

# Assessing the Effectiveness of Synthetic Control Methods in Climate Change Analysis

Jessica Cremonese

March 2023

École Normale Supérieure

Scuola Galileiana di Studi Superiori

Professor Marc Fleurbaey.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Synthetic Control Methods</b>	<b>4</b>
2.1	Setting the method . . . . .	4
2.2	Estimation . . . . .	5
2.2.1	Bias properties of SCM . . . . .	7
2.2.2	Inference and robustness checks in SCM . . . . .	7
2.3	A detour on methodological advantages . . . . .	8
2.4	Contextual and data requirements for a credible application . . . . .	9
2.4.1	Contextual requirements . . . . .	9
2.4.2	Data requirements . . . . .	10
<b>3</b>	<b>Effectiveness of SCM: challenges and remedies</b>	<b>10</b>
3.1	Kaul et al. (2015) on the inclusion of lagged dependent variables . . . . .	11
3.2	Klößner et al. (2018) on the uniqueness of weights when utilizing cross-validation techniques . . . . .	12
3.3	Kuosmanen et al. (2021) on the true nature of numerical instability . . . . .	13
3.4	Ferman et al. (2020) on specification searching opportunities . . . . .	14
3.5	Klößner and Pfeifer (2017) on using SCM as a forecasting technique . . . . .	15
<b>4</b>	<b>Applying SCM to the analysis of extreme natural events</b>	<b>16</b>
4.1	Cavallo et al. (2013) on catastrophic natural disasters and economic growth . . . . .	16
4.1.1	Strategy . . . . .	16
4.1.2	Data . . . . .	17
4.1.3	Results . . . . .	17
4.2	Barone, Mocetti (2014) on growth and institutions after earthquakes in Italy . . . . .	18
4.2.1	Strategy . . . . .	18
4.2.2	Data . . . . .	18
4.2.3	Results . . . . .	19
4.3	Coffman, Noy (2011) on measuring the long term impact of Hurricane Iniki . . . . .	19
4.3.1	Strategy . . . . .	20
4.3.2	Data . . . . .	20
4.3.3	Results . . . . .	20

<b>5</b>	<b>Applying SCM to evaluate environmental policy</b>	<b>20</b>
5.1	Cole, Elliott, Liu (2020) on the impact of lockdowns on Wuhan air pollution and health . . . . .	20
5.1.1	Strategy . . . . .	21
5.1.2	Data . . . . .	21
5.1.3	Results . . . . .	22
5.2	Steinkraus (2018) on the green paradox . . . . .	22
5.2.1	Strategy . . . . .	22
5.2.2	Data . . . . .	22
5.2.3	Results . . . . .	23
5.3	Andersson (2019) on the study of carbon taxes and CO <sub>2</sub> emis- sions . . . . .	23
5.3.1	Strategy . . . . .	23
5.3.2	Data . . . . .	24
5.3.3	Results . . . . .	24
5.4	Maamoun (2019) on empirical evidence of the success of the Kyoto Protocol . . . . .	24
5.4.1	Strategy . . . . .	25
5.4.2	Data . . . . .	25
5.4.3	Results . . . . .	26
5.5	He et al. (2022) on Free Trade Zones in China . . . . .	26
5.5.1	Strategy . . . . .	26
5.5.2	Data . . . . .	27
5.5.3	Results . . . . .	27
<b>6</b>	<b>Discussion and onclusion</b>	<b>28</b>
<b>7</b>	<b>References</b>	<b>31</b>

# 1 Introduction

Climate change poses a serious threat to our planet, with the potential for devastating consequences for societies, economies, and ecosystems. To mitigate the impacts of climate change and prepare for extreme natural events such as hurricanes, droughts, and floods, policymakers are increasingly turning to various policy measures. However, evaluating the effectiveness of these policy interventions is challenging, as there is often no clear counterfactual or comparison group to assess the impact of the policy.

In recent years, synthetic control methods (SCM) have emerged as a powerful tool to address this challenge. SCM involves constructing a synthetic control group that closely matches the characteristics of the treated unit or group, using a combination of pre-treatment data and a set of control units. This allows us to estimate the counterfactual impact of the policy intervention by comparing the treated unit to its synthetic counterpart.

SCM has been successfully applied in a range of fields, including economics, public policy, and epidemiology. Athey and Imbens (2017) describe it as “arguably the most important innovation in the policy evaluation literature in the last 15 years”. In this paper, I focus on the application of SCM to climate policy and extreme natural events analysis. Specifically, I examine how SCM can be used to evaluate the effectiveness of climate policies such as carbon taxes, international agreements, and emissions regulations. I also explore how SCM can be used to estimate the impact of extreme natural events on economic outcomes and human welfare, and to assess the effectiveness of adaptation measures.

