Causal Inference of Socioeconomic Environments and Health Interventions for Influencing the Spread of COVID-19 in the US 2020-2021

XIAOHAN LIU and SON NGUYEN*

Abstract:

The Covid-19 outbreak has dramatically changed people lives. In this work, we perform causal discovery on both scalar and temporal variables and estimate the causal effects of removing and reissuing face mask policies on time series of Covid-19 cases. Among the demographic risk factors, we find both the population density and median income have strong causation on the new cases for the entire first year of COVID-19. When it comes to temporal risk factors, we implement the causal discovery on both the federal level and state level. Our tests demonstrate that the new cases per 100,000 dominantly hinges on the active cases per 1000 people for the entire nation and 37 states, whereas the leading temporal risk factor ranges from the percentage of work-from-home to imported cases for the rest states. We also discover that removing face mask policies do not affect the new cases per 100,000 population. However, reissuing mandates do lower the number of new cases per 100,000 population in Illinois and Louisiana.

1 INTRODUCTION

With the lack of effective drugs other than vaccines, we rely heavily on non-pharmaceutical interventions to contain the spread of COVID19. The successful discovery of the causal link between some common risk factors for a pandemic and the number of newly infected cases, or a reliable estimate of the impact of the non-pharmaceutical intervention, could provide valuable information for policymakers on both the federal and state levels during the pandemic. We will implement several causal inference practices that based on additive noise models, Granger causality and Transfer entropy for both the static and temporal risk factors to examine their involvement in the spread patterns of COVID-19 in terms of causal discovery. We will use Google's CausalImpact approach which leverage other time series as control groups to estimate the counterfactual time series of interest.

2 PROBLEMS DESCRIPTIONS AND SOLUTIONS

2.1 Problems descriptions

The primary goal of this research project is to perform COVID-19 related casual inference tasks, both in terms of causal discovery and treatment effect estimation, from both scalar and time-series data. The specific questions we are to address along the course are as follows:

- For all 50 states, identify the impacting static risk factors among 6 candidates that can cause the growth of new cases for every 3 months of the first year of COVID-19.
- For each state, identify the top temporal risk factors among mobility and social distancing, COVID-19 and other health records and economic impacting ones that have causation on the new cases per 100,000 population for the first year of COVID-19.
- Estimate the causal effect of the policy decision to lift the state-wide face mask mandate of the state of Illinois on a time series of newly diagnosed cases per 100,000 population.
- Estimate the causal effect of the policy decision to reissue state-wide face mask mandate of the state of Illinois on a time series of newly diagnosed cases per 100,000 population.

Authors' address: Xiaohan Liu, xliu98@uic.edu; Son Nguyen, snguye65@uic.edu.

2.2 Solutions

In order to identify the influential static demographic risk factors, we utilize the causal discovery technique based on additive noise models, whereas Granger causality tests and Transfer entropy are performed to identify the impacting temporal risk factors towards the new COVID cases. For estimating the causal effect of the policy decisions, we will use Google's CausalImpact approach which leverage other time series as control groups to estimate the counterfactual time series of interest. The code of our implementation can be found at https://github.com/Dr-Spicy/Causal-Inference-on-Covid-19-USA.

3 RELATED WORK

Our approach to discover the causal scalar risk factors on the new cases with continuous additive noise models (ANMs) [Peters et al. 2014] is inspired by this work [Li et al. 2021] but with some modifications and implemented in our codes. From the literature review, we also find a Toolbox on ANMs [Kalainathan et al. 2020] that we are in the processing of validating our current results from another perspective.

When it comes to causal discovery between one temporal response and another temporal causal factor, we need to recourse to the good old Granger causality[Granger 1969]. There are a few well-known approaches, including the multivariate linear Granger causality test based on the vector error correction models with mild assumptions [Sims 1980; Wen et al. 2013] and nonlinear Granger causality test based on kernel density estimation and data-sharpening functions [Diks and Wolski 2016; Hmamouche 2020] despite that the Granger causality is originally defined under a linear framework. Meanwhile, since the introduction of Transfer entropy [Schreiber 2000], it has been recognized as an important tool in the analysis of causal relationships between time series data in nonlinear systems [Hlaváčková-Schindler et al. 2007].

As of estimating the causal impact of the non-time series intervention, Brodersen et al. [Brodersen et al. 2015] at Google applied Bayesian Structural Time Series [Scott and Varian 2014] model for doing the task. They use this approach to infer Google's advertisement campaign on the number of times a user was directed to the advertiser's website. Because our tasks of estimating causal impacts of with or without face mask policy on the number of Covid-19 new cases are similar to their problem, we will use the python version

of Brodersen at al. [Brodersen et al. 2015] CausalImpact R-package to do the task.

If the intervention is also time series, we attempt to quantify the causal effect by an epidemic-inspired marginal structural model [Bonvini et al. 2021]. A caveat on this model is that it imposes extreme assumptions, and its highly sophisticated implementation is hastily narrated without many details.

4 DATA DESCRIPTION AND CLEANING

The data are gathered from the University of Maryland's COVID-19 Team, John Hopkins Coronavirus Resource Center, mobility data from Google that records newly confirmed COVID-19 cases from March 22, 2020, to August 30, 2021, 16 scalar (non-temporal) and 15 temporal socioeconomic environmental and health intervention risks factors every day from March 22, 2020, to March 20, 2021, as the first year of this pandemic in the US. The daily data will then be converted to weekly to mitigate various reporting issues for more consistent results. All the scalar variables have been standardized before being used in regressions, and all the temporal ones have been converted to stationary cointegrated $I(1)$ variables by first differences. Lastly, we extract effective dates of the policies for all states from ballotpedia.org.

