Synthetic Controls with Multiple Outcomes: Estimating the Effects of Non-Pharmaceutical Interventions in the COVID-19 Pandemic*

Wei Tian [a] Seojeong Lee [b,†] Valentyn Panchenko [c]

April 6, 2023

[a,c] *School of Economics, University of New South Wales, Sydney NSW Australia*
[b] *Department of Economics, Seoul National University, Seoul, South Korea*

Abstract

We propose a generalization of the synthetic control method to a multiple-outcome framework, which improves the reliability of treatment effect estimation. This is done by supplementing the conventional pre-treatment time dimension with the extra dimension of related outcomes in computing the synthetic control weights. Our generalization can be particularly useful for studies evaluating the effect of a treatment on multiple outcome variables. To illustrate our method, we estimate the effects of non-pharmaceutical interventions (NPIs) on various outcomes in Sweden in the first 3 quarters of 2020. Our results suggest that if Sweden had implemented stricter NPIs like the other European countries by March, then there would have been about 70% fewer cumulative COVID-19 infection cases and deaths by July, and 20% fewer deaths from all causes in early May, whereas the impacts of the NPIs were relatively mild on the labor market and economic outcomes.

Keywords: Synthetic control; Policy evaluation; Causal inference; Public health

JEL codes: C32, C54, I18

*The authors thank Barton Lee, Stanley Cho, Artem Prokhorov, Tatsushi Oka, Tue Gorgens, Firmin Doko Tchatoka, and seminar participants at UNSW, University of Sydney, the Australia and New Zealand Econometrics Study Group (ANZESG2020) at Monash University, and University of California Irvine, for their constructive suggestions and feedback. Tian acknowledges that this research was supported by the Australian Government Research Training Program (RTP) Scholarship.

†Corresponding author: s.jay.lee@snu.ac.kr

1 Introduction

The Synthetic Control (SC) method (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015; Abadie, 2021) is a popular method for estimating the effect of a policy or intervention on an aggregate unit, such as a country or a city. The procedure constructs a synthetic control unit as a convex combination of the control units to minimize the difference between the outcome of the treated unit and the synthesized unit before the treatment, and then estimates the per-period treatment effects using the differences between the observed outcome of the treated unit and the counterfactual SC outcome after the treatment. Abadie et al. (2010) show that the bias of the SC estimator is bounded by a function that is inversely proportional to the number of pre-treatment periods, provided that the treated unit is perfectly approximated by the synthetic control in the pre-treatment periods. Ferman and Pinto (2021) propose a demeaned version of the SC estimator allowing for imperfect pre-treatment fit.

Our first contribution is methodological. We generalize the conventional single-outcome SC method to a multiple-outcome framework by supplementing the time dimension with the extra dimension of related outcomes in the same domain. This in effect incorporates additional pre-treatment outcomes as matching variables and reduces the risk of overfitting. We show that the bias of the multiple-outcome SC method is inversely proportional to not only the number of pre-treatment periods but also the number of related outcome variables. Therefore, the order of bias is of a smaller order than the conventional single-outcome SC estimator. This implies that using the same number of pre-treatment periods, the multiple-outcome SC estimator can be less biased than the conventional SC estimator.

In addition to the theoretical advantage, the multiple-outcome SC estimator has important practical advantages. Since the SC weights are calculated by matching both the related outcome variables and the pre-treatment periods, it can be applied even when the number of pre-treatment periods is small.[1] The multiple-outcome SC can be used when we only observe multiple related outcomes in a single pre-treatment period. This is demonstrated

[1]Some previous studies have used the conventional SC method by matching on a single outcome with a few pre-treatment periods (e.g., Billmeier and Nannicini, 2013 and Cavallo et al., 2013), which may produce biased estimates due to overfitting (Abadie et al., 2015).

in the treatment backdating exercise in Section 4.4.

Furthermore, our method can alleviate the concern on matching over a long period of time, which might be subject to misspecification error due to structural breaks in the relationship between the outcome of interest and the underlying predictors (Abadie, 2021).

Other studies that analyze multiple outcomes using the SC method include Robbins et al. (2017), Klößner and Pfeifer (2018) and Samartsidis et al. (2020).[2] Robbins et al. (2017) consider a setting with high-dimensional data at a granular level, with multiple treated units and a large number of control units, whereas we focus on typical synthetic control settings with a small number of aggregate units. Klößner and Pfeifer (2018) differs from our study in that they take account of inter-dependencies among multiple outcomes using vector autoregression models rather than factor models, and that their method requires a large number of pre-treatment periods even with multiple outcomes, whereas our method allows the number of pre-treatment periods to be small in the presence of multiple related outcomes. Samartsidis et al. (2020) adopt the Bayesian framework.

Apart from the methodological innovation, our paper also contributes empirically to understanding the impacts of non-pharmaceutical interventions (NPIs) in the COVID-19 pandemic. Examples of NPIs include closing school and workplaces, restricting travels and gatherings, public information campaigns, and so on. Since the outbreak of the pandemic, most countries have resorted to NPIs to reduce the spread of the virus and to prevent the health system from being overwhelmed, in the absence of vaccines. Understanding the effect of NPIs not only on public health related outcome variables but also on labor market and the economy is important for scholars and policy-makers to be better prepared for a possible future pandemic.

In the empirical application, we illustrate the multiple-outcome SC method by extending the analysis of Born et al. (2020) and Cho (2020) to multiple public health, labor market, and economic outcomes. Our results suggest that had Sweden implemented stricter NPIs as the other European countries by March, 2020, the cumulative numbers of COVID-19 infection cases and deaths would have been reduced by 70% and 68% respectively by July, and deaths from all causes would have been 20% fewer in early May. The impact

[2] Cattaneo et al. (2021) mention that researchers may want to match multiple outcomes, but they do not provide a formal treatment of this issue as their focus is on prediction intervals.

on mortality was larger for males and most visible for people older than 60. As for the labor market, we find that stricter NPIs would increase absence from work by almost 76% mainly through temporary layoffs, and reduce total hours worked by about 12%, in the second quarter of 2020. The impacts would quickly vanish in the third quarter, and there would be no discernible effect on the employment rate throughout. In terms of the economy, we find that stricter NPIs would shrink the volume of retail sales by 5%-13% from March to May, almost exclusively due to reduced sales in non-food products. However, we do not find any statistically significant effects of stricter NPIs on the other economic outcomes including GDP, import, export, industrial production, and CPI, which indicates that almost all of the contraction in the economy was due to the pandemic itself rather than the NPIs.

Our empirical results are in line with the existing studies, which largely find that the NPIs had statistically and economically significant effects on the public health outcomes, but a much smaller role in the labor market and economic downturns.[3] There are interesting new findings as well. First, we find that the potential reduction in deaths from all causes due to the NPIs is slightly smaller than the reduction in COVID-19 deaths. One probable explanation is that people with existing conditions or weaker immune systems are more susceptible to COVID-19, and thus mortality that would have been attributed to other causes may have been counted as COVID-19 deaths instead. The NPIs may also have limited the spread of other transmissible diseases through promoting good hygiene behaviors and reducing face-to-face interactions. In either case, the estimated reduction in deaths from all causes due to the NPIs may serve as a lower bound for the reduction in COVID-19 deaths. Second, the NPIs would have significant effects on temporary absence from work and total hours worked among the employed, but no effect on the employment rate. This is likely due to the various employment support policies that were carried out across European countries, with the aim of preserving jobs through income support and wage subsidies (OECD, 2020).

The rest of the paper is organized as follows. Section 2 describes the theoretical framework for the the multiple-outcome synthetic control method. Section 3 compares the multiple-outcome synthetic control method with the conventional single-outcome synthetic

[3] A more detailed literature review can be found in the supplementary appendix.

control method using Monte Carlo simulations. Section 4 presents the empirical results and robustness checks. Section 5 concludes and discusses the limitations. The technical details and proofs are collected in Appendix A.

2 Theoretical Framework

2.1 Multiple outcomes framework

Suppose that we observe K outcomes in domain $\mathbb{K} = \{1, 2, \ldots, K\}$ for $J + 1$ units over T time periods, where a domain refers to a collection of related outcomes driven by the same set of observed and unobserved predictors (or factors). For example, the economic domain contains different measures of the economic performance, such as GDP, industrial production, retail sales, and CPI, which can be assumed to depend on the same set of underlying predictors such as infrastructure, technology, natural resources, demographic composition, work ethic, etc.

Without loss of generality, we assume that the first unit ($i = 1$) receives the treatment at period $T_0 + 1 \leq T$ and remains treated afterwards, while all the other J units ($i = 2, \ldots, J + 1$) are untreated throughout the window of observation. Denoting the binary treatment status for unit i at time t as D_{it}, we have $D_{it} = 1$ for $i = 1$ and $t > T_0$, and $D_{it} = 0$ otherwise. We consider fixed J, large T and K asymptotics.

We are interested in the effect of the treatment on a single or multiple outcomes in domain $\mathbb{K}$ for the treated unit after the treatment:

$$\tau_{1t,k} = Y^1_{1t,k} - Y^0_{1t,k}, \quad t > T_0, \ k \in \mathbb{K}, \tag{1}$$

where $Y^1_{1t,k}$ is the potential outcome under the treatment, and $Y^0_{1t,k}$ is the potential outcome without the treatment, so that the observed outcome can be written as $Y_{1t,k} = D_{1t}Y^1_{1t,k} + (1 - D_{1t})\, Y^0_{1t,k}$.[4] Since we only observe the treated potential outcome but not the untreated potential outcome for unit 1 at $t > T_0$, we need to predict the counterfactual outcome $Y^0_{1t,k}$. This counterfactual outcome is constructed as a convex combination (weighted average) of the control units.

[4] As in Abadie et al. (2010), we treat $\tau_{1t,k}$ as given once the sample is drawn.

Suppose that the untreated potential outcome $k \in \mathbb{K}$ for unit i at time t is given by an interactive fixed effects model

$$Y_{it,k}^0 = \delta_{t,k} + \boldsymbol{Z}_i'\boldsymbol{\theta}_{t,k} + \boldsymbol{\mu}_i'\boldsymbol{\lambda}_{t,k} + \varepsilon_{it,k}, \tag{2}$$

where $\delta_{t,k}$ is the time trend in outcome k, $\boldsymbol{Z}_i$ and $\boldsymbol{\mu}_i$ are the $r \times 1$ and $f \times 1$ vectors of observed and unobserved predictors of $Y_{it,k}^0$ with outcome-specific coefficients $\boldsymbol{\theta}_{t,k}$ and $\boldsymbol{\lambda}_{t,k}$, respectively, and $\varepsilon_{it,k}$ is the idiosyncratic transitory shock. The model in (2) is similar to that of Abadie et al. (2010) and Abadie (2021) but is allowed to have the outcome-specific constant and coefficients of observable and unobservable predictors.

Assuming that the individual-specific unobserved predictor $\boldsymbol{\mu}_i$ is common to the outcomes in the same domain is the key assumption in our model. If the outcomes depend on different sets of unobserved predictors (thus $\boldsymbol{\mu}_{i,k}$), then we lose the benefit of matching on multiple related outcomes in terms of having the same order of bias with the conventional SC method. Nevertheless, we argue that our assumption is no different from the standard factor analysis where a low number of common factors underlying related variables is assumed. Moreover, our model accommodates the case when the unobserved predictors are separable: $\boldsymbol{\mu}_{i,k} = \boldsymbol{\mu}_i + \boldsymbol{u}_k$ where $\boldsymbol{u}_k$ is outcome-specific predictors.