This paper makes several contributions to the literature. First, I provide an overview of the principles and methods of SCM, and discuss its advantages and limitations compared to traditional evaluation methods. Second, I review the existing literature on the application of SCM to climate policy and extreme natural events analysis, highlighting the key findings and insights. Third, I present several case studies that illustrate the potential of SCM to inform climate policy and natural disaster management. Finally, I discuss the implications of our findings for policymakers and researchers, and identify areas for future research.

The paper is organized as follows. Section 2 introduces the synthetic control method and its estimation techniques, as well as the key requirements for a successful application. Section 3 explores the criticalities of the method and provides an update on recent improvements to the basic model. In Section 4, I illustrate the application of the method to the study of extreme natural events, while in Section 5, I showcase its use in evaluating environmental policies. Finally, Section 6 and 7 discuss and summarize the main findings

and implications of the paper.

## 2 Synthetic Control Methods

Synthetic Control Methods (SCM) were originally proposed by Abadie and Gardeazabal (2003) and Abadie et al. (2010) to estimate the effects of aggregate interventions. The key idea behind the method is that, when units are a few aggregate entities, a better counterfactual than using any single unit can be derived by computing a combination of the untreated units that closely resembles the treated one, i.e., a “synthetic control”. The selection of the “donor units” which contribute to the creation of a synthetic counterfactual is formalized with a datadriven procedure. Although the method was originally intended for samples with few units, it has been successfully applied in contexts with large samples, for instance in Acemoglu et al. (2016). Such a synthetic control unit is computed as a weighted average of all potential comparison units that best resemble the treated units. A good synthetic unit will resemble treated one not only in terms of the outcome variable (outer optimization), but also in terms of the pre-treatment predictors value (inner optimization).

In this section, I will introduce the method and explore its feasibility, data requirements and methodological issues. The main references are Abadie and Gardeazabal (2003) and Abadie, Diamond and Hainmueller (2010), which introduced the method in the literature, and Abadie (2021), which provides a useful guide to the application of SCM.

### 2.1 Setting the method

Suppose to have data for  $j = 1, \dots, J + 1$  units, and suppose that unit  $j = 1$  is the treated unit. The “donor pool” of untreated units which will contribute to the construction of a synthetic control for unit  $j = 1$  is then constituted by the remaining  $j = 2, \dots, J + 1$  units. Let us further assume that the data covers  $T$  periods, with periods up to  $T_0$  being the pre-intervention observations.

For each unit  $j$  at time  $t$  data is available for the outcome of interest  $Y_{jt}$ , and for a number  $k$  of predictors  $X_{1j}, \dots, X_{kj}$ . Define the  $k \times 1$  vectors  $\mathbf{X}_1, \dots, \mathbf{X}_{J+1}$  which contain values of the predictors for units  $j = 1, \dots, J + 1$ . Furthermore, let us define the  $k \times J$  matrix  $\mathbf{X}_0 = (\mathbf{X}_2, \dots, \mathbf{X}_{J+1})$  which collects values of the predictors for the untreated units. For each unit  $j$ , define the potential outcome without treatment as  $Y_{jt}^N$ . For the treated unit  $j = 1$ , define the potential response under the treatment as  $Y_{jt}^I$  in the post treatment period  $t > T_0$ . The effect of the intervention for the affected unit  $j = 1$  for

$t > T_0$  is:

$$\tau_{1t} = Y_{1t}^I - Y_{1t}^N \quad (1)$$

For the treated unit,  $Y_{1t}^I$  is observed so that  $Y_{1t} = Y_{1t}^I$ , but  $Y_{1t}^N$  is not. SCM provides a way to estimate  $Y_{1t}^N$  for  $t > T_0$ , i.e., how the outcome of interest would have been in the absence of treatment. Notice that  $\tau_{1t}$  is allowed to change over time.

## 2.2 Estimation

A downside of comparative case studies lies in the attempt to select the control units by informally arguing for an affinity between the treated and the untreated before the intervention. However, when using data from aggregate units such as countries or regions, it can be difficult to find a proper counterfactual. SCM offers a formal procedure to select and combine the comparison units in order to create a synthetic counterfactual scenario where unit  $j = 1$  was unaffected by treatment.

Define  $\mathbf{W} = w_2, \dots, w_{J+1}$  as a  $J \times 1$  vector of nonnegative weights that sum to one. The  $\mathbf{W}$  vector attributes a weight to each pre-treatment unit in the donor pool  $j = 2, \dots, J + 1$  and characterizes its contribution to the synthetic unit.

