption holds in practice, we consider $N = 185, T = 150, K = 2$, with 185 countries as units. We consider number of confirmed cases and number of deceased as two metrics over 150 days between January 22, 2020 and June 20, 2020. For assumption to hold, data matrix corresponding to number of confirmed cases and number of deceased and their combination should be approximated by a low-rank matrix.
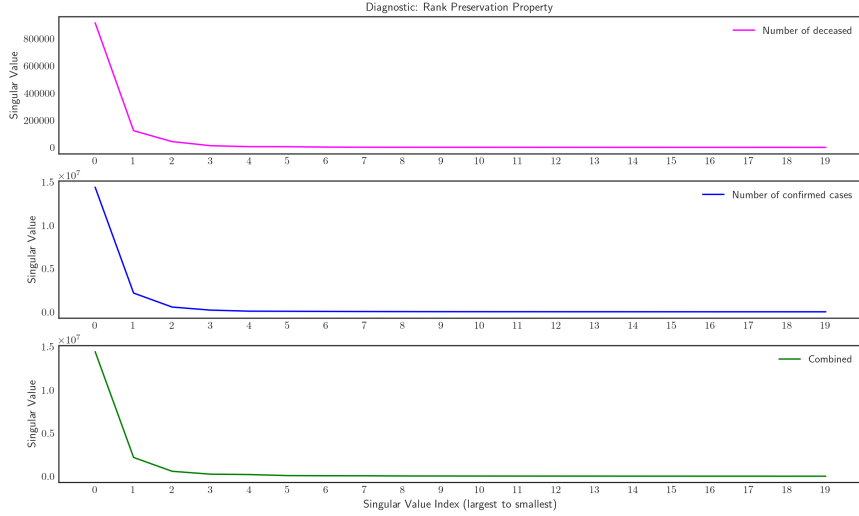

Figure 1: The Singular Value spectrum for all countries (of dimensions 185 × 150, showing top 20 singular values, in descending order.

As shown in Figure 1, the spectrum of the top 20 singular values (sorted in descending order) for each matrix. The plots clearly support the implications that most of the spectrum is concentrated within the top 5 principal components. Same conclusion holds true when units are states of United States, and when we consider only countries in European Union.

Let $\mathcal{Z} \in \mathbb{R}^{(N-1) \times T \times K}$ corresponding to donor units, and $X_1 \in \mathbb{R}^{1 \times T_0 \times K}$ correspond to unit under intervention. We obtain $\hat{\mathcal{M}}$ from $\mathcal{Z}$ after applying a hard singular value thresholding. Subsequently, weights are learned using linear regression by computing

$$\hat{\beta} = \underset{v \in \mathbb{R}^{(N-1)}}{\arg \min} \left\| X_1 - v^T \hat{\mathcal{M}}_{T_0} \right\|_2^2$$

For every $k \in [K]$, the corresponding estimated counterfactual means for the treatment unit is then defined as

$$\hat{\mathcal{M}}_1^{(k)} = \hat{\beta}^T \hat{\mathcal{M}}^{(k)}$$

2 Methods

2.1 Overview

As described in Section 1, we use Multi-dimensional Robust Synthetic Control to construct a synthetic control for the treatment unit using data from multiple control units or donor group using pre-intervention period data. The synthetic control is then used for estimating the counterfactual in the post-intervention period. In our setup intervention date is typically the date when a stay-home order or lock-down was declared for the treatment unit. However, government policy may have been applied over time with different level of stringency measures.

To understand this we use stringency and policy indices data from Ox-CGRT Hale et al. (2020), which records the strictness of policies that restrict people's behavior and includes 8 different measures - e.g. school and workplace closure, cancellation of public events, restrictions on gathering size etc.

3 Results

4 Discussion

5 Concluding Remarks

References

A. Abadie, A. Diamond, and J. 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. doi: 10.1198/jasa.2009.ap08746.

M. Amjad, D. Shah, and D. Shen. Robust synthetic control. *Journal of Machine Learning Research*, 19(22):1–51, 2018. URL http://jmlr.org/papers/v19/17-777.html.

M. Amjad, V. Misra, D. Shah, and D. Shen. Mrsc: Multi-dimensional robust synthetic control. *Proc. ACM Meas. Anal. Comput. Syst.*, 3(2), 6 2019. doi: 10.1145/3341617.3326152. URL https://doi.org/10.1145/3341617.3326152.

T. Hale, S. Webster, A. Petherick, T. Phillips, and B. Kira. Oxford covid-19 government response tracker. 2020.