5 EXPERIMENT RESULTS

5.1 CAUSAL DISCOVERY OF SCALAR RISK FACTORS

5.1.1 Setup. We are using the additive noise models (ANMs) to test of causation between the response Y (average number of new cases over one of the four quarters) and six potential scalar causal variables X for all 50 states. Each state is assumed to be a sample that satisfies exchangeability. Let us further assume the effects of confounding, selection bias, and feedback are also negligible. Consider the relationship between X and Y can be correctly modeled by a nonlinear mapping function

$$Y = f_Y(X) + E_Y, X \perp\!\!\!\perp E_Y, \tag{1}$$

where E_Y is an independent additive noise. We then regard the joint density $P_{X,Y}$ to be induced by the ANM from X to Y $X \to Y$. In some other cases, we may have the alternative direction of ANM $Y \to X$, when the relationship between X and Y should be modeled by

$$X = f_X(Y) + E_X, Y \perp\!\!\!\perp E_X, \tag{2}$$

where E_X is an independent additive noise. If the joint density $P_{X,Y}$ is induced by the ANM $X \to Y$ but not by the ANM $Y \to X$, we call the ANM $X \to Y$ identifiable.

ANMs highly depend on the independent relation between the causal variable and the noises. We pick the Hilbert-Schmidt norm of the cross-variance operation's finite-dimensional approximation, the Hilbert-Schmidt independence criterion (HSIC), to evaluate the level of independence between the residuals and causal variables for its mathematical elegance and easy interpretation. To be specific, it is defined as $HSIC(X,Y)^2 = ||C_{XY}||^2_{HS}$, where $||\cdot||_{HS}$ is the Hilbert-Schmidt norm and the covariance operator C_{XY} is a generalization of the finite-dimensional covariance matrix to infinite-dimensional feature space. It has been proven that $HSIC(X,Y) = 0$ if and only if X and Y are statistically independent [Gretton et al. 2007].

For the sake of model flexibility, we evenly divide the entire year into 4 quarters as shown in Figure 1 and study them separately. The red indicates the first quarter (March 22 to June 20, 2020), yellow is for the second quarter (June 21 to September 19, 2020), green for the third quarter (September 20 to December 19, 2020), and blue for the fourth quarter (December 20, 2020, to March 20, 2021). In this way, we only need to suppose the models are stationary within each quarter instead of the entire year, and we gain the benefit of observing whether the causation will change as time lapses for some factors.

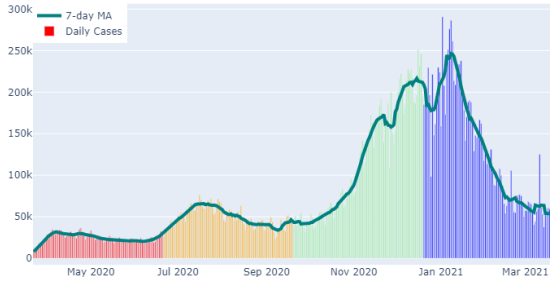

Fig. 1. Four quarters partition of the first year of COVID-19 in US.

We now can summarize the detailed procedure for testing causality between the average number of new cases Y and each one of those potential X's in every time period as follows:

Step 1: Randomly partition all 50 states into a training set of size n $\{(Y_1, X_1), \ldots, (Y_n, X_n)\}$ and a testing set of size m $\{(\tilde{Y}_1, \tilde{X}_1), \ldots, (\tilde{Y}_m, \tilde{X}_m)\}$. Due to the lack of high homogeneity across the states and the small sample size, we recommend to use $n = 45$ and $m = 5$ in practice.

Step 2: Applying the smoothing spline regression method (local polynomial regression method generates similar results) over the training set to estimate the non-linear mapping functions mentioned in (1) and (2) as $\widehat{f_Y}$ and $\widehat{f_X}$. The hyperparameters like the knots, the smoothing bandwidths, the penalty coefficients *etc.* are chosen by Cross-Validation. Other nonparametric regression methods such as neural networks can be used here as well.

Step 3: Obtain the predicted residuals over the testing set by

$$\widehat{E}_{\tilde{Y}} = (\widehat{E}_{\tilde{Y}_1}, \ldots, \widehat{E}_{\tilde{Y}_m})^T \text{ , where } \widehat{E}_{\tilde{Y}_j} = \tilde{Y}_j - \widehat{f_Y}(\tilde{X}_j).$$

$$\widehat{E}_{\tilde{X}} = (\widehat{E}_{\tilde{X}_1}, \ldots, \widehat{E}_{\tilde{X}_m})^T \text{ , where } \widehat{E}_{\tilde{X}_j} = \tilde{X}_j - \widehat{f_X}(\tilde{Y}_j).$$

Step 4: Calculate $HSIC^2(\tilde{X}, \widehat{E}_{\tilde{Y}})$ as

$$HSIC^2(\tilde{X}, \widehat{E}_{\tilde{Y}}) = \frac{1}{m^2} tr(K_{\tilde{X}} H K_{\widehat{E}_{\tilde{Y}}})$$

$$K_{\tilde{X}} = \begin{bmatrix} k(\tilde{X}_1, \tilde{X}_1) & \cdots & k(\tilde{X}_1, \tilde{X}_m) \\ \vdots & \ddots & \vdots \\ k(\tilde{X}_m, \tilde{X}_1) & \cdots & k(\tilde{X}_m, \tilde{X}_m) \end{bmatrix}, K_{\widehat{E}_{\tilde{Y}}} = \begin{bmatrix} k(\widehat{E}_{\tilde{Y}_1}, \widehat{E}_{\tilde{Y}_1}) & \cdots & k(\widehat{E}_{\tilde{Y}_1}, \widehat{E}_{\tilde{Y}_m}) \\ \vdots & \ddots & \vdots \\ k(\widehat{E}_{\tilde{Y}_m}, \widehat{E}_{\tilde{Y}_1}) & \cdots & k(\widehat{E}_{\tilde{Y}_m}, \widehat{E}_{\tilde{Y}_m}) \end{bmatrix},$$

where $H = I - \frac{1_n 1_n^T}{n}$ and $k(\cdot,\cdot)$ is the Gaussian kernel in our implementation. $HSIC^2(\tilde{Y}, \widehat{E}_{\tilde{X}})$ is calculated similarly.