For a set of weights,  $\mathbf{W}$ , the estimators of  $Y_{1t}^N$  and  $\tau_{1t}$  are:

$$\hat{Y}_{1t}^N = \sum_{j=2}^{J+1} w_j Y_{jt} \quad (2)$$

$$\hat{\tau}_{1t} = Y_{1t} - \hat{Y}_{1t}^N \quad (3)$$

Nonnegative weights ensure a convex combination of the pre-treatment values of donor units, so that the resulting control can be interpreted as a weighted average of the control units with typically sparse weights. Furthermore, it ensures comparability of the outcome variable by giving the synthetic control outcome the same scale of the intervention unit. Abadie (2021) notes that, when using weights that sum to one, variables in the data should be rescaled to correct for differences in size between the units, for instance by using per capita GDP instead of level GDP. This correction is not needed if variables are already comparable, for instance in the case of prices. Allowing for weights outside of the  $[0, 1]$  interval may provide a more accurate synthetic control by placing negative emphasis on some donor units that are dissimilar to the treated one. However, negative unbounded weights may introduce extrapolation, where the assigned weights are used to extrapolate beyond

the observed range of data to estimate the effect of treatment. This can lead to biased estimates and reduced precision, and makes the interpretation of weights less straightforward.

The core of the SCM estimation lies in the definition of the weights. The approach in Abadie et al. (2021) is to choose weights  $w_2, \dots, w_{J+1}$  such that the synthetic unit resembles pre-treatment values of the treated unit in terms of the outcome variable. Given the nonnegative constant vector  $\mathbf{V} = (v_1, \dots, v_k)$  that represents the relative importance of the  $k$  predictors, the optimal weight vector  $\mathbf{W}^*$  is the one that, for some positive constants  $v_h$ , minimizes:

$$\|\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W}\| = \left( \sum_{h=1}^k v_h (X_{h1} - w_2 X_{h2} - \dots - w_{J+1} X_{hJ+1})^2 \right)^{1/2} \quad (4)$$

Subject to  $\sum_{j=2}^{J+1} w_j$  and  $w_j \geq 0$ . That is, minimizing the distance between the treated unit  $\mathbf{X}_1$  and the weighted combination of the control units  $\mathbf{X}_0 \mathbf{W}$ . The output  $\mathbf{W}^* = (w_2^*, \dots, w_{J+1}^*)'$  is used in the estimation of the treatment effect for the treated unit across  $t = T_0 + 1, \dots, T$  as:

$$\hat{\tau}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt} \quad (5)$$

Any possible choice of weights in  $\mathbf{V}$  produces a different set of optimal weights, which is effectively a different synthetic control  $\mathbf{W}(\mathbf{V}) = (w_2(\mathbf{V}), \dots, w_{J+1}(\mathbf{V}))'$ . For this reason the choice of  $\mathbf{V}$  is a key issue. An initial approach could be to divide the predictor weights equally across the  $k$  predictors included in the model. However, more elegant solutions are proposed by the authors.

The proposed solution is to choose  $\mathbf{V}$  such that the synthetic control  $\mathbf{W}(\mathbf{V})$  minimizes the mean squared prediction error (MSPE) of the synthetic control with respect to  $Y_{1t}^N$  over the pre treatment period  $\mathcal{T}_0 = 1, \dots, T_0$ :

$$\sum_{t \in \mathcal{T}} (Y_{1t} - w_2(\mathbf{V}) Y_{2t} - \dots - w_{J+1}(\mathbf{V}) Y_{J+1t})^2 \quad (6)$$

We could choose  $\mathbf{V}$  via out-of-sample validation, which requires substantial pre-treatment observations. This exploits the observed pre-treatment  $Y_{1t}^N$  to gauge the predictive power of the variables  $X_{1j}, \dots, X_{kj}$  and assign coherent weights  $\mathbf{V}$ . To use out-of-sample validation start by dividing the  $\mathcal{T}_0$  period in a *training* period and a *validation* period. The lengths of the two periods may depend on data availability and frequency of measurement

of outputs and variables. Then, for every value  $\mathbf{V}$ , compute the synthetic control weights on the training period and call them  $\tilde{w}_2(V), \dots, \tilde{w}_{J+1}(V)$ . The MSPE over the validation period will be:

$$\sum_{t=t_1+1}^{T_0} (Y_{1t} - \tilde{w}_2(V)Y_{2t} - \dots - \tilde{w}_{J+1}(V)Y_{J+1t})^2 \quad (7)$$

Then, compute  $\mathbf{V}^* \in \mathcal{V}$  such that MSPE is minimized over this training period, with  $\mathcal{V}$  being the set of all potential  $\mathbf{V}$ . Then, check the synthetic control's ability to emulate the treated unit behavior on the remaining validation period observations by using  $\mathcal{V}^*$  to compute  $\mathcal{W}^* = \mathcal{W}(V)^*$ .

For an analysis of the shortcomings of cross-validation as defined in this section, see subsection 2.3.

### 2.2.1 Bias properties of SCM

Appendix B of Abadie et al. (2010) provides an analysis of the bias properties of synthetic control estimators in the case of a linear factor model and a vector autoregressive (VAR) model. According to their findings, the estimator's bias in a factor model scenario can be constrained by a function that goes to zero as the amount of pre-treatment periods increases. In an VAR scenario, the synthetic control estimator is unbiased.