A few more remarks on the model. The model does not exclude the possibility that some outcomes in the domain depend on other predictors that are serially uncorrelated and independent from the included predictors and the treatment status, which can thus be treated as part of the transitory shocks. In addition, the coefficients may contain zero so that the corresponding predictors may affect some outcomes in some periods, but not all outcomes in all periods, as long as there is enough variation in the coefficients across different pre-treatment periods or outcomes, as specified in Condition 2.

Although we do not have a formal test for the validity of our model, we may find some suggestive evidence from the empirical applications. For example, we show in the Supplementary Appendix, B.2., that the synthetic control for West Germany constructed by matching on multiple economic variables in 1989 alone can track the trajectory of West Germany's GDP closely for 30 years, and the trajectory for the multiple-outcome synthetic control after the treatment is also very similar to that of the synthetic control constructed by matching on 30 years of pre-treatment GDP.

A synthetic control is constructed using a convex combination of the control units such that the synthetic control matches the treated unit in terms of the observed predictors and the pre-treatment values of the K related outcomes. This can be achieved if the matching variables of the treated unit is in the convex hull of those of the control units.

Suppose there is a set of weights $(\hat{w}_2, \ldots, \hat{w}_{J+1})$ such that $\hat{w}_j \geq 0$ for $j = 2, \ldots, J+1$, $\sum_{j=2}^{J+1} \hat{w}_j = 1$ for all $t \leq T_0$ and $k \in \mathbb{K}$

$$\sum_{j=2}^{J+1} \hat{w}_j \boldsymbol{Z}_j = \boldsymbol{Z}_1, \qquad \sum_{j=2}^{J+1} \hat{w}_j Y_{jt,k} = Y_{1t,k}. \tag{3}$$

Eq. (3) is analogous to the pre-treatment fit condition in Eq. (2) of Abadie et al. (2010), which is a convenient technical condition to show the asymptotic unbiasedness of the synthetic control estimator. As the authors discuss, this condition can be satisfied only approximately in practice.[5] The multiple-outcome synthetic control estimator for $\tau_{1t,k}$ is then constructed as

$$\widehat{\tau}_{1t,k} = Y_{1t,k} - \sum_{j=2}^{J+1} \hat{w}_j Y_{jt,k}. \tag{4}$$

To facilitate a straightforward comparison between the multiple-outcome SC estimator and the conventional single-outcome SC estimator based only on the kth outcome, let $\{\tilde{w}_j^{(k)}\}_{j=2}^{J+1}$ be the single-outcome SC weights such that $\tilde{w}_j^{(k)} \geq 0$ for $j = 2, \ldots, J+1$, $\sum_{j=2}^{J+1} \tilde{w}_j^{(k)} = 1$, $\sum_{j=2}^{J+1} \tilde{w}_j^{(k)} \boldsymbol{Z}_j = \boldsymbol{Z}_1$ and $\sum_{j=2}^{J+1} \tilde{w}_j^{(k)} Y_{jt,k} = Y_{1t,k}$ for all $t \leq T_0$. The conventional SC estimator for $\tau_{1t,k}$ is

$$\widetilde{\tau}_{1t,k} = Y_{1t,k} - \sum_{j=2}^{J+1} \tilde{w}_j^{(k)} Y_{jt,k}. \tag{5}$$

The following proposition shows that the bias of the multiple-outcome SC method is reducing faster than that of the conventional single-outcome SC method.

[5] Ferman and Pinto (2021) formally shows that under typical large T asymptotic if this condition does not hold, the bias of the conventional single-outcome SC estimator does not disappear as the number of pre-treatment periods increases. Nonetheless, they show that a demeaned version of the SC estimator can still perform better than the difference-in-difference estimator. Ferman (2021) relaxes this assumption by using large T, large J asymptotic. We do not pursue this setting with our multiple-outcome SC, but leave it for the future research.

Proposition 1. *Suppose that J is fixed, whereas T_0 and K are increasing, and the technical conditions*[6] *are satisfied, then, for any $t > T_0$,*

$$|\mathbb{E}(\widetilde{\tau}_{1t,k}) - \tau_{1t,k}| = O\left(\frac{1}{\sqrt{T_0}}\right),$$
$$|\mathbb{E}(\widehat{\tau}_{1t,k}) - \tau_{1t,k}| = O\left(\frac{1}{\sqrt{KT_0}}\right).$$

Since the bias of the multiple-outcome SC estimator reduces as the number of pre-treatment periods and the number of related outcomes increases, in practice we can use the SC method even when the number of pre-treatment periods is small, if multiple related outcomes are available and the treated unit can be closely approximated by the synthetic control for these outcomes in the pre-treatment periods. In Appendix A, we compare the multiple-outcome SC weights with the single-outcome SC weights and find that they are generically different.

The estimation and inference procedures are similar to those of the conventional SC method (Abadie, 2021), by including the multiple related outcomes in the pre-treatment periods as matching variables. To save space, we have placed the detailed description of the estimation and inference procedures in the Supplementary Appendix. The Supplementary Appendix, B.2, also illustrates the proposed method on the classic application of Abadie et al. (2015), the effect of the 1990 German reunification on West Germany's GDP per capita.

2.2 Extensions

We discuss two extensions that will be useful for the empirical application. One is about using demeaned outcomes, i.e., outcomes in differences with respect to their pre-treatment averages, to allow better pre-treatment fits when the differences in the level of the outcomes for different units are relatively stable over time. This may be especially helpful in the multiple-outcome framework, where the relative positions of the units may vary across different outcomes. The other extension, which is discussed in Appendix A, is to include an additional assumption on the functional form of the treatment effect, which is necessary

[6]The technical conditions and proofs are given in Appendix A.

if we want to estimate the treatment effect on the untreated using the synthetic control method in cases with many treated units.

2.2.1 Adjusting for differences in levels

The conventional SC method requires the treated unit to be in the convex hull of the control units in terms of the pre-treatment matching variables. However, there are cases where the treated unit is extreme in the values of the matching variables, such that no convex combination of the control units can closely approximate the outcome of the treated unit in the pre-treatment periods. In particular, it is often the case in practice that there are relatively stable differences in the level of the outcome across units before the treatment. In such cases, Ferman and Pinto (2021) and Abadie (2021) suggest constructing the synthetic control using demeaned outcomes, i.e., outcomes measured in differences with respect to their pre-treatment means. This enables the synthetic control to track the dynamics in the outcome of the treated unit over time, while allowing the levels to differ by a constant amount, which is similar to the "parallel trends" assumption in the difference-in-differences method.[7] Apart from allowing a better pre-treatment fit for the treated unit, using demeaned outcomes has the additional merit that it helps correct the size distortion of the permutation test based on the post-to-pre-treatment RMSPE (Root Mean Squared Prediction Error) ratios obtained from permuting the treatment status among all units.

In the multiple-outcome framework, the relative position of the units can vary across different outcomes, making it difficult to match on multiple outcomes simultaneously. The demeaning process would be helpful in these circumstances by improving the pre-treatment fits.

Provided that $T_0 \geq 2$, replacing the outcomes, $Y_{is,k}$ in (3) with demeaned outcomes $\dot{Y}_{it,k} = Y_{it,k} - \frac{1}{T_0}\sum_{s=1}^{T_0} Y_{is,k}$, we obtain a new set of weights, $(\hat{w}_2, \ldots, \hat{w}_{J+1})$. We can then construct the multiple-outcome synthetic control estimator for $\tau_{1t,k}$ using the demeaned

[7] Using demeaned outcomes is also similar to a proposal in Doudchenko and Imbens (2017), which includes an intercept when minimizing the difference between the synthetic control and the treated unit in the matching variables.

outcomes as

$$\widehat{\tau}_{1t,k}^{\mathrm{DM}} = \dot{Y}_{1t,k} - \sum_{j=2}^{J+1} \hat{w}_j \dot{Y}_{jt,k} \tag{6}$$

to account for the differences in the level of the outcomes. Similarly with Proposition 1, the bias of this estimator can be shown to shrink as the number of related outcomes and pre-treatment periods goes to infinity.

Corollary 1. *Suppose that J is fixed, whereas T_0 and K are increasing, the demeaned outcomes are used, and the technical conditions are satisfied, then for any $t > T_0$,*

$$\left|\mathbb{E}(\widehat{\tau}_{1t,k}^{DM}) - \tau_{1t,k}\right| = O\left(\frac{1}{\sqrt{KT_0}}\right).$$

3 Monte Carlo Simulations

In this section, we conduct Monte Carlo simulations to compare the multiple-outcome SC method and the conventional single-outcome SC method.

We fix the number of post-treatment periods at 1, and the number of units at 30, with a single treated unit and 29 control units. The treatment effect is set to 0, so that the treated potential outcomes are the same with the untreated potential outcomes.[8] The data generating process (DGP) is as follows:

$$Y_{it,k} = \delta_{t,k} + \boldsymbol{Z}_i'\boldsymbol{\theta}_{t,k} + \boldsymbol{\mu}_i'\boldsymbol{\lambda}_{t,k} + \varepsilon_{it,k}, \tag{7}$$

where $\boldsymbol{Z}_i$ is the vector of 2 observed predictors and $\boldsymbol{\mu}_i$ is the vector of 4 unobserved predictors. The observed and unobserved predictors, $\boldsymbol{Z}_i$ and $\boldsymbol{\mu}_i$, are drawn independently from the uniform distribution $U[-1, 1]$ for the control units, and $U[-d, d]$ with $d \in [0, 1]$ for the treated unit. When $d = 1$, the treated unit is equally likely to obtain extreme values in the outcomes as the control units. When d is smaller, the treated unit is more likely to be in the convex hull of the control units, and thus have better pre-treatment fits.

To generate outcomes that are closer to real data, where there are often clear differences in the level of the outcomes across units and that the levels are relatively stable over time,

[8] The zero treatment effect is set to investigate the size of the test. It does not affect the bias and the standard derivation of the SC estimator.

we set the variance of the mean of the coefficients to be large relative to the variance of the coefficients and the transitory shocks. As such, the time trend ($\delta_{t,k}$) and the coefficients ($\boldsymbol{\theta}_{t,k}$ and $\boldsymbol{\lambda}_{t,k}$) are drawn independently from the normal distribution $N(\omega_k, 1)$ with $\omega_k \sim N(0, 10^2)$, and the transitory shocks, $\varepsilon_{it,k}$, are drawn independently from the standard normal distribution.[9]

Figure 1 displays the trajectories for two of the related outcomes from a typical simulated sample with $d = 1$ and $T_0 = 10$. The trajectories for the treated unit are in black, and the control units in gray. As intended, there are visible and stable differences in the level of the outcomes, and the levels for the treated unit are different across different outcomes. Although the two outcomes share the same underlying predictors, they may appear different due to differences in the time trends, the coefficients on the predictors and the transitory shocks. Overall, the trajectories of the outcomes generated in our simulation look very similar to those in real data, e.g., as seen in Figure 3 in the empirical application.