$$H_0 : X \to Y \text{ or } Y \to X$$

Step 5: Infer causal direction:

$$X \to Y \text{ if } HSIC^2(\tilde{X}, \widehat{E}_{\tilde{Y}}) < HSIC^2(\tilde{Y}, \widehat{E}_{\tilde{X}})$$

$$Y \to X \text{ if } HSIC^2(\tilde{X}, \widehat{E}_{\tilde{Y}}) > HSIC^2(\tilde{Y}, \widehat{E}_{\tilde{X}})$$

Here, we set the null hypothesis H_0 : no $X \to Y$ nor $Y \to X$. But there is no analytical forms for the asymptotic null distribution of the $HSIC$ and we use permutation/bootstrap to calculate the p-values of the causal test statistics defined as:

$$T_c = |HSIC^2(\tilde{X}, \widehat{E}_{\tilde{Y}}) - HSIC^2(\tilde{Y}, \widehat{E}_{\tilde{X}})|$$

over $B = 100$ permutations where the correspondence between the testing set $\tilde{Y}$ and $\tilde{X}$ are randomly assigned to simulate an empirical distribution. The p-values are defined as the proportion of the statistic $\tilde{T}_c$ larger or equal to the one computed on the original testing set.

Step 6: Repeat step 1-5 to produce a p-value 100 times, report the average after removing outliers for a more robust conclusion.

5.1.2 Results. For each of the four time intervals/quarters, the response is the average new cases of COVID-19. The six scalar risk factors considered for causality are the population density, percent of males, median income, percent of seniors, percent of African Americans and percent of Hispanic Americans. Since the sample size is too small, we choose a traditional significance level of 0.05 without Bonferrroni correction for multiple tests. The p-values are summarized in Table 1. We observe that population density is indubitably the most important risk facts for all four quarters as it is supposed to be by any epidemic theory. Besides, the median income is verified to have a considerable impact most likely it portraits the socioeconomic status distribution of the target state's population, which could affect an individual's attitude and capability against the pandemic. Moreover, the percent of males and the percent of seniors are not causal at all. Interesting things happen on the percent of African Americans and Hispanic Americans, which are non-causal in the first two quarters, then become causal or possibly causal in the third quarter, and non-causal in the last quarter. We hypothesis that this phenomenon could be a result of either the minority groups are more vulnerable during the outbreak of the second wave or the frequencies of social movements and protests like Black Lives Matter coincides with the ethnic profiles and it is a common belief that those gatherings will increase the spread of virus. The summarized causal graphs between the average new cases and all six static demographic risk factors can be found in Figure 2.

Table 1. p-values for testing 6 scalar potential causes for the number of new cases. *(Significance codes:'**' 0.01, '*' 0.05, '.' 0.1.)*

Risk factor	p-value			
	quarter 1	quarter 2	quarter 3	quarter 4
Population Density	0.0072**	0.0033**	0.0029**	0.0019**
Percent of Males	0.4125	0.4178	0.2431	0.4229
Median Income	0.0296*	0.0274*	0.0164*	0.0195*
Percent of Seniors	0.3301	0.3607	0.2343	0.3401
Percent of African Americans	0.2835	0.2796	0.0578 .	0.2392
Percent of Hispanic Americans	0.2717	0.3051	0.0449*	0.2399

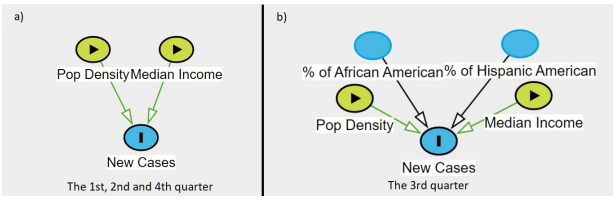

Fig. 2. The causal graphs between the average new cases per quarter and all static demographic risk factors. a), for the first, second and fourth quarter; b), for the third quarter.

5.2 Causal discovery of temporal risk factors

5.2.1 Setup. There are two popular concepts that have been widely used to infer causal relationships between time series, Granger causality [Granger 1969] and Transfer entropy [Schreiber 2000]. The core ideas behind temporal causality is firstly the cause must occur before the effect, and secondly a time series X is said to cause Y if it contains unique information to predict the current value of Y but not included in the history of Y and all other information in the universe. Besides, Transfer entropy measures the information flow from one time series X to another Y. And it detects directional and dynamical information flows while not assuming any particular parametric form to describe interactions within the system [Faes et al. 2014].

Firstly, we are using the classical Granger causality test [Granger 1980] to verify the existence of Granger causality between two time series. This has been materialized by considering two VAR(p) (Vector Auto-Regressive) models to predict Y_t, one with its own history and the other with the history of X. Finally, the difference between these two situations are evaluated to check if the additional X information has some effects on the prediction.

$$\text{Model}_1 \quad Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-1} + U_t,$$

$$\text{Model}_2 \quad Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-1} + \sum_{i=1}^{p} \beta_i X_{t-1} + U_t,$$

where p is the lag parameter and U is the white noise error term. To quantify the causality, we need to evaluate the statistical difference between the variances of the errors of Model_1 and Model_2 as σ_1^2 and σ_2^2 with the F-test, where the statistic is defined as:

$$F = \frac{(RSS_1 - RSS_2)/p}{RSS_2/(n - 2p - 1)} \sim F(p, n - 2p - 1),$$

where RSS_1 and RSS_2 are the residual sum of squares related to Model_1 and Model_2, and n is the size of the lagged variables. Two hypotheses will be considered:

- H_0 : $\forall i \in \{1, \ldots, p\}, \beta_i = 0$,
- H_1 : $\exists i \in \{1, \ldots, p\}, \beta_i = 0$.

H_0 here represents that time series X does not Granger causes Y.