### 2.2.2 Inference and robustness checks in SCM

Usual inferential techniques that rely on regression standard errors would overstate significance levels when used in SCM applications. Abadie et al. (2010) propose inferential methods based on the derivation of permutation distributions (also termed falsification exercises, or randomization tests). Assuming only one treated unit and many donor units, one should proceed by reassigning the treatment to each unit in the donor pool to compute placebo effects. Then, pool together all the estimated effects and consider the effect of intervention on the treated unit as significant if it is extreme relative to the placebos. A concern with this procedure is that the placebo estimations may not possess as good a pre-treatment fit as the actually treated unit does. Abadie et al. (2010) propose a measure of goodness of fit based on the ratio of post-intervention quality of fit relative to pre-intervention quality of fit.

First, define the root mean squared prediction error (RMSPE) of the synthetic estimate. For  $0 \leq t_1 \leq t_2 \leq T$  and  $j = \{1, \dots, J+1\}$ :

$$R_j(t_1, t_2) = \left( \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} (Y_{jt} - \hat{Y}_{jt}^N)^2 \right)^{1/2} \quad (8)$$



With  $\hat{Y}_{jt}^N$  being the synthetic outcome for time  $t$  and unit  $j$  when unit  $j$  is considered treated and the remaining  $J$  units are used as donors. Then, compute the goodness of fit estimate as the ratio:

$$r_j = \frac{R_j(T_0 + 1, T)}{R_j(1, T_0)} \quad (9)$$

The permutation distribution of  $r_j$  may be used for inference. A  $p$ -value to be used for inference may be computed as follows:

$$p = \frac{1}{J+1} \sum_{j=1}^{J+1} I_+(r_j - r_1) \quad (10)$$

With  $I_+$  being an indicator function with output 1 if the argument is non-negative, zero otherwise. This  $p$ -value represents the likelihood to find a treatment effect which is at least the same magnitude as that of the treated unit, if treatment were randomly assigned across donor units. As stated in Abadie (2021), significance is evaluated with respect to a benchmark distribution for the assignment process, which Abadie et al. (2010) implement using the uniform case.

Typical robustness checks would make sure that results are not excessively dependent on few control units. Usual checks consist in excluding from the donor pool units that are exposed to conditions that might emulate treatment, or exclude the donors with the highest weights from the donor pool to test the stability of the results. To give an example of the former, consider the case where we are investigating the effect of local policy aimed at containing polluting emissions. In this case, we would want to test robustness by excluding from the donor pool any unit where similar policies have been implemented in the relevant time window.

In sections 4 and 5 I will present a number of relevant papers, with most of them applying the mentioned robustness checks and inferential techniques.

## 2.3 A detour on methodological advantages

SCM has advantages relative to other competing methods, and in this subsection, I will emphasize these features with respect to linear regression estimators.

*Transparency and goodness of fit.* While both synthetic control estimation and regression estimators are applied to panel data, SCM is more transparent relative to the discrepancy between the treated unit and the synthetic unit.

Furthermore, the goodness of fit of the synthetic control unit can be easily evaluated through analysis of the pre treatment period. SCM should not be used if the fit in the pre treatment period is unsatisfactory.

*Extrapolation prevention.* Another advantage of synthetic controls comes from the restrictions placed on the weights, which prevent extrapolation. Instead, regression weights may lie outside of the  $[0, 1]$  interval (for a comparison of the outcomes of SCM and regression see Abadie, Diamond and Hainmueller (2015)).

*Discourages specification searching.* Synthetic control weights can be computed before the observation of post-treatment outcomes. Therefore, a researcher can decide on the study design without knowing how it would affect the conclusions.

*Weight sparsity.* When estimating weights for the control group, regression estimators typically provide non-zero weights to all control units. Instead, sythetic control weights are sparse. The contribution of each donor unit is straightforward and allows for a geometric interpretation of the sparse weights: the sythetic unit represents a point that lies in the convex hull generated by the donor units with non zero weights. Note that with many treated units the weights are not necessarily unique nor that sparse. Abadie and L’Hour (2019) offer a penalized version of SCM that provides unique and sparse weights under some conditions.

## 2.4 Contextual and data requirements for a credible application

Most of the data and contextual requirements for successful application of SCM are applicable to any other comparative case study. Following Abadie (2021), I will start from contextual requirements, and eventually move on to data requirements.

### 2.4.1 Contextual requirements

The first observation is on the volatility of outcomes. Excessive random noise in the outcome variable increases the likelihood of over-fitting. Therefore it is advised to filter out such noise before implementation; otherwise, the effect of treatment may be drowned out. This concern does not come from common volatility between the units, but from unit-specific volatility.

The second observation pertains to the donor pool. For a unit to be a suitable candidate for the donor units, it should not be affected by or subject to the treatment provided to the unit of interest, and it should have common features with the treated unit (typically, units in a similar region and/or

context). This entails the elimination from the donor pool of any unit which may have suffered a shock to the outcome that would not have happened in the absence of treatment in unit of interest.