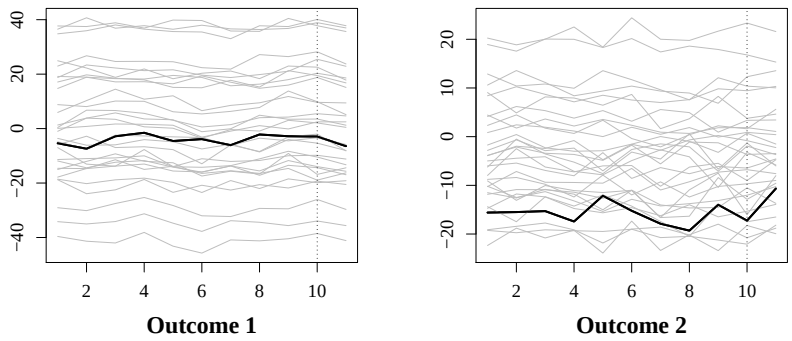


Figure 1: A Simulated Example

There are 5000 simulations for each pair of d and T_0. In each simulation, the observed and unobserved predictors are drawn only once so that the outcomes are related by sharing the same underlying predictors, while the time trend, the coefficients and the transitory shocks are drawn independently across different outcomes. We compare the conventional

[9]When ω_k is a constant, there would be no distinguishable levels in the outcomes for different units, and demeaning the outcomes would not be useful in this case.

single-outcome SC estimator and the multiple-outcome SC estimators constructed using $K = 1$, 3, and 10 demeaned outcomes, respectively. The only difference between the single-outcome SC method and the multiple-outcome SC method when $K = 1$ is the use of the demeaned outcomes. To measure their performances, we estimate the treatment effect on outcome 1 for the treated unit at $t = T_0 + 1$, and compute the average absolute bias and standard deviation of the estimators as well as the average rejection rate of the 10% test in 5000 simulations. The null hypothesis of zero treatment effect is rejected in each simulation, if the RMSPE ratio for the treated unit is ranked among the largest 10%, i.e., top 3 in our sample, in the permutation test. When the estimation improves with better pre-treatment fits and larger numbers of pre-treatment periods and related outcomes, we expect the average absolute bias and standard deviation of the estimators to be closer to $\sqrt{\frac{2}{\pi}} \approx 0.8$ and 1, which are the mean and standard deviation of the standard normal distribution folded at the mean (half-normal distribution).[10] When the distributions of the RMSPE ratio for the treated unit and the control units are close, so that there is little size distortion in the permutation test, we expect the average rejection rate of the 10% test to be close to the nominal rejection rate at 10%. The pre-treatment RMSPE is also reported as a measure of pre-treatment fit.

Several findings emerge from the results of the simulations, which are reported in Table 1. First, when the number of pre-treatment periods increases or the number of related outcomes increases, the bias decreases and becomes closer to the expected value for all estimators. Similar patterns are observed for the standard deviation of the estimators. Furthermore, the results clearly show that matching on more pre-treatment variables alleviates overfitting and improves out-of-sample prediction. Note that overfitting is a finite-sample problem due to erroneously matching on transitory shocks as well as the outcome variables and should be distinguished from the perfect fit condition (3) which is assumed to hold for a large T_0 and K.

Second, when the support of the predictors for the treated unit is the same with that for the control units ($d = 1$), demeaning substantially improves estimation in terms of both bias and standard deviation, as the conventional single-outcome SC method is likely

[10] Note that the terms from the bias decomposition in the proof of Proposition 1 are all close to 0, except the post-treatment transitory shock, which follows a standard normal distribution in our simulation.

to perform poorly when the treated unit is far from the convex hull of the control units, whereas demeaning adjusts for the differences in the levels and improves the pre-treatment fit. Meanwhile, the rejection rate of the 10% test is close to the nominal size, with or without demeaning in this case, since the RMSPE ratio for the treated unit is not conditional on a good pre-treatment fit. When d is smaller, the probability of obtaining a good pre-treatment fit increases for the treated unit while staying unchanged for the other units. As a result, the bias and standard deviation for both the conventional SC estimator and the demeaned SC estimator are smaller, and the improvement in estimation by demeaning the outcomes becomes less pronounced. In contrast, the distortion in the size of the test increases drastically, and demeaning alleviates the size distortion by improving the pre-treatment fits for all units.

Third, a larger number of pre-treatment periods or related outcomes also reduces the size distortion, since the pre-treatment RMSPE for the treated unit is less likely to be very close to 0. Overall, the results show that the multiple-outcome SC method outperforms the conventional single-outcome method, when there are multiple related outcomes with stable differences in the level of the outcomes.

Table 1: Simulation

		Conventional SC				Multi-Outcome SC ($K=1$)				Multi-Outcome SC ($K=3$)				Multi-Outcome SC ($K=10$)			
d	T_0	Pre-fit	Bias	SD	Rej.	Pre-fit	Bias	SD	Rej.	Pre-fit	Bias	SD	Rej.	Pre-fit	Bias	SD	Rej.
1	5	1.65	1.94	2.91	0.10	0.51	1.43	1.81	0.10	0.82	1.32	1.67	0.10	0.99	1.22	1.54	0.10
1	10	1.63	1.64	2.47	0.10	0.83	1.27	1.61	0.10	1.04	1.19	1.50	0.10	1.14	1.12	1.40	0.10
1	20	1.62	1.52	2.36	0.10	1.03	1.18	1.49	0.10	1.15	1.11	1.41	0.10	1.20	1.08	1.36	0.10
0.5	5	0.44	1.10	1.40	0.36	0.23	1.16	1.47	0.32	0.56	1.08	1.36	0.15	0.77	1.01	1.26	0.12
0.5	10	0.71	1.03	1.29	0.24	0.54	1.08	1.35	0.19	0.80	1.01	1.26	0.14	0.91	0.95	1.18	0.12
0.5	20	0.86	0.95	1.20	0.17	0.77	0.99	1.25	0.15	0.92	0.92	1.16	0.12	0.99	0.89	1.11	0.10
0	5	0.24	1.05	1.32	0.57	0.15	1.09	1.37	0.48	0.48	1.04	1.31	0.19	0.71	0.99	1.23	0.13
0	10	0.54	0.98	1.23	0.34	0.45	1.03	1.29	0.25	0.72	0.96	1.20	0.15	0.86	0.90	1.13	0.13
0	20	0.73	0.92	1.16	0.23	0.68	0.96	1.21	0.18	0.86	0.90	1.13	0.14	0.93	0.87	1.09	0.12

Note: This table compares the pre-treatment fit, average absolute bias, standard deviation, and rejection rate of the 10% test for the single-outcome SC estimator, and the multiple-outcome SC estimators constructed using 1, 3 and 10 demeaned outcomes respectively, with varying d and T_0, based on 5000 simulations for each setting.

4 Empirical Application

In order to curb the spread of COVID-19, most countries in Europe had implemented strict non-pharmaceutical interventions (NPIs) by late March, 2020. As an exception, Sweden opted against a general lockdown and implemented much lighter NPIs. For example, social distancing and working from home were advised but not mandated, bars, restaurants, and schools for children under 16 were kept open, quarantines for infected cases were not enforced, and facemasks were not recommended outside health care (Ludvigsson, 2020). Building upon the empirical strategy in Born et al. (2020) and Cho (2020), we exploit this natural experiment to estimate the impacts of the NPIs on various public health, labor market and economic outcomes, using the multiple-outcome SC method. To minimize the risk of structural breaks, we start the pre-treatment periods from 2019. And to examine the dynamics of the impacts of the NPIs over time, we estimate the treatment effects in the first 3 quarters of 2020, before the second wave of COVID-19 cases in Europe that started from October.[11]

4.1 Data

There are 26 countries in the treatment group for Sweden, including the other European Union members (excluding Cyprus, Luxembourg, and Malta due to their small sizes) as well as Norway, Switzerland and the United Kingdom. The NPIs implemented in each country usually consist of a bundle of individual policies with varying duration and magnitudes. To compare the strictness of the NPIs across countries, we employ the Government Stringency Index, which is a composite measure of the strictness of government responses based on several individual metrics, and ranges from 0 to 100, with 100 representing the strictest response (Hale et al., 2020).

Figure 2 depicts the stringency index for each country in our sample in the first 3 quarters of 2020, with Sweden shown in black and the others in gray. We see that most countries rapidly tightened their intervention policies in March, and kept them in place

[11] This section presents the main results of the empirical application. More details about the data, the estimation and inference procedures, and the additional results can be found in the supplementary appendix.

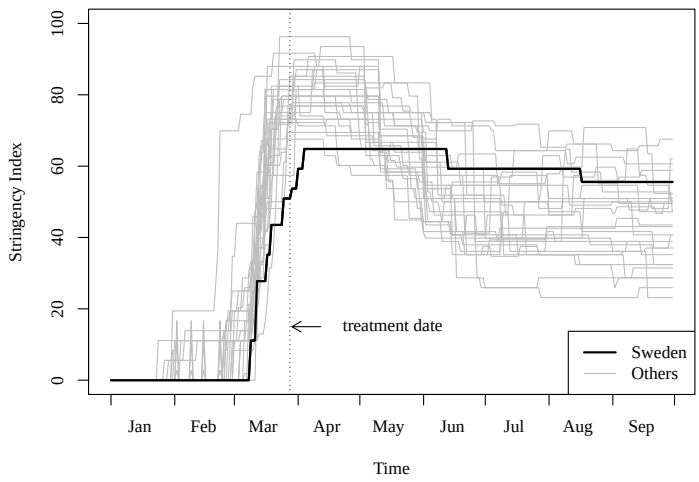


Figure 2: Stringency Index

until May, when the strict NPIs began to be gradually eased. The stringency index of Sweden almost always stayed as the lowest in the sample during this period, consistent with the earlier observation that Sweden implemented much lighter NPIs than the other European countries. To derive the binary treatment status from the stringency index, we pick the average date that the stringency index peaked for each country in the treatment group, March 28th, as the treatment date, which is denoted by the vertical dotted line in Figure 2.[12]

We are interested in the impacts of the NPIs in three domains, namely, public health, the labor market, and the economy. Specifically, we examine 3 public health outcomes: the cumulative numbers of COVID-19 infection cases and deaths, and the number of weekly deaths from all causes; 3 labor market outcomes: employment rate, absence from work and total hours worked among the employed; and 6 economic outcomes: GDP, import, export, industrial production, retail sales, and CPI. As the outcomes are observed in different

[12]One concern about this choice is that the NPIs have already started to be implemented prior to this date, albeit at lower levels compared with the peaks, and matching on outcomes that were realized after some NPIs have been in place may attenuate the estimates of the treatment effects. However, as we show in Section 4.4, the results of our analysis remain virtually the same even if we backdate the treatment to October 1, 2019, before the first case of COVID-19 in the world was reported.

frequencies, the numbers of periods in which they are observed are also different.[13]

Figure 3 visualizes the outcomes over time for all countries in our sample, with the trajectories for Sweden in black and the other countries in gray.[14]

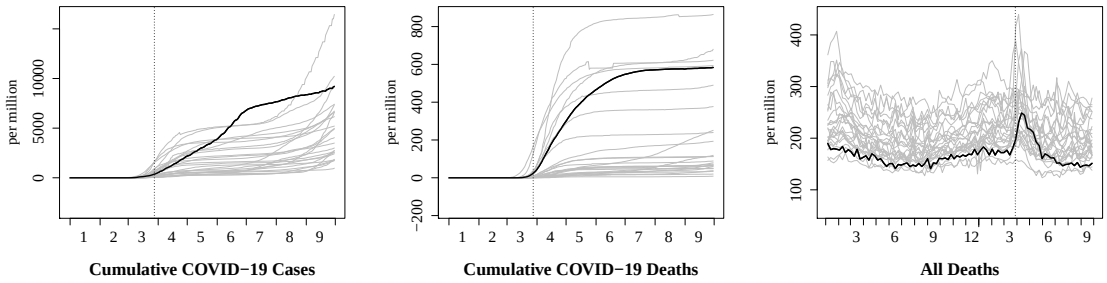
In the public health domain, we see that the numbers of COVID-19 cases and deaths began to rise quickly for countries in the treatment group from mid-March, before flattening out in early May. In comparison, although starting at lower levels and rates, the numbers of COVID-19 cases and deaths in Sweden continued growing to become among the highest in the sample, before slowing down only from June. Similarly, the number of weekly deaths from all causes in Sweden stayed near the bottom throughout 2019, but came close to the sample mean with a sharp spike in April, despite rises in other countries as well in this period, and only fell back to its usual level after June.