In practice, we implemented the Granger test with the help of aforementioned R package "NlinTS". The bi-variate Granger causality will be measured both ways between all 15 temporal risk factors and the new cases per 100K people as the temporal response. We set

the lag parameter $p = 2$. This is due to fact that the most contracted people have observable symptoms range from 2 to 12 days after infection, which is almost two weeks.

On the other hand, Transfer entropy is rooted in information theory and based on the classical concept of Shannon entropy as a measure of uncertainty [Shannon 1948]. In order to measure the information flow between two time series X and Y, we couple Shannon entropy with the Kullback-Leibler (KL) distance under the assumption that the underlying stochastic processes are both Markovian [Schreiber 2000]. In the bi-variate case, the information flow is measured by quantifying the deviation from the generalized Markov property under the KL distance metric. The Transfer entropy from X to Y based on Shannon entropy is given by:

$$\begin{aligned} TE_{X\to Y}^{(p,q)} &= H(Y_t|\{Y_{t-1},\dots,Y_{t-q}\}) \\ &\quad - H(Y_t|\{(Y_{t-1},X_{t-1}),\dots,(Y_{t-q},X_{t-p})\}) \\ &= \sum_{Y_t,Y_t^q,X_t^p} Prob(Y_t,Y_t^q,X_t^p) log(\frac{Prob(Y_t|Y_t^q,X_t^p)}{Prob(Y_t|Y_t^q)}), \end{aligned}$$

where H stands for the Shannon entropy, p and q are the lag parameters for X and Y *resp.* The measure for the information flow from Y to X as $TE_{Y\to X}^{(q,p)}$ can be derived analogously. However, Transfer entropy's estimates are biased when the sample size is small, and our sample size $n = 52$ is on the small side. We would construct a bias correction as the effective Transfer entropy as

$$Eff.TE_{X\to Y}^{(p,q)} = TE_{X\to Y}^{(p,q)} - TE_{X_{shuffled}\to Y}^{(p,q)},$$

where $TE_{X_{shuffled}\to Y}^{(p,q)}$ indicates Transfer entropy using a shuffled version of the time series X. Shuffle here implies randomly drawling values from X and realigning them to generate a new time series, which destroys the time dependence of X as well as the statistical dependencies between X and Y. In order to achieve consistent estimators, shuffling is repeated and the average of the resulting shuffled Transfer entropy will serve as a bias corrected estimator on Transfer entropy for our small sample size.

Note that Transfer entropy is an asymmetric measure, thus it allows the estimation of the directional coupling between time series as the net information flow defined as:

$$\widehat{N.I.F.} = Eff.TE_{X\to Y}^{(p,q)} - Eff.TE_{Y\to X}^{(q,p)}.$$

One can interpret this estimated quantity as a measure of the dominant direction of the information flow. That is, a positive result indicates a dominant information flow from X to Y compared to the other direction or, similarly, it indicates which model provides more predictive information about the other model if we are interpreting it with the two models introduced in Granger causality [Nichols et al. 2013]. To access the statistical significance of Transfer entropy estimates, we rely on a Markov block bootstrap. In contrast to shuffling, it preserves the time dependencies within X and Y, but eliminates the statistical dependencies between them. Then, repeated estimation of Transfer entropy provides the distribution of the estimator under the null hypothesis,

- H_0 : There is no net information flow.,
- H_1 : There is net information flow..

The associated p-value is given by $1 - \hat{q}_\tau$, where $\hat{q}_\tau$ denotes the quantile of the simulated distribution that is determined by the respective Transfer entropy estimates. Finally, if one null hypothesis is rejected, the direction of net information flow will be decided by the sign of $\widehat{N.I.F.}$.

The estimations and tests associated with Transfer entropy are materialized with R package "RTransferEntropy" similarly to our Granger causality test both ways between all 15 temporal risk factors and the new cases per 100K people as the temporal response. When it comes to the lag parameters, they are not necessarily the same as required by the Granger causality test. But, to keep everything consistent and assume there is no mutual information, we enforce $p = q = 2$ for our analysis.

5.2.2 Results. We can divide the 15 temporal risk factors into three categories: A, mobility and social distancing including working trip per person, non-work trip per person, miles travelled per person, social distancing index, unemployment rate and percentage of working-from-home; B, COVID record and health including active cases per 1000 people, test done per 1000 people and imported cases; C. economic impacts including change of retail consumption, change of grocery and pharmacy consumption, change of parks and recreational consumption, change of transit stations related consumption, change of workplace related consumption and change of residential related consumption. Since the sample size is still considered as small, we use a significance level of 0.05 without Bonferrroni correction.

Table 2. p-values testing 15 temporal risk factors Granger causes the number of new cases per 100K people. *(Significance codes:'**' 0.01, '*' 0.05, '.' 0.1.)*

Risk factor	p-value			
	USA	**CA**	**IL**	**TX**
Work trip/ppl	0.0810 .	0.0686 .	0.7461	0.2427
Non-work trip/ppl	0.9828	0.6451	0.6491	0.6265
Miles/ppl	0.7389	0.9948	0.7547	0.1125
Social distancing index	0.8876	0.7159	0.4195	0.6205
Unemployment rate	0.5491	0.6372	0.3283	0.5034
% of WFH	0.0147**	0.0135**	0.2601	0.1738
Active cases/1000 ppl	0.0000**	0.0039**	0.0000**	0.0135*
Test done/1000 ppl	0.8376	0.3514	0.7937	0.5390
Imported cases	0.0000**	0.1216	0.0001**	0.0176*
Retail	0.6501	0.8098	0.5465	0.5485
Grocery	0.1210	0.3784	0.1181	0.1378
Parks	0.5668	0.8806	0.7595	0.8938
Transit	0.4108	0.8846	0.4537	0.5345
Workplaces	0.4422	0.2329	0.6695	0.0168*
Residential	0.5729	0.8137	0.1313	0.3016