Third, a typical concern is anticipatory behavior by the units which would introduce bias in the SCM estimates. Depending on data availability, a solution would be to backdate the period that marks the start of treatment. Since SCM allows for differential exposure to treatment across time, the initial periods that are barely affected by treatment may show very small effects, whereas the subsequent periods will show larger effects.

Spillover effects are another point of concern that stems from the selection of an ideal donor unit. Usually, a donor unit is valid if it has common features with the treated one, while also being unaffected by the shock. As a consequence, donor units tend to come from the same regions or be exposed to the same context. Spillover effects are especially an issue if units come from the same geographical area. In order to understand the bias introduced by interference, the researcher may carry out and compare the estimates with and without the affected donor units. Moreover, due to the transparency of the synthetic counterfactual and the sparsity of weights, the researcher may reason as to the potential direction of the bias and account for that in their analysis.

#### **2.4.2 Data requirements**

For the method to accurately and credibly track the treated unit, there must be availability of a window of pre-treatment observations from all units. Using large periods of time may present the issue of structural breaks in the pre-treatment span, affecting the structural stability of the model. Therefore, we must be aware of how using too long a window may compromise the predictive ability of the synthetic unit.

Extensive post-treatment information is crucial for the evaluation of the effects, especially if they are expected to intensify or dissipate over time.

### **3 Effectiveness of SCM: challenges and remedies**

Similar to any other technique, the application of synthetic controls demands careful consideration of the proper data and contextual prerequisites that must be fulfilled, along with fulfilling the required conditions for its use to claim a causal explanation of outcomes. Various features are currently being debated and enhanced in the literature, while some areas remain unexplored.

For instance, an area that still needs to be investigated is the computational aspects of synthetic control methods, which is still relatively under-researched and holds significant potential for improvement.

### 3.1 Kaul et al. (2015) on the inclusion of lagged dependent variables

The predictors used in a synthetic control application can include lagged outcome variables and other economically relevant variables that have explanatory power for the dependent variable. Generally, the lagged outcomes tend to have greater predictive power than the covariates included in the model. Although the intuitive approach would be to include all pre-treatment lagged outcome variables and covariates, doing so could nullify the effect of the covariates for both nested and regression optimization approaches. It's important to be mindful of the potential bias arising from this when deciding what to include in the model.

To illustrate this, write predictor matrices as:

$$X_1 = \begin{pmatrix} C_1 \\ Z_1 \end{pmatrix}, \quad X_0 = \begin{pmatrix} C_0 \\ Z_0 \end{pmatrix} \quad (11)$$

Here,  $C_i$  represents covariates, and  $Z_i$  represents the dependent variable elements. Similarly, we can rewrite matrix  $V$  as:

$$V = \begin{pmatrix} V_C & 0 \\ 0 & V_Z \end{pmatrix} \quad (12)$$

So that  $V_C$  contains the covariate weights, and  $V_Z$  contains the lagged dependent variable weights.

For given predictor weights  $V$ , we can write the inner optimization to find  $W(V)^*$  as:

$$\min_W \sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)} \quad (13)$$

And the outer optimization to find the optimal  $V$  that minimizes the MSPE as:

$$\min_V (Z_1 - Z_0 W^*(V))' (Z_1 - Z_0 W^*(V)) \quad (14)$$

If we denote  $W^{**}$  as:

$$W^{**} := \arg \min_W (Z_1 - Z_0 W)' (Z_1 - Z_0 W) \quad (15)$$

Then, for any  $V$ , we would have:

$$(Z_1 - Z_0 W^{**})' (Z_1 - Z_0 W^{**}) \leq (Z_1 - Z_0 W^*(V))' (Z_1 - Z_0 W^*(V)) \quad (16)$$

However, if we choose  $V$  such that  $V_C$  is zero and  $V_Z$  is the identity matrix, we get:

$$W^*(V^*) = W^{**} \quad (17)$$

The implication is that the outer optimization would want null predictor weights even if the inner optimization would assign them positive weights. As a result, the synthetic control would end up ignoring the covariates. Consequently, even if positive  $V$  elements are reported, they are effectively ignored. Bias emerges from this dynamic if the covariates are relevant explanatory elements of the dependent. To solve for this, it's recommended to exclude some outcomes lags in the inner optimization so that positive weights emerge for other predictors.

The authors illustrate this by replicating a paper by Billmeier and Nannicini (2013), and highlight a stark difference in estimated treatment effects when using all pre-treatment dependent variables observation as opposed to not doing so.

This dynamic may be exploited to understand the predictive power of predictors via the comparison of a full lagged dependent variable model and a reduced lagged dependent variable one. The discrepancy between results is indicative of whether to include such covariates.

### **3.2 Klößner et al. (2018) on the uniqueness of weights when utilizing cross-validation techniques**

Some authors noticed numerical instability in the assigned predictor weights when replicating results from Abadie and Gardeazabal (2003). Klößner et al. (2018) investigate this ambiguity by looking at validation weights. When using the out-of-sample validation technique, predictor weights are not necessarily uniquely defined, so that when replicating Abadie et al. (2015) the authors find different yet equivalent solutions for the weights depending on the software package used (specifically, R with the SYNTH package versus STATA) and on the specific donor unit ordering (alphabetical versus custom).