The labor market outcomes present a more varied picture. Due to various protective measures to contain employment losses, the employment rate only experienced modest dips in the second and third quarters of 2020 for all countries in the sample. However, there were much more visible changes in absence from work and total hours worked among the employed. Compared with the 2019 levels, the percentage of the employed who were temporarily absent from work more than doubled, and the total hours worked dropped by about 10-20% in the second quarter for countries in the treatment group, and quickly returned to their normal levels in the third quarter when the NPIs were relaxed. In contrast, there were only very mild changes in these two outcomes for Sweden.

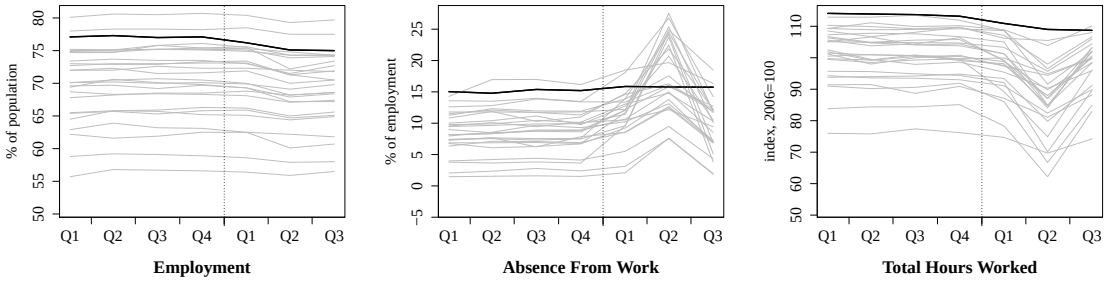
When it comes to the economic domain, all outcomes except CPI were adversely im-

[13] Although the frequencies are different for different outcomes, they are similar for outcomes in the same domain. Using outcomes of different frequencies is similar to using linear combinations of the outcomes as in Abadie et al. (2010), e.g., quarterly data can be considered as averages of monthly data. Since we have an abundance of pre-treatment variables to match on, we do not include any additional covariates. See Botosaru and Ferman (2019) for a discussion on the role of the covariates in the SC method.

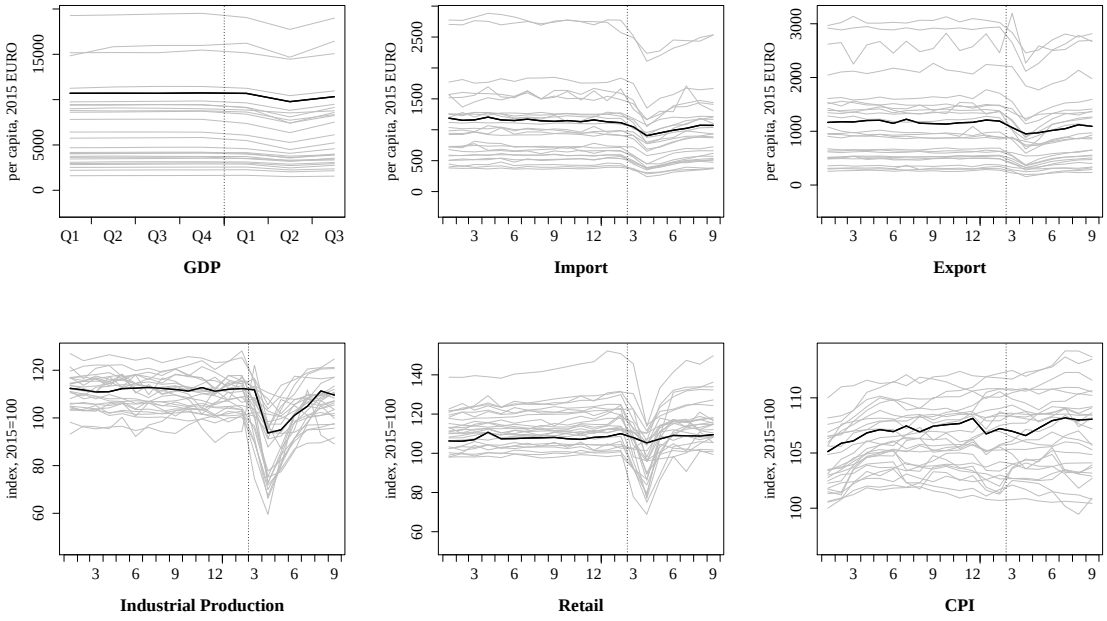
[14] Since the outcomes are observed in different frequencies, the vertical dotted lines are used as delimiters to visually separate the pre-treatment and posttreatment periods for the outcomes, and may not sit on the treatment date exactly. Specifically, we only include the daily and weekly outcomes observed on and before Mar 28th, 2020, the monthly outcomes observed before March 2020, and the quarterly outcomes observed in 2019, in the pre-treatment matching variables, none of which contain observations after the treatment date. The vertical dotted lines are positioned to reflect these choices.



(a) Public Health Outcomes



(b) Labor Market Outcomes



(c) Economic Outcomes

Figure 3: Descriptive Graphs

pacted by the pandemic in the second quarter and somewhat recovered in the third. The changes were relatively modest in GDP, import and export, and more dramatic in industrial production and retail sales for countries in the treatment group. The economic outcomes in Sweden experienced similar changes, with the noticeable exception that the drop in the volume of retail sales was much smaller.

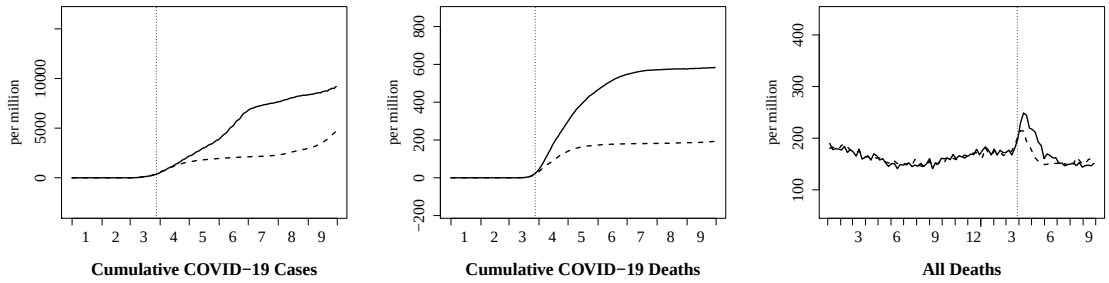
It is also worth noting that the relative position for Sweden differs across the outcomes, highlighting the necessity of adjusting for the differences in the level of the outcomes through demeaning.

4.2 Mutiple-Outcome Synthetic Control

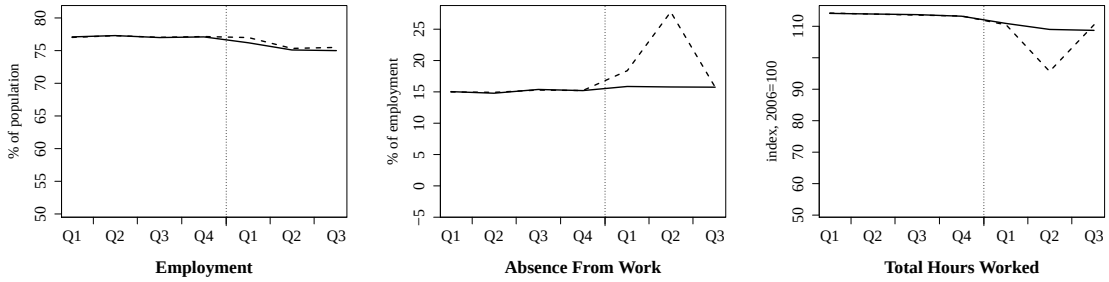
As outcomes in different domains may depend on different sets of predictors, we construct a synthetic Sweden in each of the three domains, using a convex combination of the countries in the comparison group, to closely approximate the dynamics of the outcomes in the domain for Sweden before the treatment.

Figure 4 compares the trajectories of each outcome for Sweden and the synthetic Sweden. In the left-most figure of Panel (a), the actual Sweden starts to accumulate more COVID-19 cases and deaths than the synthetic Sweden from April. The gaps between the two widen rapidly thereafter, before stabilising from July. Similarly, we observe more deaths from all causes in the actual Sweden than the synthetic Sweden from April to July, with the gap peaking in May. A back-of-the-envelope calculation suggests that had Sweden implemented stricter NPIs like the other European countries in March, the cumulative COVID-19 cases and deaths could have been reduced by about 5300 and 390 per million population respectively by July, amounting to 70% and 68% drops from the realized levels in Sweden, and there could have been 20% fewer weekly deaths from all causes in early May. For comparison, we also compute the cumulative deaths from all causes since April for both the synthetic Sweden and the actual Sweden, and find that the cumulative deaths from all causes could have been reduced by 364 per million population by July, representing an 11% drop from the realized level in Sweden.[15]

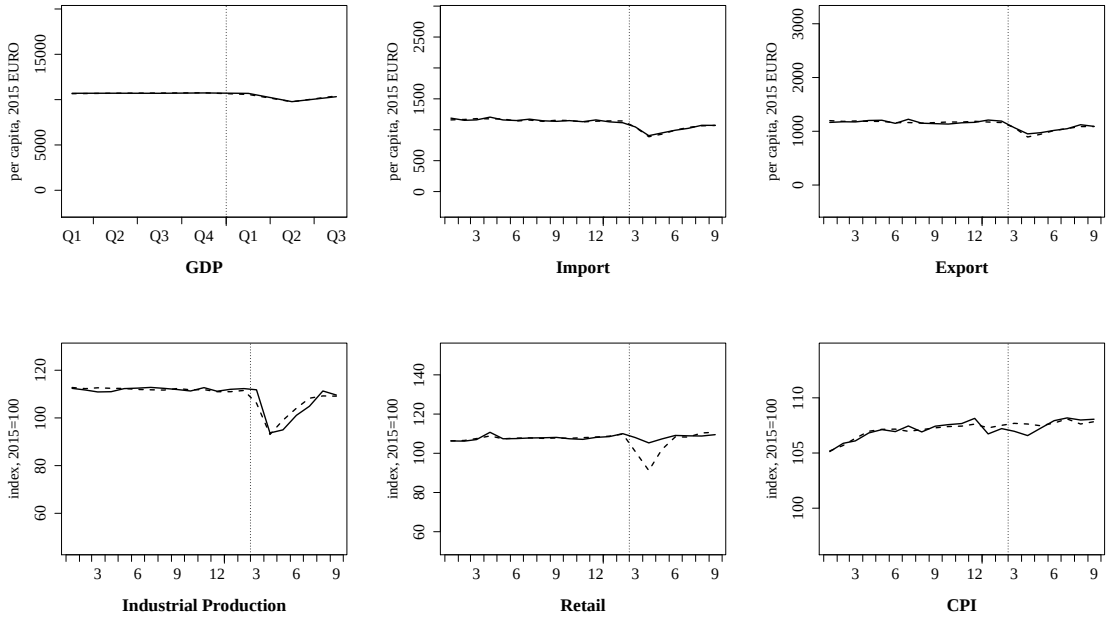
[15]Note that the percentage drop in deaths from all causes is much smaller than that in COVID-19 deaths, because the base number is much larger.