Although we have establish tests for all 50 states and the country as a whole, we only demonstrate a few of them due the limitation of space. But they are enough to show the causal structure can vary drastically from state to state. Table 2 summarized the p-values testing 15 temporal risk factors Granger cause the number of new case per 100,000 people as the response for the entire country (USA) along with three exemplary states like California (CA), Illinois (IL)

Table 3. p-values testing the number of new cases per 100K people Granger causes the 15 temporal risk factors. *(Significance codes:'**' 0.01, '*' 0.05, '.' 0.1.)*

Risk factor	p-value			
	USA	**CA**	**IL**	**TX**
Work trip/ppl	0.0252*	0.0050**	0.4075	0.1330
Non-work trip/ppl	0.9888	0.5402	0.1032	0.9679
Miles/ppl	0.6824	0.6638	0.8026	0.9200
Social distancing index	0.8946	0.6682	0.0265*	0.7822
Unemployment rate	0.5295	0.6634	0.7875	0.4556
% of WFH	0.0266*	0.0019**	0.2076	0.1682
Active cases/1000 ppl	0.0418*	0.0000**	0.0000**	0.0000**
Test done/1000 ppl	0.0063**	0.0001**	0.0120*	0.0016**
Imported cases	0.0000**	0.4315	0.0000**	0.0118*
Retail	0.0414*	0.0876 .	0.0067**	0.0576 .
Grocery	0.0975 .	0.0249*	0.6160	0.0056**
Parks	0.0396*	0.3274	0.0686 .	0.0041**
Transit	0.0080**	0.0388*	0.0456 .	0.2335
Workplaces	0.3651	0.0367*	0.5365	0.9744
Residential	0.1042	0.0783 .	0.0395*	0.4426

and Texas (TX). On the other hand, Table 3 summarized the p-values testing the number of new case per 100,000 people as the response Granger cause 15 temporal risk factors for USA, CA, IL and TX. These causal discoveries can also be verified with Transfer entropy analysis, so we don't show their tests' p-values to trim down redundancy. On the national level, one can see the Granger causality structure in Figure 3. We can see the national new COVID cases per 100K people has positive mutual Granger causality with the imported cases and active case per 1000 people and negative mutual Granger causality with the percentage of working-from-home. We should note that Granger causality is not necessarily the classical cause-and-effect but in terms of prediction as discussed earlier. A positive mutual Granger causality means if one time series goes up the other one will also tend to increase and *vice. versa*. In addition, there are five other one-directional Granger causalities. An increase in the national new COVID cases per 100K people is likely to induce drops in the work trips per person, retail consumption, parks and recreational consumption and transit stations related consumption. At the same time, increment in the national new COVID cases per 100K people will probably stimulates more people to get COVID tests per 1000 people.

When it comes to specific state, we observe diffident Granger causal structures versus the national cases. We will use California's Granger casual diagram shown in Figure 4 as an example. At first, the positive mutual Granger causality still exist between the Californian new COVID cases per 100K people and the active case per 1000 people, whereas the negative mutual Granger causality exist between the new COVID cases per 100K people and the percentage of work-from-home. Moreover, an increase in the Californian new COVID cases per 100K people is likely to induce drops in the work trips per person, grocery and pharmacy consumption, workplace related consumption and transit stations related consumption. Meanwhile, the increment in the Californian new COVID cases per

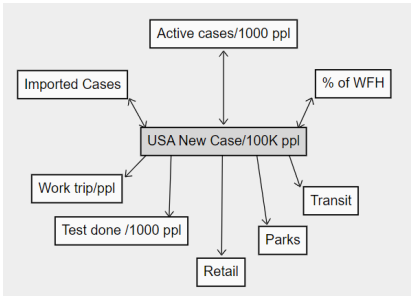

Fig. 3. The Granger causal graphs between all possible temporal risk factors Granger cause the number of new case per 100,000 people as the response for USA.

100K people will probably stimulates more people to get COVID tests per 1000 people. Similar yet different Granger causal stories can be told for Illinois (Figure 5) and Texas (Figure 6). Eventually, we summarize the leading Granger causal temporal risk factor to the new cases per 100K people of each state in Table 4. The non-work trip per person is the leading risk factor for Missouri's new case, miles travelled per person for Kansas, unemployment rate for Iowa, imported cases for New Jersey, North Carolina, West Virginia, Wisconsin and Wyoming. For the rest states and the entire country, the active cases per 1000 people remains the most influential causal risk factor. We believe the discrepancies truly reflect the intrinsic natures of the COVID spread pattern in each state and nation-wide and can provide valuable information to the policy makers of both the federal and state levels.

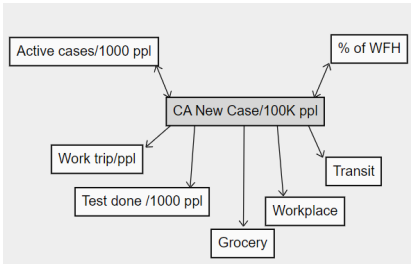

Fig. 4. The Granger causal graphs between all possible temporal risk factors Granger cause the number of new case per 100,000 people as the response for California.

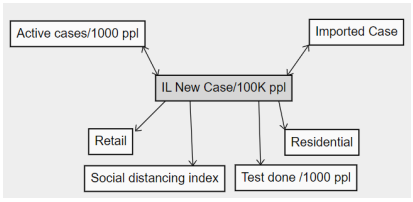

Fig. 5. The Granger causal graphs between all possible temporal risk factors Granger cause the number of new case per 100,000 people as the response for Illinois.

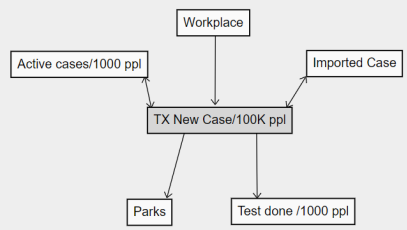

Fig. 6. The Granger causal graphs between all possible temporal risk factors Granger cause the number of new case per 100,000 people as the response for Texas.