Cross-validation requires the sectioning of the pre-treatment observations in a training period and validation period, and a two-step estimation procedure that computes predictor and donor weights. This is opposed to the standard one step estimation procedure of SCM and generates ambiguity on weights. The authors develop a rule of thumb to gauge the extent of this ambiguity in the estimates of the treatment effect that is centered on the difference  $k - \alpha$ , where  $k$  is the number of predictors and  $\alpha$  is the number of donor units that obtains positive weights in the training period. If  $k - \alpha > 0$ ,

the predictor weights are not unique and there will be proportionally increasing ambiguity.

On the other hand, the one-step procedure to estimate donor weights  $W^*(V^*)$  provides a well-defined estimator that outputs unique donor weights. The authors stress that the ill-defined cross-validation synthetic control estimator is not a failure of the method as such, and that the ambiguity may be negligible, although a positive  $k - \alpha$  is the typical scenario.

### 3.3 Kuosmanen et al. (2021) on the true nature of numerical instability

In contrast to the findings of Klößner et al. (2018), who attribute numerical instability to the cross-validation method, Kuosmanen et al. (2021) attribute it to the commonly used *Synth* package algorithm available for R, STATA, and Matlab, as well as the MSCMT algorithm, which produce unstable and suboptimal weights that fail to converge to the existing optimum unique solution. Therefore, different ordering of the donors and predictors affect the results.

The authors emphasize that numerical instability is not the underlying flaw but more of a symptom. They demonstrate that the model has a tendency for corner solutions, which is the true design flaw and is caused by the joint optimization of donor and predictor weights. The computational complexity of this NP-hard bilevel optimization problem causes other packages to fail.

Their proposed solution is to determine predictor and donor weights by applying a two-step algorithm that optimizes the donor weights when the predictor weights are given (note that the *Synth* package fails to derive optimal weights even when predictor weights are given), an approach that builds on previous work by Malo et al. (2020). The authors provide the updated code and technical documentation for this approach at their GitHub page (see p.4 of Kuosmanen et al. (2021) for the link).

An explicit restatement of the SCM problem provided by Malo et al. (2020) reveals the computational difficulty of problem. Following the paper's notation:

$$\min_{\mathbf{v}, \mathbf{w}} L_V = \frac{1}{T^{pre}} (y_1^{pre} - Y_0^{pre} w)' (y_1^{pre} - Y_0^{pre} w) \quad (18)$$

subject to

$$\mathbf{w} = \operatorname{argmin} L_W = (x_1 - X_0 w)' \mathbf{V} (x_1 - X_0 w) \quad (19)$$

$$\mathbf{1}' \mathbf{w} = 1$$

$$\mathbf{1}'\mathbf{v} = 1$$

$$\mathbf{w} \geq 0, \mathbf{v} \geq 0$$

Where (15) is the outer optimization problem, and (16) is the inner optimization problem.  $T^{pre}$  is the pre intervention period. This could be interpreted as a social planner playing a Stackelberg game where the leader chooses  $\mathbf{v}$  and the follower chooses  $\mathbf{w}$  cooperatively.

The proposed solution is detailed next.

*Step 1:* Solve the quadratic optimization problem.

$$\min_w L_Q = (x_q - X_0 w)' V^* (x_1 - X_0 w)$$

subject to

$$\mathbf{1}'\mathbf{w} = 1$$

$$\mathbf{w} \geq 0$$

*Step 2:* Given the step 1 optimal solution for  $L_W^*$ , solve the convex optimization problem.

$$\min_w L_V = (y_1^{pre} - Y_0^{pre} w)' (y_1^{pre} - Y_0^{pre} w)$$

subject to

$$\mathbf{1}'\mathbf{w} = 1$$

$$\mathbf{w} \geq 0$$

This procedure allows for multiple optima in Step 1, unlike *Synth*.

### 3.4 Ferman et al. (2020) on specification searching opportunities

The authors pose the issue of specification searching behavior in synthetic control applications stemming from a lack of consensus about which specifications should be used in SCM applications which has lead to a wide variety of choices in the literature. If different model specifications lead to different synthetic control units, the researcher may be tempted to discretionarily select the ones that lead to stastically significant results. Via Monte Carlo and placebo simulations they find that the probability of detecting a false positive is decreasing in the number of available pre-treatment periods, but still large when the time window is much greater than what is usually used in the literature.

They explore opportunities for specification searching when selecting the number of pre-treatment periods to include in the model. In accordance with

Kaul et al. (2015), they suggest to not include all pre-treatment periods if the model includes relevant covariates. If there are no covariates, the inclusion of all pre-treatment periods is the MSPE minimizing choice. The recommendation is to always include multiple specifications in synthetic control applications, and use the full pre-treatment period model as a benchmark.