(a) Public Health Outcomes



(b) Labor Market Outcomes



(c) Economic Outcomes

Figure 4: Outcomes for Sweden (solid line) and the Synthetic Sweden (dashed line)

Since COVID-19 is more easily contracted by people with existing conditions or weaker immune systems, and NPIs may have reduced the transmission of other diseases by promoting good hygiene behaviors and reducing face-to-face interactions, the estimated reduction in deaths from all causes, which amounts to a difference of 64% compared with the realized number of COVID-19 deaths, may serve as a lower bound for the reduction in COVID-19 deaths.

As for the labor market outcomes in Panel (b), we find that stricter NPIs would increase absence from work among the employed by almost 76%, and reduce total hours worked by about 12%, in the second quarter of 2020. The impacts would disappear in the third quarter, and there would be no visible effect on the employment rate throughout the first 3 quarters in 2020.

In terms of the economy as shown in Panel (c), we find that stricter NPIs would shrink the volume of retail sales by 5%-13% from March to May, whereas the effects on the other economic outcomes including GDP, import, export, industrial production, and CPI, were all close to 0. This suggests that almost all of the contraction in the economy was due to the pandemic itself rather than the NPIs.[16]

4.3 Inference

Given the common perception that the implementation of NPIs would benefit public health by reducing the spread of the virus, while harming the labor market and the economy, we test the null hypothesis for outcome k at $t > T_0$, $H_0 : \tau_{1t,k} = 0$, against the alternative hypothesis, $H_1 : \tau_{1t,k} < 0$, so that if the null hypothesis is rejected, then we can conclude that the implementation of strict NPIs would have statistically significant negative effect on outcome k for Sweden at time t.[17] Following the one-sided inference procedure in Abadie (2021), we permute the treatment status among all the units, and compute the pre-treatment and post-treatment RMSPE's for unit i, outcome k in period t. Based

[16]Since we are interested in the treatment effects on multiple outcomes in particular domains, we may also follow Kling et al. (2007) and summarize the estimated treatment effects in domain $\mathbb{K}$ using the aggregate treatment effect. The results are included in the supplementary appendix.

[17]We reverse the sign of the treatment effect on absence from work, so that it has the same expected direction with the other outcomes.

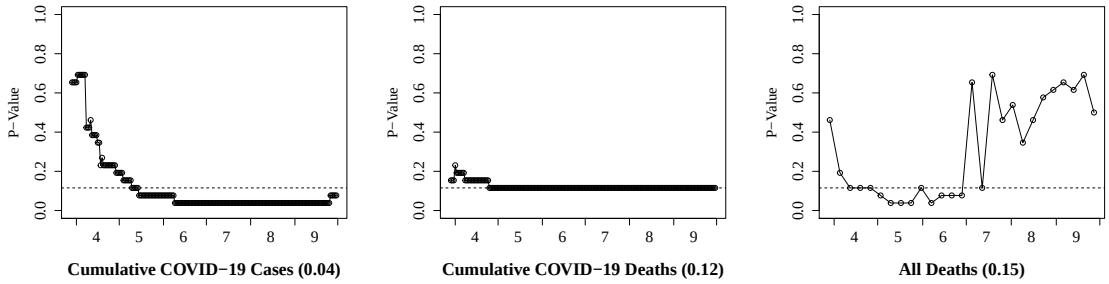
on the ranking of the post-to-pre-treatment RMSPE ratios, we can then compute the corresponding p-values.

Figure 5 shows the per-period p-values for each outcome, with the p-values in the posttreatment periods overall reported in the parentheses. The horizontal dotted line represents the significance level at $\alpha = 3/(J+1)$ (top 3 in the ranking of the post-to-pre-treatment RMSPE ratios in the sample), with $J = 25$ ($\alpha \approx 0.12$) in the public health and labor market domains, and $J = 26$ ($\alpha \approx 0.11$) in the economic domain. We find that the effects of the NPIs on COVID-19 cases and deaths were statistically significant at the 12% level from May, and the effect on deaths from all causes was significant from April to June. There were significant effects on absence from work and total hours worked in the second quarter, whereas the treatment effect on the employment rate remained insignificant throughout. As for the economic outcomes, we only find significant effect on retail sales in March, and we cannot reject the null hypotheses that the NPIs had no effects on the other economic outcomes in the first three quarters of 2020.

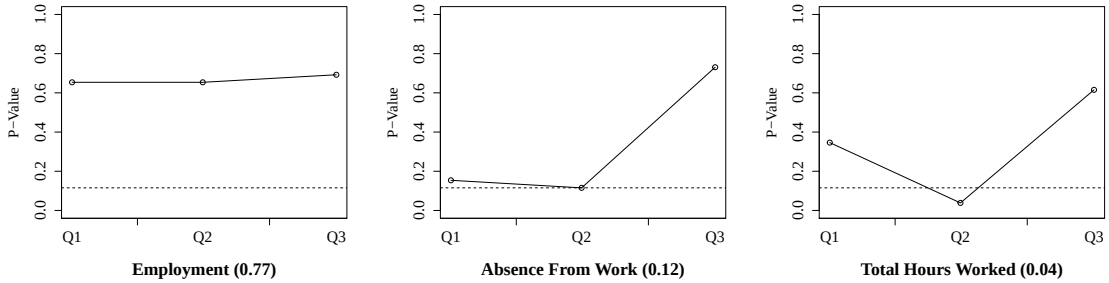
4.4 Robustness Checks

In our benchmark analysis, we pick March 28, 2020, the average date on which the stringency index peaked for each country in the treatment group, as the treatment date. As mentioned earlier, this choice may lead to attenuated estimates of the treatment effects if the outcomes were affected by the treatment before the chosen date either because countries had started implementing the NPIs before this date, or because individuals had anticipated the implementation of the NPIs and had acted in advance.

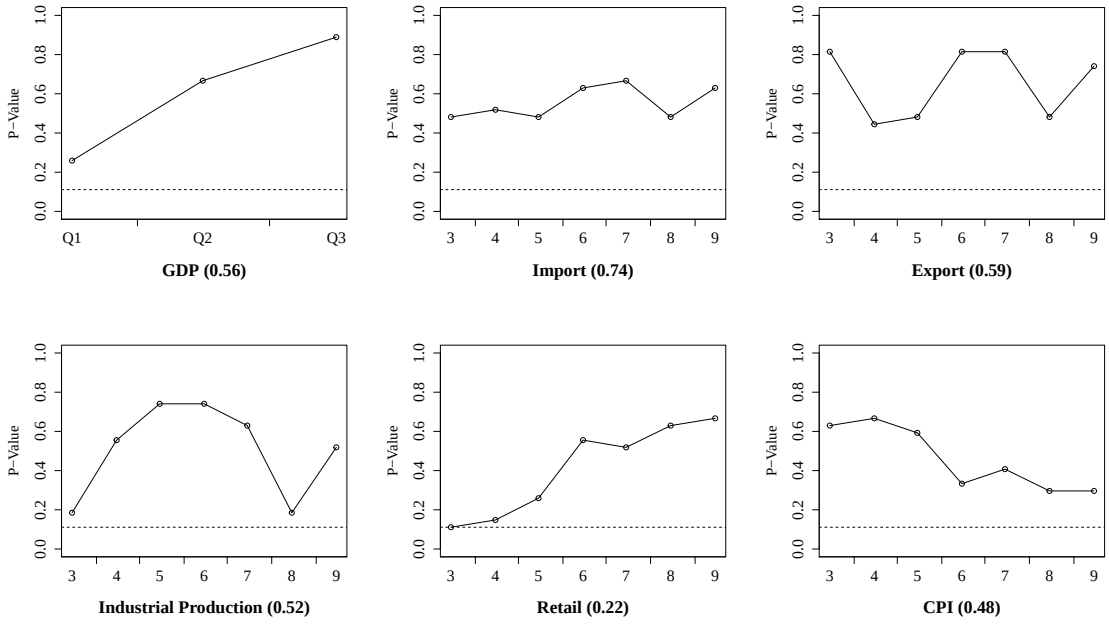
We thereby backdate the treatment to October 1, 2019, to see whether we can obtain results that are similar to those in the benchmark specification. This alternative date was before the first case of COVID-19 infection in the world was detected, and almost 4 months before the first case was reported in Europe, so that we can be confident that the outcomes observed before this date were by no means contaminated. In addition, the backdating exercise is also useful to evaluate the credibility of the SC estimator by assessing whether the synthetic Sweden can reproduce the outcomes of the actual Sweden in the absence of the treatment.



(a) Public Health Outcomes



(b) Labor Market Outcomes



(c) Economic Outcomes

Figure 5: P-Values of the One-sided Tests of H_0: strict NPIs would have no effect on the specific outcome at the corresponding posttreatment period

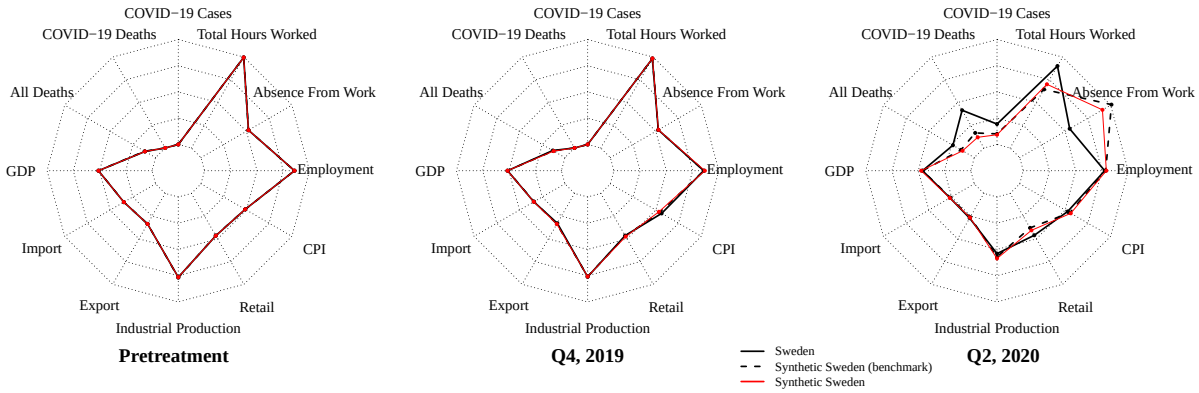


Figure 6: Radar Chart of Outcomes (Backdating)

To save space, we summarize the results succinctly using the radar charts.[18] Figure 6 compares the synthetic Sweden constructed by matching only on outcomes observed before the placebo treatment date, October 1, 2019, with the actual Sweden at three different stages. The left figure shows the outcomes averaged in the first three quarters of 2019, before the placebo treatment date, the middle figure shows the outcomes in the fourth quarter of 2019, which was after the placebo treatment date but before the actual treatment date, and the right figure shows the outcomes in the second quarter of 2020, after the actual treatment assignment.