Table 4. Leading Granger causal temporal risk factor to the new cases per 100K people of each state.

Top Risk factor	Region
Non-work trip/ppl	MO
Miles/ppl	KS
Unemployment rate	IA
% of WFH	AL, ME, RI, VA,
Active cases/1000 ppl	USA,AK, AZ, AR, CA, CO, CT, DE, FL, GA, HI, ID, IL, IN,KY, LA, MD, MA, MI, MN, MS, MT, NE, NV, NH, NM, NY, ND,OH,OK, OR, PA, SC, SD, TN, TX, UT, VT, WA
Imported cases	NJ, NC, WV, WI, WY

5.3 Estimating the causal effect of the face mask policy decisions

5.3.1 Setup. As mentioned above, we focus on using CausalImpact to estimate the causal effect of the face mask policy decisions on number of cases per 100,000 population. We first identify our treated states, which were affected by the interventions, lifting or reissuing state-wide mask orders. We focus on analyzing the lifting mask mandate orders of highly populated states such as Illinois, Michigan, and Texas and the reissuing mask mandate order of Illinois, Louisiana, and Oregon. In the future, we can extend this analysis to more states. Because the time frames of the interventions of different states are different, we can only have one treated state for each analysis as the treated group. After that, we identify pre-period and post-period for each treated state. For simplicity's sake, pre-periods of all states start from the week of March 16, 2020. Because the symptoms for Covid-19 appear approximately two weeks after exposure, we expect that there will be a lag of approximately the same window for the effect of the mask mandate removal or reissue to be observed. Therefore, the end date of the states' pre-periods will be extended by approximately two weeks. The post-periods for the treated start a week right after the end dates of the pre-periods and end after eight weeks. We could extend this period further to evaluate the full effect; however, the longer the window is used, the more difficult it is to isolate the effect of one treatment with other factors. Table 5 and 6 provide details of the above steps for removing mask orders and reissuing mask orders respectively.

We then find control states as the control groups for each treated group based on their similarities in the pre-period until the intervention. There are multiple approaches to do this. The most straightforward approach is to use the Euclidean distance. However, this approach implicitly imposes a penalty if the relationship between the markets shifts temporarily. While it is desirable to have a consistent agreement between the treated state and the control states, the occasional historical shifts should not preclude viable candidates for the reference market. We can also calculate pairwise correlations between each of the treated and its control groups directly on the number of cases per 100,000 population series, but this would ignore the size and produce spurious correlations caused by trends. Therefore, we calculate the pairwise correlations between each of the treated and controls based on the percentage changes of the number of cases per 100,000 population series and choose the states which are highly correlated (>70%) with our treated. Another popular approach is to use Dynamic Time Warping to find the distance along the warping curve - the best alignment between the time series; however, this will not be covered in this paper and is considered as future work.

Lastly, for each treated state, we estimate the counterfactual series of the treated states in the post-period by training predictive BSTS model to learn the relationship between it and its control groups in the pre-period. After that, we can estimate the average treatment effect (ATE) by taking the difference between the observed series and the estimated counterfactual series. The BSTS model is written as:

$$
\begin{aligned}
y_t &= \mu_t + \gamma_t + \beta^T x_t + \epsilon_t, \\
\mu_t &= \mu_{t-1} + \delta_{t-1} + u_t, \\
\delta_t &= \delta_{t-1} + v_t, \\
\gamma_t &= -\sum_{s=1}^{S-1} \gamma_t + w_t,
\end{aligned}
$$

where y_t is the time series of interest, μ_t is the trend term or local level over time, γ_t is the seasonal term of periodic events, and $\beta^T x_t$ are regressor terms of the other time series that are predictive of y_t. In this case, $\beta^T x_t$ are the control states. This method is flexible and we it also has built-in feature selection using spike-and-slab prior. An important arguments that we need to be aware is the standard deviation of the Gaussian random walk of the local level. We keep it as the default option of 0.01 because we expect the treated series are well behaved and stable with low residual volatility after regressing out known predictors. This method of estimating the ATE with matching works with three main assumptions: ignorability, stable unit treatment value assumption (SUTVA), and predictability. Ignorability assumption implies that there are no unobserved confounders. The SUTVA assumption implies that control series do not get affected by the treatment of the treated series. Predictability implies that it is possible to model the treated series using the control series.

Table 5. Removing mask orders' dates, pre-periods, and post-periods for Illinois, Michigan, and Texas.

State	IL	MI	TX
Removal date	June 11, 2021	June 22, 2021	March 10, 2021
Pre-period	March 16,2020 - Jun 21, 2021	March 16,2020 - Jun 28, 2021	March 16, 2020 - March 15, 2021
Post-period	Jun 28, 2021 - August 23, 2021	July 05, 2021 - August 30, 2021	March 22, 2021 - May 17, 2021

Table 6. Reissuing mask orders' dates, pre-periods, and post-periods for Illinois, Louisiana, and Oregon.

State	IL	LA	OR
Reissue date	Aug 30, 2021	Aug 2, 2021	Aug 13, 2021
Pre-period	Mar 16,2020 - Sep 6, 2021	Mar 16,2020 - Aug 9, 2021	Mar 16, 2020 - Aug 23, 2021
Post-period	Sep 13, 2021 - Nov 08, 2021	Aug 16, 2021 - Oct 11, 2021	Aug 30, 2021 - Oct 25, 2021

Table 7. Removing mask orders' average absolute affect with 95% CI, average relative effect with 95% CI, cumulative absolute effect with 95% CI, cumulative relative effect with 95% CI, tail-area probability p, probability of a causal impact, and most importance variables based on inclusion probability of Illinois, Michigan, and Texas.