A commonly used criterion to choose between different specifications is to select the ones that minimize the MSPE over the validation period. Some examples of this in practice are Dube and Zipperer (2015) and Donohue, Aneja and Weber (2019).

### 3.5 Klößner and Pfeifer (2017) on using SCM as a forecasting technique

Although the computation of a synthetic control does not strictly require post treatment values, not much work has been done on the use of SCM as a forecasting technique. In fact, Klößner and Pfeifer (2017) is the only academic work I could find on the topic and is mostly an introduction to the possibility of using SCM in this context. The authors compare the performance of SCM in forecasting US one period GDP growth to other well-known forecasting methods such as Holt-Winters, Random Walk and ARMA(p,q) with  $p, q \in \{0, 1, 2, 3, 4\}$  are chosen with Akaike and Bayes's information criteria.

The authors introduce a discount factor  $\beta$  in the outer optimization problem so that more recent observations weigh more:

$$\min_V \sum_{t=1}^T \beta^{T-t} \left( Y_t^1 - \sum_{j=2}^{J+1} W^*(V)_j Y_t^j \right)^2 \quad (20)$$

The inner optimization remains unchanged. In the model, the aim is to forecast the first quarter of 2015 by using as potential donors the lags of US GDP going from  $1, \dots, H$ , where  $H \in \{8, 12, 16, 20, 24, 28, 32, 36, 40\}$  is the maximum lag included. No covariates are included in the model.

To compare the quality of the forecast, the authors use different criteria, among which we find an average rank of the estimator according to forecast precision, mean absolute prediction error (MAPE) and root mean squared error (RMSPE). They find that RMSPE slightly favours SCM over other methods. When SCM is the winner, a higher  $\beta$  results in more precise estimates. Overall, SCM performs competitively with respect to other methods.



## 4 Applying SCM to the analysis of extreme natural events

Synthetic control methods have been applied to the study of right-to-carry laws, legalized prostitution, immigration policy, corporate connections, minimum wages and much more. This section focuses on reviewing the literature that applies synthetic control methods to the investigation of natural disasters and extreme natural events. For each of the reviewed works, I will provide a description the data and of which variation of SCM has been applied, as well as a brief review of the outcomes.

### 4.1 Cavallo et al. (2013) on catastrophic natural disasters and economic growth

This paper looks for empirical evidence on whether natural disasters affect economic growth. Economic theory does not provide a clear cut answer to the question, with neoclassical growth model stating that the destruction of capital following a natural disaster should not affect the rate of technological progress, hence it should only affect short term growth, and endogenous growth models even stating that growth could be higher as a result of reinvestment and upgrading of capital (think of the Schumpeterian creative destruction).

#### 4.1.1 Strategy

For each country affected by a natural disaster, the authors compute a counterfactual via SCM. The synthetic controls are computed from data about unaffected countries with the aim of producing a counterfactual for a country affected by a large, non-recurring event. Only disasters before the year 2000 are considered.

The model is applied on a sample of  $J + 1$  countries, where the first country  $J = 1$  is affected by the disaster and the remaining  $J$  countries are the donor pool. The aim is to estimate

$$\alpha_{it} = Y_{it}^I - Y_{it}^N$$

where  $Y_{it}^I$  is GDP per capita for country  $i$  at time  $t$  if the country is exposed to the disaster, and  $Y_{it}^N$  is GDP per capita for country  $i$  at time  $t$  in the absence of a disaster. While  $Y_{it}^I$  is observed, we need to find an estimate for  $Y_{it}^N$ . Following Abadie et al. (2010), they proceed by estimating  $W^*$  and  $V^*$  such that the MSPE is minimized using the out-of-sample validation approach.

#### 4.1.2 Data

The authors use a dataset covering 196 countries between 1970 and 2008 which includes:

- Data on real GDP per capita and purchasing power parity (PPP) from the World Bank and World Development Indicators (WDI).
- GDP predictors for trade openness (from WDI), capital stock (from Penn World Tables (PWT)), land area in km<sup>2</sup>, population, secondary educational attainment (from Lutz et al. (2007)), latitude in absolute value, Polity 2 indicator (from the Polity IV database).
- Data on natural disasters, their human and economic impacts from the EM-DAT database collected by the Centre for Research on the Epidemiology of Disasters (CRED). This database has information on natural disasters from 1900 at the condition that at least 10 or more people have died, 100 people have been affected, a state of emergency is declared or international assistance is called for. It provides two measures of the magnitude of a disaster: number of people killed and amount of direct damage in US dollars.

The authors only focus on large disasters, defined as exceeding the cutoffs of 99th, 90th and 75th percentiles of the world distribution of number of people killed. The 99th percentile corresponds to more than 233 deaths per 1 million inhabitants, the 90th to 17 deaths per 1 million, and the 75th to 7 deaths per 1 million (the mortality of Hurricane Katrina in 2005). Using these cutoffs, the subsample of disasters happening before 2000 with full data are of 4 events for the 99th percentile cutoff, 18 for the 90th percentile and 22 for the 74th percentile.