The results show that the synthetic Sweden is very close to the actual Sweden not only before October 1, 2019, but after the placebo treatment and before the true treatment as well. The absence of estimated effects before the actual treatment date shows that the synthetic Sweden can reliably reproduce the untreated potential outcomes of the actual Sweden. The differences between the two emerges after the actual treatment date and the results are almost identical to our benchmark results. The close resemblance of our benchmark results and the results produced by backdating the treatment to almost 6 months earlier provides compelling evidence that our estimates for the treatment effects are reliable. Note that we would not have been able to do this had we been focusing on estimating the effect of the NPIs on COVID-19 infection cases or deaths using the conventional single-outcome SC method, like in Born et al. (2020) or Cho (2020), since COVID-19 cases and

[18]The full detailed graphs showing the outcomes for Sweden and the synthetic Sweden in all periods as well as the results for the other robustness checks can be found in the supplementary appendix.

deaths were not observed before this placebo treatment date.[19]

The treatment backdating exercise, along with the reanalysis exercise in the Supplementary Appendix, B.2, highlights one of the advantages of the multiple-outcome SC method over the conventional single-outcome SC method, i.e., our method allows the users to construct a reliable synthetic control when we only observe the outcomes in very few pre-treatment periods, or in the extreme case, when we do not observe the outcome of interest before the treatment at all, as long as the synthetic control can closely approximate the unit of interest in multiple related outcomes before the treatment.

We also conduct several additional robustness checks to assess the sensitivity of our results to changes in the design of the study, e.g., when we exclude any particular country or outcome, when we match on single outcomes rather than multiple related outcomes, and when we use the original outcomes rather than the demeaned outcomes. The results of these exercises, which are included in the supplementary appendix, are all similar to our benchmark results.

4.5 Further Analysis

In the supplementary appendix, we also conduct subgroup analysis when data permit, to investigate whether there is heterogeneity in the effects of the NPIs across different groups in the population or different industries in the economy.

The analysis on deaths from all causes in different gender and age groups suggests that the impact of the NPIs was larger for males in the 60-79 age group, and larger for females in the group with age 90 or over, possibly as a result of differences in lifestyles or risk preferences. The analysis on reasons for absence from work suggests that absence from work due to temporary layoffs would quintuple and absence from work due to other reasons would almost double in the second quarter of 2020, whereas absence from work due to holiday and illness would not be affected. The analysis on retail sales suggests that the reduction in retail sales due to the NPIs would be mainly through reduced sales of non-food products.

[19] Cho (2020) backdates the treatment by 3 days in one of the robustness checks, which may not be sufficient to address the concern that the pre-treatment outcome may have been affected by the treatment.

We also examine the effects of the NPIs on the labor market outcomes by gender, the effect on the employment rate by occupation, education and age, as well as the effects on the individual components of GDP and CPI. However, we do not find any heterogeneity in the treatment effects for these subgroups.

5 Conclusion

This paper generalizes the conventional single-outcome SC method to a multiple-outcome framework, where the number of pre-treatment periods is supplemented with the number of related outcomes, improving the SC methods in the situations when the number of available pre-treatment periods is small or when potential structural breaks impose limitations on the number of pre-treatment periods. We show that the bias of the multiple-outcome SC estimator diminishes with the number of pre-treatment periods and the number of related outcome variables. Our simulations show that the multiple-outcome SC estimator improves upon the conventional single-outcome SC estimator in terms of bias, standard deviation, and the size.

We investigate the effects of the non-pharmaceutical interventions (NPIs) on various outcomes in the public health, labor market, and economic domains using the multiple-outcome SC method, where we construct a synthetic Sweden in each domain using the other European countries that implemented much stricter NPIs. We find that the NPIs would significantly reduce the cumulative numbers of COVID-19 cases and deaths as well as deaths from all causes, increase temporary absence from work and reduce total hours worked among the employed, but would have limited impacts on the employment rate and the economy, other than shrinking the volume of retail sales in the early stage.

There are several limitations in our empirical analysis. First, the data on COVID-19 cases and deaths may suffer from reporting issues such as measurement errors or time lags, due to differences in the reporting standards in different countries and time constraints. This issue is alleviated to some extent in our analysis by matching on deaths from all causes, which is more accurately measured. Second, the complexities in the NPIs due to different timings and contents in different countries are not likely to be accurately or fully captured by the stringency index or the binary treatment indicator. It would be

worthwhile to conduct more detailed analysis on the effects of the individual policies, as in Castex et al. (2021) and Chernozhukov et al. (2021). Third, there may be spill-over effects. For example, risk-loving individuals from other European countries might have fled to Sweden, which would drive the number of COVID-19 cases up in Sweden, and cause the public health effects of the NPIs to be over-estimated. Note that this is different from the voluntary social distancing behaviors, which does not invalidate our results.

Finally, a few words may be said to applied researchers for whom our method may be useful. Just as Abadie et al. (2015) cautioning users of the original SC method that "we do not recommend using this method when the pre-treatment fit is poor or the number of pre-treatment periods is small", our method is best suited for applications where 1) there are enough pre-treatment variables to match on (i.e., there is either a large number of related outcomes or a large number of pre-treatment periods, or both), so overfitting is unlikely,[20] and 2) the synthetic control constructed using a few control units can closely approximate the pre-treatment variables of the treated unit, which is unlikely to happen with a large number of outcome variables if they do not share the common linear structure.[21]

Appendix A Technical Derivations and Proofs

Comparison of multi-outcome and single-outcome SC weights

For better understanding of the multiple-outcome SC, we compare the multiple-outcome SC weights with the single-outcome SC weights. Let $\boldsymbol{w}$ be the $J \times 1$ vector and $\boldsymbol{Y}^c_{t,k}$ be the $J \times 1$ vector of the outcome variable of the control units. Ignoring the facts that the elements of $\boldsymbol{w}$ sum to one and nonnegative, the conventional single-outcome SC weights

[20] The number of pre-treatment variables required to avoid overfitting depends on the size of the donor pool. In the usual cases where the number of control units (donors) is in dozens, preferably at least 10 pre-treatment variables are needed.

[21] There has not been an accepted rule of thumb to determine whether the pre-treatment fit is good other than relying on visual inspections. Gardeazabal and Vega-Bayo (2017) consider matches that have a pre-treatment mean absolute error smaller than 20% of the mean values in the outcome variable as good matches, but this may not apply to cases where the mean of the outcome is close to 0.

for the kth outcome are obtained by minimizing

$$\sum_{t=1}^{T_0}(Y_{1,t,k} - \boldsymbol{w}'\boldsymbol{Y}^c_{t,k})^2. \tag{A.1}$$

The solution $\widetilde{\boldsymbol{w}}_k$ is simply the least squares estimator:

$$\widetilde{\boldsymbol{w}}_k = \left(\sum_{t=1}^{T_0}\boldsymbol{Y}^c_{t,k}\boldsymbol{Y}^{c'}_{t,k}\right)^{-1}\sum_{t=1}^{T_0}\boldsymbol{Y}^c_{t,k}Y_{1,t,k}, \tag{A.2}$$

provided that $\sum_{t=1}^{T_0}\boldsymbol{Y}^c_{t,k}\boldsymbol{Y}^{c'}_{t,k}$ is invertible. For the multiple-outcome SC weights, we further stack $\boldsymbol{Y}^c_{t,k}$ over the K outcomes. Then the criterion function is

$$\sum_{k=1}^{K}\sum_{t=1}^{T_0}(Y_{1,t,k} - \boldsymbol{w}'\boldsymbol{Y}^c_{t,k})^2 \tag{A.3}$$

and the solution is

$$\widehat{\boldsymbol{w}} = \left(\sum_{k=1}^{K}\sum_{t=1}^{T_0}\boldsymbol{Y}^c_{t,k}\boldsymbol{Y}^{c'}_{t,k}\right)^{-1}\sum_{k=1}^{K}\sum_{t=1}^{T_0}\boldsymbol{Y}^c_{t,k}Y_{1,t,k}, \tag{A.4}$$

provided that $\sum_{k=1}^{K}\sum_{t=1}^{T_0}\boldsymbol{Y}^c_{t,k}\boldsymbol{Y}^{c'}_{t,k}$ is invertible. The actual implementation of the SC method restricts the weights to sum one and to be non-negative, so the solutions will be different from (A.2) and (A.4). Nonetheless, the least-squares estimators (A.2) and (A.4) show that $\widehat{\boldsymbol{w}}$ is not a simple average of $\widetilde{\boldsymbol{w}}_k$ across the K outcomes. In addition, the minimized criterion functions are exactly zero, so both $\widetilde{\boldsymbol{w}}$ and $\widehat{\boldsymbol{w}}$ give the optimal fit for each outcome variable.

Treatment effect on the untreated

In the discussion of the theoretical framework, we have been focusing on the setting with a single treated unit and many control units, where we estimate the treatment effects on the treated. However, there are cases with many treated units, and we may wish to estimate the treatment effects on the untreated by constructing a synthetic unit for the untreated unit using the treated units.

Without loss of generality, suppose that unit 1 remains untreated within the window of observation, while all the other units are treated from $t = T_0 + 1$ onwards. Recall

that the treated potential outcome is $Y^1_{it,k} = Y^0_{it,k} + \tau_{it,k}$. Since we have not imposed any assumption on the treatment effects except treating them as fixed given the sample, the treated potential outcomes may not have the interactive fixed effects functional forms or depend on the same predictors as the untreated potential outcomes. As a consequence, a synthetic unit that matches the untreated unit in the pre-treatment matching variables may not credibly reproduce its counterfactual outcomes after the treatment, even if it is similar to the untreated unit in the underlying predictors of the untreated potential outcomes. Therefore, in order to estimate the treatment effects on the untreated, we assume that the treatment effects are determined by the same predictors of the untreated potential outcomes, that is[22]

$$\tau_{it,k} = \alpha_{t,k} + \boldsymbol{Z}_i'\boldsymbol{\beta}_{t,k} + \boldsymbol{\mu}_i'\boldsymbol{\gamma}_{t,k}, \quad \forall i, t, k, \tag{A.5}$$

where $\alpha_{t,k}$ is the time trend in $\tau_{it,k}$, and $\boldsymbol{\beta}_{t,k}$ and $\boldsymbol{\gamma}_{t,k}$ are the outcome-specific coefficients of the observed and unobserved predictors respectively.

The multiple-outcome synthetic control estimator for $\tau_{1t,k}$, the treatment effect on the untreated unit, can then be constructed as

$$\check{\tau}_{1t,k} = \sum_{j=2}^{J+1} \hat{w}_j Y_{jt,k} - Y_{1t,k}. \tag{A.6}$$

Note that Eq. (A.5), together (1) and (2), implies that the treated potential outcome k for unit i at time t is given by

$$Y^1_{it,k} = (\delta_{t,k} + \alpha_{t,k}) + \boldsymbol{Z}_i'\left(\boldsymbol{\theta}_{t,k} + \boldsymbol{\beta}_{t,k}\right) + \boldsymbol{\mu}_i'\left(\boldsymbol{\lambda}_{t,k} + \boldsymbol{\gamma}_{t,k}\right) + \varepsilon_{it,k}. \tag{A.7}$$

Since the treated potential outcome $Y^1_{it,k}$ has an interactive fixed effects structure, we can similarly show that the bias of $\check{\tau}_{1t,k}$ vanishes as the number of pre-treatment periods or related outcomes grows.

Corollary 2. *Suppose that J is fixed, whereas T_0 and K are increasing, the demeaned outcomes are used, (A.5) holds, and the technical conditions are satisfied, then for any*

[22]Note, this assumption is more general than assuming that the treatment effects are constant. A similar assumption is discussed in Athey et al. (2021), where the treatment effect is assumed to have a low-rank pattern.

$t > T_0$,

$$|\mathbb{E}(\check{\tau}_{1t}) - \tau_{1t,k}| = O\left(\frac{1}{\sqrt{KT_0}}\right).$$

Technical conditions

We impose the technical conditions similar to those in Abadie et al. (2010) and Botosaru and Ferman (2019).