State	IL	MI	TX
Average Absolute Effect	-2.1 [-55, 59]	36 [-12, 92]	-44 [-106, 20]
Average Relative Effect	-2.3% [-60%, 64%]	114% [-37%, 291%]	-37% [-88%, 17%]
Cumulative Absolute Effect	-19.3 [-495, 528]	324 [-106, 832]	-396 [-950, 180]
Cumulative Relative Effect	-2.3% [-60%, 64%]	114% [-37%, 291%]	-37% [-88%, 17%]
Tail-area probability p	0.44076	0.07071	0.09907
Probability of a causal impact	56%	93%	90%
Most importance variables	IA, IN, MN, PA, & ND	MN, PA, OH, & IN	FL, KY, SC, VA, NC, & AR

Table 8. Reissuing mask orders' average absolute affect with 95% CI, average relative effect with 95% CI, cumulative absolute effect with 95% CI, cumulative relative effect with 95% CI, tail-area probability p, probability of a causal impact, and most importance variables based on inclusion probability of Illinois, Louisiana, and Oregon.

State	IL	LA	OR
Average Absolute Effect	-114 [-157, -68]	-121 [-222, 8.3]	80 [23, 130]
Average Relative Effect	-42% [-57%, -25%]	-24% [-44%, 1.7%]	42% [12%, 69%]
Cumulative Absolute Effect	-1030 [-1411, -612]	-1092 [-1997, 74.7]	717 [203, 1173]
Cumulative Relative Effect	-42% [-57%, -25%]	-24% [-44%, 1.7%]	42% [12%, 69%]
Tail-area probability p	0.00101	0.0404	0.00403
Probability of a causal impact	99.8993%	95.96%	99.59677%
Most importance variables	MN, MA, ND, WI, MD & CO	FL, MS, TN, & NC	CO, MS, HI, KS, IA, & IL

5.3.2 Results. Overall, we can see that removing mask orders does not have causal effects on Illinois, Michigan, and Texas. The probabilities of obtaining these effect by chance are 0.44, 0.07, 0.1, respectively, which mean the impact may be spurious and would generally not be considered statistically significant. This may be the case because people may not practice the policy. For states that pro-mask mandate, people will still wear them. For states that are anti-mask mandate, they do not wear them. Table 7 provide detailed results of removing mask orders.

On the other hand, Reissuing mask orders do have causal effects on the three states Illinois, Louisiana, and Oregon. The average absolute effects of Illinois and Louisianas' reissuing mandates are -114 and -121. They reduce the new cases per 100,000 population by 114 and 121 cases, respectively, by reissuing mask mandates.

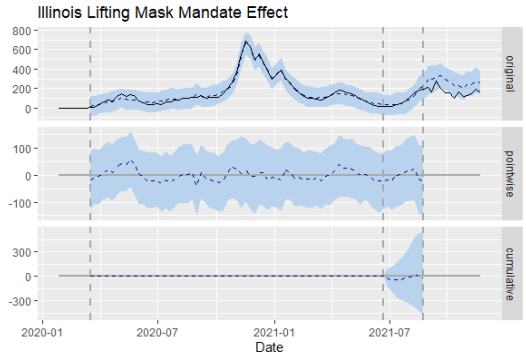

Fig. 7. Original, pointwise, and cumulative plots of removing mask order. The first vertical line is the start of the pre-period, the second vertical line is the intervention, and the last one is the end of the post-period which is 8 week from the intervention shifted by 2 week. The first panel is the real series (solid black line) vs. the estimated counterfactual series (dotted line). The second panel is the pointwise effect, the different between the 2 lines of the first panel. The last panel is the cumulative effect.

However, Oregon has an unexpected average absolute effect of 80, which means Oregon increases 80 cases per 100,000 population by reissuing mask mandates. This is counterintuitive because we don't expect that reissuing mask mandates would increase the number of cases. Because we use other states to predict the counterfactual series of Oregon, there are maybe anomalies in the control variables that may have caused an overly optimistic expectation of what should have happened in the response variable in the absence of the intervention. Therefore, more investigation is needed to detect and eliminate possible anomalies. We can also use other time series matching methods such as Dynamic Time Warping to find the control groups better. Table 8 provide detailed results of reissuing mask orders.

Because of page limits, we do not include all results plots for this section. We will provide these plots upon request. However, we do think that Tables 7 and 8 already provide detail results.

6 DISCUSSIONS AND NEXT STEPS

Although we have achieved great success on identifying all of the causal risk factors to the spread pattern of COVID-19, we probably need some collaborate with experts in epidemiology and social science to study the better explain those causalities. For instance, can we claim the state governments of New Jersey, North Carolina, West Virginia, Wisconsin and Wyoming perform better than other states because the leading risk factor of new cases in those states is the imported cases. Or, it is merely because of many infected people crossed their state borders to escape from outbreak centrals like NYC. Moreover, what assumptions and restrictions should we further consider to reduce the likelihood of spurious causation. In the future, as of the causal discovery on temporal factors, we plan to use some nonlinear Granger causality tests with Artificial Neural Networks which frees up the linear signal assumption within the time series. However, this comes at the cost of larger sample sizes which is not readily available to us right now. In addition, we plan

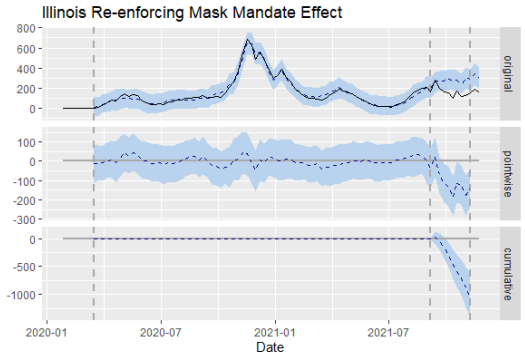

Fig. 8. Original, pointwise, and cumulative plots of reissuing mask order. The first vertical line is the start of the pre-period, the second vertical line is the intervention, and the last one is the end of the post-period which is 8 week from the intervention shifted by 2 week. The first panel is the real series (solid black line) vs. the estimated counterfactual series (dotted line). The second panel is the pointwise effect, the different between the 2 lines of the first panel. The last panel is the cumulative effect.

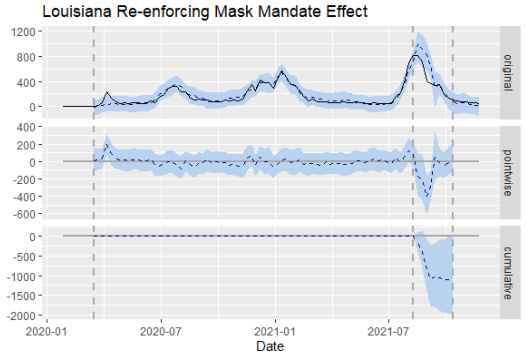

Fig. 9. Original, pointwise, and cumulative plots of reissuing mask order. The first vertical line is the start of the pre-period, the second vertical line is the intervention, and the last one is the end of the post-period which is 8 week from the intervention shifted by 2 week. The first panel is the real series (solid black line) vs. the estimated counterfactual series (dotted line). The second panel is the pointwise effect, the different between the 2 lines of the first panel. The last panel is the cumulative effect.

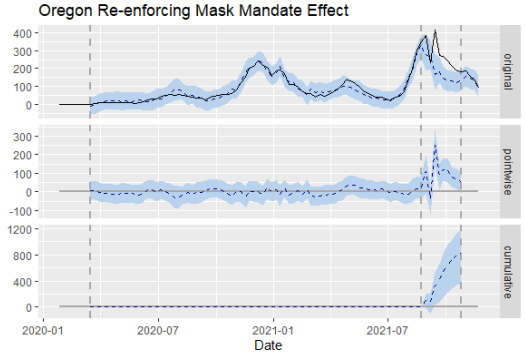

Fig. 10. Original, pointwise, and cumulative plots of reissuing mask order. The first vertical line is the start of the pre-period, the second vertical line is the intervention, and the last one is the end of the post-period which is 8 week from the intervention shifted by 2 week. The first panel is the real series (solid black line) vs. the estimated counterfactual series (dotted line). The second panel is the pointwise effect, the different between the 2 lines of the first panel. The last panel is the cumulative effect.

to perform estimations on the causal effect of a temporal risks factor on the new cases per 100,000 people, which is also in the time series form on both the federal and state levels. For example, we can attempt to answer counterfactual questions like if the percentage of work-from-home was decreased by 1% in the first week of 2021 across the entire country, how large the increment of new infected case should we expect to see.

As of the estimating causal impact of face mask policies, we still need to do more experiments on matching the control states with the treated states. One promising approach is Dynamic Time Wrapping. We will then attempt to expand our analysis to all states with mask policies. We can try out other methods to estimate causal effect such as Difference in Differences and Synthetic Control. We will also consider using deep learning model such as using Vertex AI AutoML instead of BSTS.

REFERENCES

Matteo Bonvini, Edward Kennedy, Valerie Ventura, and Larry Wasserman. 2021. Causal Inference in the Time of Covid-19. *arXiv preprint arXiv:2103.04472* (2021).

Kay H. Brodersen, Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L. Scott. 2015. Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics* 9 (2015), 247–274.

Cees Diks and Marcin Wolski. 2016. Nonlinear granger causality: Guidelines for multivariate analysis. *Journal of Applied Econometrics* 31, 7 (2016), 1333–1351.

Luca Faes, Daniele Marinazzo, Alessandro Montalto, and Giandomenico Nollo. 2014. Lag-specific transfer entropy as a tool to assess cardiovascular and cardiorespiratory information transfer. *IEEE Transactions on Biomedical Engineering* 61, 10 (2014), 2556–2568.

Clive WJ Granger. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society* (1969), 424–438.

Clive WJ Granger. 1980. Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and control* 2 (1980), 329–352.

Arthur Gretton, Kenji Fukumizu, Choon Hui Teo, Le Song, Bernhard Schölkopf, Alexander J Smola, et al. 2007. A kernel statistical test of independence.. In *Nips*, Vol. 20. Citeseer, 585–592.

Katerina Hlaváčková-Schindler, Milan Paluš, Martin Vejmelka, and Joydeep Bhattacharya. 2007. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports* 441, 1 (2007), 1–46.

Youssef Hmamouche. 2020. NlinTS: An R Package For Causality Detection in Time Series. *R J.* 12, 1 (2020), 21.

Diviyan Kalainathan, Olivier Goudet, and Ritik Dutta. 2020. Causal Discovery Toolbox: Uncovering causal relationships in Python. *J. Mach. Learn. Res.* 21 (2020), 37–1.

Zhouxuan Li, Tao Xu, Kai Zhang, Hong-Wen Deng, Eric Boerwinkle, and Momiao Xiong. 2021. Causal Analysis of Health Interventions and Environments for Influencing the Spread of COVID-19 in the United States of America. *Frontiers in Applied Mathematics and Statistics* 6 (2021), 69. https://doi.org/10.3389/fams.2020.611805

Jonathan M Nichols, Frank Bucholtz, and Joe V Michalowicz. 2013. Linearized transfer entropy for continuous second order systems. *Entropy* 15, 8 (2013), 3186–3204.

Jonas Peters, Joris M Mooij, Dominik Janzing, and Bernhard Schölkopf. 2014. Causal discovery with continuous additive noise models. (2014).

Thomas Schreiber. 2000. Measuring information transfer. *Physical review letters* 85, 2 (2000), 461.

Steven L Scott and Hal R Varian. 2014. Predicting the present with Bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation* 5, 1-2 (2014), 4–23.

Claude Elwood Shannon. 1948. A mathematical theory of communication. *The Bell system technical journal* 27, 3 (1948), 379–423.

Christopher A Sims. 1980. Macroeconomics and reality. *Econometrica: journal of the Econometric Society* (1980), 1–48.

Xiaotong Wen, Govindan Rangarajan, and Mingzhou Ding. 2013. Multivariate Granger causality: an estimation framework based on factorization of the spectral density matrix. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371, 1997 (2013), 20110610.