Condition 1. *Transitory shocks*

1) $\varepsilon_{it,k}$ *are independent across* i, t, k;[23]

2) $\mathbb{E}(\varepsilon_{it,k} \mid \boldsymbol{Z}_j, \boldsymbol{\mu}_j) = 0$ *for all* i, j, t *and* k;

3) $\mathbb{E}|\varepsilon_{jt,k}|^p < \infty$ *for all* $j = 2, \ldots, J+1$, $t \leq T_0$, $k \in \mathbb{K}$ *and some even integer* $p \geq 2$.

Condition 2. *The smallest eigenvalue of* $\frac{1}{KT_0}\sum_{k=1}^{K}\sum_{t=1}^{T_0}\boldsymbol{\lambda}_{t,k}\boldsymbol{\lambda}_{t,k}'$ *is bounded from below by some positive number* $\underline{\xi}$.

Proofs

Proof of Proposition 1. To provide a unified framework for the biases of the multiple-outcome synthetic control estimator and the single-outcome synthetic control estimator, without loss of generality, we can write

$$\tilde{w}_j^{(k)} = \hat{w}_j + \tilde{v}_j^{(k)}, \ j = 2, \ldots, J+1, \tag{A.8}$$

where $-1 \leq \tilde{v}_j^{(k)} \leq 1$. We emphasize that both sets of weights provide perfect fit for the pre-treatment values of the kth outcome and the observed predictors.

[23] Note that despite the i.i.d. condition on $\varepsilon_{it,k}$, the unobserved interactive fixed effects may account for the correlations along the corresponding dimensions. Technically, this condition is required so that the $B_{2t,k}$ and $B_{3t,k}$ terms in the proof have zero mean. In practice, the outcomes may have different scales or volatilities so that the transitory shocks may be clustered at the outcome level. This complexity is dealt with by standardizing each outcome in each period before matching in the implementation.

Under the restrictions $\sum_{j=2}^{J+1} \hat{w}_j \boldsymbol{Z}_j = \boldsymbol{Z}_1$, the bias of the multiple-outcome synthetic control estimator for outcome k and $t > T_0$ is the expectation of

$$Y_{1t,k}^0 - \sum_{j=2}^{J+1} \hat{w}_j Y_{jt,k} = \left(\boldsymbol{\mu}_1 - \sum_{j=2}^{J+1} \hat{w}_j \boldsymbol{\mu}_j \right)' \boldsymbol{\lambda}_{t,k} + \varepsilon_{1t,k} - \sum_{j=2}^{J+1} \hat{w}_j \varepsilon_{jt,k}, \tag{A.9}$$

and the bias of the single-outcome synthetic control estimator is the expectation of

$$\begin{aligned} Y_{1t,k}^0 - \sum_{j=2}^{J+1} \tilde{w}_j^{(k)} Y_{jt,k} &= \left(\boldsymbol{\mu}_1 - \sum_{j=2}^{J+1} \tilde{w}_j^{(k)} \boldsymbol{\mu}_j \right)' \boldsymbol{\lambda}_{t,k} + \varepsilon_{1t,k} - \sum_{j=2}^{J+1} \tilde{w}_j^{(k)} \varepsilon_{jt,k} \\ &= \left(\boldsymbol{\mu}_1 - \sum_{j=2}^{J+1} \hat{w}_j \boldsymbol{\mu}_j \right)' \boldsymbol{\lambda}_{t,k} - \sum_{j=2}^{J+1} \tilde{v}_j^{(k)} \boldsymbol{\mu}_j' \boldsymbol{\lambda}_{t,k} + \varepsilon_{1t,k} - \sum_{j=2}^{J+1} \tilde{w}_j^{(k)} \varepsilon_{jt,k}. \end{aligned} \tag{A.10}$$

We can stack observations over the pre-treatment periods to replace the first term on the RHS of (A.9) and the first two terms on the RHS of (A.10) with functions of the coefficients and the individual transitory shocks from the stacked expressions. Stacking the pre-treatment outcomes $Y_{it,k}$ over the T_0 pre-treatment periods, we have

$$\boldsymbol{Y}_{i,k} = \boldsymbol{\delta}_k + \boldsymbol{\theta}_k \boldsymbol{Z}_i + \boldsymbol{\lambda}_k \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_{i,k}, \tag{A.11}$$

where $\boldsymbol{Y}_{i,k}$, $\boldsymbol{\delta}_k$, and $\boldsymbol{\varepsilon}_{i,k}$ are $T_0 \times 1$, and $\boldsymbol{\theta}_k$ and $\boldsymbol{\lambda}_k$ are $T_0 \times r$ and $T_0 \times F$, respectively.

Since $\sum_{j=2}^{J+1} \tilde{w}_j^{(k)} \boldsymbol{Z}_j = \boldsymbol{Z}_1$, the restrictions $\sum_{j=2}^{J+1} \tilde{w}_j^{(k)} \boldsymbol{Y}_{j,k} = \boldsymbol{Y}_{1,k}$ can be simplified to

$$\boldsymbol{\lambda}_k \left(\boldsymbol{\mu}_1 - \sum_{j=2}^{J+1} \tilde{w}_j^{(k)} \boldsymbol{\mu}_j \right) = \sum_{j=2}^{J+1} \tilde{w}_j^{(k)} \boldsymbol{\varepsilon}_{j,k} - \boldsymbol{\varepsilon}_{1,k}. \tag{A.12}$$

Similarly, using the multiple-outcome synthetic control weights, we have

$$\boldsymbol{\lambda}_k \left(\boldsymbol{\mu}_1 - \sum_{j=2}^{J+1} \hat{w}_j \boldsymbol{\mu}_j \right) = \sum_{j=2}^{J+1} \hat{w}_j \boldsymbol{\varepsilon}_{j,k} - \boldsymbol{\varepsilon}_{1,k}, \tag{A.13}$$

which further simplifies (A.12) to

$$-\boldsymbol{\lambda}_k \sum_{j=2}^{J+1} \tilde{v}_j^{(k)} \boldsymbol{\mu}_j = \sum_{j=2}^{J+1} \tilde{v}_j^{(k)} \boldsymbol{\varepsilon}_{j,k}. \tag{A.14}$$

Since the K outcomes are determined by the same set of predictors in our multiple-outcome framework, we can further stack (A.11) over the K outcomes to get

$$\boldsymbol{Y}_i = \boldsymbol{\delta} + \boldsymbol{\theta} \boldsymbol{Z}_i + \boldsymbol{\lambda} \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i, \tag{A.15}$$

where $\boldsymbol{Y}_i$, $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}_i$ are $KT_0 \times 1$, and $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ are $KT_0 \times r$ and $KT_0 \times f$, respectively. And the restrictions $\sum_{j=2}^{J+1} \hat{w}_j \boldsymbol{Y}_j = \boldsymbol{Y}_1$ can be simplified to

$$\boldsymbol{\lambda}\left(\boldsymbol{\mu}_1 - \sum_{j=2}^{J+1} \hat{w}_j \boldsymbol{\mu}_j\right) = \sum_{j=2}^{J+1} \hat{w}_j \boldsymbol{\varepsilon}_j - \boldsymbol{\varepsilon}_1. \tag{A.16}$$

Condition 2 states that the $f \times f$ matrix $\boldsymbol{\lambda}'\boldsymbol{\lambda}$ has full rank, thus pre-multiplying $(\boldsymbol{\lambda}'\boldsymbol{\lambda})^{-1}\boldsymbol{\lambda}'$ on both sides of (A.16), we have

$$\left(\boldsymbol{\mu}_1 - \sum_{j=2}^{J+1} \hat{w}_j \boldsymbol{\mu}_j\right) = (\boldsymbol{\lambda}'\boldsymbol{\lambda})^{-1}\boldsymbol{\lambda}'\left(\sum_{j=2}^{J+1} \hat{w}_j \boldsymbol{\varepsilon}_j - \boldsymbol{\varepsilon}_1\right), \tag{A.17}$$

so that (A.9) can be written as

$$\begin{aligned} Y_{1t,k}^0 - \sum_{j=2}^{J+1} \hat{w}_j Y_{jt,k} =& \boldsymbol{\lambda}_{t,k}'(\boldsymbol{\lambda}'\boldsymbol{\lambda})^{-1}\boldsymbol{\lambda}' \sum_{j=2}^{J+1} \hat{w}_j \boldsymbol{\varepsilon}_j & (B_{1t,k}) \\ & - \boldsymbol{\lambda}_{t,k}'(\boldsymbol{\lambda}'\boldsymbol{\lambda})^{-1}\boldsymbol{\lambda}'\boldsymbol{\varepsilon}_1 + \varepsilon_{1t,k} - \sum_{j=2}^{J+1} \hat{w}_j \varepsilon_{jt,k}. & (B_{2t,k}) \end{aligned}$$

Similarly, pre-multiplying $\boldsymbol{\lambda}_{t,k}'(\boldsymbol{\lambda}_k'\boldsymbol{\lambda}_k)^{-1}\boldsymbol{\lambda}_k'$ to both sides of (A.14) gives

$$-\boldsymbol{\lambda}_{t,k}' \sum_{j=2}^{J+1} \tilde{v}_j^{(k)} \boldsymbol{\mu}_j = \boldsymbol{\lambda}_{t,k}'(\boldsymbol{\lambda}_k'\boldsymbol{\lambda}_k)^{-1}\boldsymbol{\lambda}_k' \sum_{j=2}^{J+1} \tilde{v}_j^{(k)} \boldsymbol{\varepsilon}_{j,k}. \tag{A.18}$$

Replacing (A.17) and (A.18) into (A.10), we have

$$\begin{aligned} Y_{1t,k}^0 - \sum_{j=2}^{J+1} \tilde{w}_j^{(k)} Y_{jt,k} =& \boldsymbol{\lambda}_{t,k}'(\boldsymbol{\lambda}'\boldsymbol{\lambda})^{-1}\boldsymbol{\lambda}' \sum_{j=2}^{J+1} \hat{w}_j \boldsymbol{\varepsilon}_j & (B_{1t,k}) \\ & - \boldsymbol{\lambda}_{t,k}'(\boldsymbol{\lambda}'\boldsymbol{\lambda})^{-1}\boldsymbol{\lambda}'\boldsymbol{\varepsilon}_1 + \varepsilon_{1t,k} - \sum_{j=2}^{J+1} \hat{w}_j \varepsilon_{jt,k} & (B_{2t,k}) \\ & - \sum_{j=2}^{J+1} \tilde{v}_j^{(k)} \varepsilon_{jt,k} & (B_{3t,k}) \\ & + \boldsymbol{\lambda}_{t,k}'(\boldsymbol{\lambda}_k'\boldsymbol{\lambda}_k)^{-1}\boldsymbol{\lambda}_k' \sum_{j=2}^{J+1} \tilde{v}_j^{(k)} \boldsymbol{\varepsilon}_{j,k}. & (B_{4t,k}) \end{aligned}$$

Given Condition 1 and since the synthetic control weights are independent of the observations for $t > T_0$, $B_{2t,k}$ and $B_{3t,k}$ have zero mean, whereas $B_{1t,k}$ and $B_{4t,k}$ do not because $\hat{w}_j$ and $\tilde{w}_j^{(k)}$ are functions of $\boldsymbol{\varepsilon}_j$ (Botosaru and Ferman, 2019).

It is shown that $\mathbb{E}\left|B_{4t,k}\right| = O\left(\frac{1}{\sqrt{T_0}}\right)$ in Abadie et al. (2010). Following closely the proof in Appendix B of Abadie et al. (2010), we show that $\mathbb{E}\left|B_{1t,k}\right| = O\left(\frac{1}{\sqrt{KT_0}}\right)$. We can rewrite $B_{1t,k}$ as

$$B_{1t,k} = \sum_{j=2}^{J+1} \hat{w}_j \sum_{q=1}^{K} \sum_{s=1}^{T_0} \boldsymbol{\lambda}'_{t,k} \left(\sum_{l=1}^{K} \sum_{n=1}^{T_0} \boldsymbol{\lambda}_{n,l} \boldsymbol{\lambda}'_{n,l} \right)^{-1} \boldsymbol{\lambda}_{s,q} \varepsilon_{js,q}. \tag{A.19}$$

Let the largest element of $|\boldsymbol{\lambda}_{t,k}|$ for $t = 1, \ldots, T$ and $k = 1, \ldots, K$ be bounded from above by $\bar{\lambda}$. Under Condition 2 and using the Cauchy–Schwarz Inequality, we have

$$\begin{aligned}
&\left(\boldsymbol{\lambda}'_{t,k} \left(\sum_{l=1}^{K} \sum_{n=1}^{T_0} \boldsymbol{\lambda}_{n,l} \boldsymbol{\lambda}'_{n,l} \right)^{-1} \boldsymbol{\lambda}_{s,q} \right) \\
&\leq \left(\boldsymbol{\lambda}'_{t,k} \left(\sum_{l=1}^{K} \sum_{n=1}^{T_0} \boldsymbol{\lambda}_{n,l} \boldsymbol{\lambda}'_{n,l} \right)^{-1} \boldsymbol{\lambda}_{t,k} \right)^{\frac{1}{2}} \left(\boldsymbol{\lambda}'_{s,q} \left(\sum_{l=1}^{K} \sum_{n=1}^{T_0} \boldsymbol{\lambda}_{n,l} \boldsymbol{\lambda}'_{n,l} \right)^{-1} \boldsymbol{\lambda}_{s,q} \right)^{\frac{1}{2}} \\
&\leq \left(\frac{\bar{\lambda}^2 f}{KT_0 \underline{\xi}} \right).
\end{aligned}$$

Let $\bar{\varepsilon}_j = \sum_{q=1}^{K} \sum_{s=1}^{T_0} \boldsymbol{\lambda}'_{t,k} \left(\sum_{l=1}^{K} \sum_{n=1}^{T_0} \boldsymbol{\lambda}_{n,l} \boldsymbol{\lambda}'_{n,l} \right)^{-1} \boldsymbol{\lambda}_{s,q} \varepsilon_{js,q}$. Then by Hölder's Inequality and the norm monotonicity, we have

$$\left|B_{1t,k}\right| \leq \sum_{j=2}^{J+1} \hat{w}_j |\bar{\varepsilon}_j| \leq \left(\sum_{j=2}^{J+1} |\hat{w}_j|^q \right)^{1/q} \left(\sum_{j=2}^{J+1} |\bar{\varepsilon}_j|^p \right)^{1/p} \leq \left(\sum_{j=2}^{J+1} |\bar{\varepsilon}_j|^p \right)^{1/p},$$

with $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$.

Using Hölder's Inequality again, we have

$$\mathbb{E}\left[\sum_{j=2}^{J+1} |\bar{\varepsilon}_j| \right] \leq \mathbb{E}\left[\left(\sum_{j=2}^{J+1} |\bar{\varepsilon}_j|^p \right)^{1/p} \right] \leq \left(\mathbb{E}\left[\sum_{j=2}^{J+1} |\bar{\varepsilon}_j|^p \right] \right)^{1/p} = \left(\sum_{j=2}^{J+1} \mathbb{E}|\bar{\varepsilon}_j|^p \right)^{1/p}.$$

Then using Rosenthal's Inequality, we have

$$\mathbb{E}|\bar{\varepsilon}_j|^p \leq C(p) \left(\frac{\bar{\lambda}^2 f}{KT_0 \underline{\xi}} \right)^p \max \left\{ \sum_{q=1}^{K} \sum_{s=1}^{T_0} \mathbb{E}|\varepsilon_{js,q}|^p, \left(\sum_{q=1}^{K} \sum_{s=1}^{T_0} \mathbb{E}|\varepsilon_{js,q}|^2 \right)^{p/2} \right\},$$

where the constant $C(p) = \mathbb{E}(\phi - 1)^p$ with ϕ being a Poisson random variable with parameter 1.

Let $\bar{m}_p = \max_j \frac{1}{KT_0} \sum_{q=1}^{K} \sum_{s=1}^{T_0} \mathbb{E}|\varepsilon_{js,q}|^p$, then we have

$$\mathbb{E}|B_{1t,k}| \leq C\,(p)^{1/p} \left(\frac{\bar{\lambda}^2 f}{\underline{\xi}}\right) J^{1/p} \max\left\{\frac{\bar{m}_p^{1/p}}{(KT_0)^{1-1/p}}, \frac{\bar{m}_2^{1/2}}{(KT_0)^{1/2}}\right\}. \tag{A.20}$$

Therefore, $\mathbb{E}\,|B_{1t,k}| = O\left(\frac{1}{\sqrt{KT_0}}\right)$, and $\mathbb{E}\,(\widehat{\tau}_{1t,k} - \tau_{1t,k}) \to 0$ as $KT_0 \to \infty$, i.e., the bias of the multiple-outcome synthetic control estimator is bounded by a function that goes to zero when the number of outcomes in the domain or the pre-treatment periods goes to infinity.

We have shown that the bias of the conventional single-outcome synthetic control method is usually $O\left(\frac{1}{\sqrt{T_0}}\right)$, but if the single-outcome synthetic control weights coincide with the multiple-outcome synthetic control weights, in which case $\tilde{v}_j^{(k)} = 0$ and $\tilde{w}_j^{(k)} = \hat{w}_j$, then the order becomes $O\left(\frac{1}{\sqrt{KT_0}}\right)$.

□

Proof of Corollary 1. The bias for the demeaned synthetic control estimator is

$$\begin{aligned}
\widehat{\tau}_{1t,k}^{\mathrm{DM}} - \tau_{1t,k} &= \dot{Y}_{1t,k} - \sum_{j=2}^{J+1} \hat{w}_j \dot{Y}_{jt,k} - Y_{1t,k}^1 + Y_{1t,k}^0 \\
&= Y_{1t,k}^1 - \frac{1}{T_0} \sum_{s=1}^{T_0} Y_{1s,k} - \sum_{j=2}^{J+1} \hat{w}_j \dot{Y}_{jt,k}^0 - Y_{1t,k}^1 + Y_{1t,k}^0 \\
&= \dot{Y}_{1t,k}^0 - \sum_{j=2}^{J+1} \hat{w}_j \dot{Y}_{jt,k}^0.
\end{aligned}$$

Notice that the demeaned equation for $Y_{it,k}^0$ retains the interactive fixed effects structure:

$$\begin{aligned}
\dot{Y}_{it,k}^0 &= Y_{it,k}^0 - \frac{1}{T_0} \sum_{s=1}^{T_0} Y_{is,k}^0 \\
&= \dot{\delta}_{t,k} + \boldsymbol{Z}_i' \dot{\boldsymbol{\theta}}_{t,k} + \boldsymbol{\mu}_i' \dot{\boldsymbol{\lambda}}_{t,k} + \dot{\varepsilon}_{it,k}.
\end{aligned}$$

We can thus follow similar steps to show that $\mathbb{E}\left(\widehat{\tau}_{1t,k}^{\mathrm{DM}} - \tau_{1t,k}\right) \to 0$ as $KT_0 \to \infty$ under the demeaned outcomes, and conditions 1 and 2.

□

Proof of Corollary 2. The proof is similar to that of Proposition 1 and thus omitted. □

SUPPLEMENTARY MATERIALS

Supplementary appendix: The supplementary appendix contains a extended review of the literature on the effects of NPIs, data sources for the empirical application, more details of the estimation and inference procedures, and additional results. (.pdf file)

Datasets and R codes: Datasets and R codes used for the simulations in Section 3, and the empirical application in Section 4, and a readme file that describes the contents and instructs the reader on their use. (.zip file)

References

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2):391–425.

Abadie, A., Diamond, A., and Hainmueller, J. (2010). 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.

Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, 93(1):113–132.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, pages 1–15.

Billmeier, A. and Nannicini, T. (2013). Assessing economic liberalization episodes: A synthetic control approach. *Review of Economics and Statistics*, 95(3):983–1001.

Born, B., Dietrich, A., and Müller, G. J. (2020). Do lockdowns work? A counterfactual for Sweden. *Covid Economics*, 16:1–22.

Botosaru, I. and Ferman, B. (2019). On the role of covariates in the synthetic control method. *The Econometrics Journal*, 22(2):117–130.

Castex, G., Dechter, E., and Lorca, M. (2021). COVID-19: The impact of social distancing policies, cross-country analysis. *Economics of Disasters and Climate Change*, 5(1):135–159.

Cattaneo, M. D., Feng, Y., and Titiunik, R. (2021). Prediction intervals for synthetic control methods. *Journal of the American Statistical Association*, 116(536):1865–1880.

Cavallo, E., Galiani, S., Noy, I., and Pantano, J. (2013). Catastrophic natural disasters and economic growth. *Review of Economics and Statistics*, 95(5):1549–1561.

Chernozhukov, V., Kasahara, H., and Schrimpf, P. (2021). Causal impact of masks, policies, behavior on early COVID-19 pandemic in the U.S. *Journal of Econometrics*, 220(1):23 – 62. Themed Issue: Pandemic Econometrics / Covid Pandemics.

Cho, S.-W. S. (2020). Quantifying the impact of non-pharmaceutical interventions during the COVID-19 outbreak–the case of Sweden. *The Econometrics Journal*.

Doudchenko, N. and Imbens, G. W. (2017). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Working paper, National Bureau of Economic Research.

Ferman, B. (2021). On the properties of the synthetic control estimator with many periods and many controls. *Journal of the American Statistical Association*, 116(536):1764–1772.

Ferman, B. and Pinto, C. (2021). Synthetic controls with imperfect pretreatment fit. *Quantitative Economics*, 12(4):1197–1221.

Gardeazabal, J. and Vega-Bayo, A. (2017). An empirical comparison between the synthetic control method and hsiao et al.'s panel data approach to program evaluation. *Journal of Applied Econometrics*, 32(5):983–1002.

Hale, T., Angrist, N., Cameron-Blake, E., Hallas, L., Kira, B., Majumdar, S., Petherick, A., Phillips, T., Tatlow, H., and Webster, S. (2020). Oxford COVID-19 government response tracker.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119.

Klößner, S. and Pfeifer, G. (2018). Synthesizing cash for clunkers: Stabilizing the car market, hurting the environment? Working paper.

Ludvigsson, J. F. (2020). The first eight months of Sweden's COVID-19 strategy and the key actions and actors that were involved. *Acta Paediatrica*, 109(12):2459–2471.

OECD (2020). Country policy tracker: Employment and social measures, OECD, Paris.

Robbins, M. W., Saunders, J., and Kilmer, B. (2017). A framework for synthetic control methods with high-dimensional, micro-level data: evaluating a neighborhood-specific crime intervention. *Journal of the American Statistical Association*, 112(517):109–126.

Samartsidis, P., Seaman, S. R., Montagna, S., Charlett, A., Hickman, M., and Angelis, D. D. (2020). A bayesian multivariate factor analysis model for evaluating an intervention by using observational time series data on multiple outcomes. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 183(4):1437–1459.