#### 4.1.3 Results

Findings show statistically significant and negative effects of very large natural disasters on GDP growth for both the short and long term. Of the four very large disaster (99th percentile cutoff), the 1979 Islamic Iranian Revolution after the 1978 earthquake and the Sandinista Nicaraguan Revolution of 1979 that followed an earthquake were followed by political revolutions. When controlling for political upheaval, the significance disappears. This dynamic is reminiscent of Acemoglu, Johnson and Robinson's (2005) work on the significance of institutions as a determinant of economic growth. For countries unaffected by revolutions and/or milder natural events, the effect of disasters is not significant.

As noted in Section 2.4, the interpretation of SCM as causal rests on the assumption that countries from the donor pool are unaffected both directly and indirectly by the treatment for unit  $J = 1$ . It is very hard to rule out this type of interaction if the donor pool has relations with the treated unit, so the authors check that the control group is not affected by the presence of a main trading partner.

## 4.2 Barone, Mocetti (2014) on growth and institutions after earthquakes in Italy

Barone and Mocetti essentially apply Cavallo's et al. (2013) approach to the case of two earthquakes in Italy, with the stated advantage that they can exactly pinpoint the geographical areas affected by the events instead of considering an entire country as impacted. A second advantage comes from the better comparability of Italian regions rather than comparing different countries. Third, they use data that is uniformly built since it comes from the same institutional sources. Finally, they can control for the amount of public transfers to each region, which play a crucial role in the long term consequences of natural disasters depending on whether these resources are misallocated or used to enhance the local economy.

### 4.2.1 Strategy

The empirical strategy is the one from Abadie et al. (2010) applied to  $J + 1$  Italian regions, of which the first unit is the treated one. Weights  $W$  are chosen to minimize the MSPE function.

### 4.2.2 Data

Data for the main outcome variable GDP per capita and a number of predictors (labor units, investments, population, sectoral shares of value added) comes from the CRENOs research institute and covers the 1951-2004 period. Then, the same data for 2004-2009 comes from the National Statistical Institute (ISTAT). Human capital data comes from ISTAT's census information.

The authors build four regional time-varying variables to control for the quality of local institutions. The first uses ISTAT data on embezzlement, bribery, corruption, infringement of public duty, fraud, fencing and bankruptcies to measure the intensity of corruption and fraudulent behavior. Then, using data from Fondazione Rodolfo de Benedetti, they build the fraction of national members of parliament appointed in each region who were involved

in scandals. The third variables captured political participation by using referenda and European election turnout for each region. The final one is the diffusion of newspaper readership using data from Accertamenti Diffusione Stampa (ADS).

Data on the number of deaths and affected people comes from the aforementioned EM-DAT. The first earthquake in the Friuli region killed 922 and affected 218,200 people, the second one in Irpinia (Campania and Basilicata regions) affected 407,700 and killed 4689.

#### **4.2.3 Results**

The method assigns sparse weights to a small number of donor pool regions in both cases, as usual. The main results indicate the absence of a significant short term shock. Accounting for fiscal stimulus in the form of financial aid highlights that it represented a sizeable chunk of regional GDP. Without financial aid, in the five years following the disasters, Friuli's GDP per capita would have been 2-4% lower and Irpinia's would have been 6-11% lower. However, in the long term the synthetic control detects two divergent trends of GDP per capita in the two regions. This can be reconnected to the quality of institutions and the consequent use of financial aid resources, with Friuli steering toward virtuous reconstruction, and Irpinia experiencing more rent seeking behaviors and misallocation of resources.

Designing financial aid in a way that is conscious of institutional quality and local dynamics favors immediate reconstruction and cushions the negative effects on GDP.

### **4.3 Coffman, Noy (2011) on measuring the long term impact of Hurricane Iniki**

Coffman and Noy study the long term impact of the 1992 category 4 hurricane Iniki on the Hawaian county of Kauai, using the unaffected islands as a donor pool to build a synthetic control. What makes the counties ideal candidates for the application of SCM is that they share similar economic structures, are affected by common shocks, they belong to the same state and thus share institutions and legal frameworks. Differences between the counties are mostly in terms of size. At the time of publication, Iniki was the strongest hurricane to hit Hawaii in recorded history causing \$7.4 billion of direct damages at 2008 prices.

### **4.3.1 Strategy**

The SCM is applied as in Abadie et al. (2010) with regression methods. To account for potential indirect effects of Iniki on the economic activity of other counties, Maui was excluded from the donor pool. This is due to tourism being the main economic activity of all islands and to Maui temporarily absorbing arrivals to the Kauai county right after the hurricane hit. Another channel for indirect effects could be that of displacement of residents to other islands; however, data shows this is not an issue.

### **4.3.2 Data**

Data is collected from 1975 to 2011, and includes information on employment, resident population, personal income, per capita income and transfer payments. Once again, data on deaths (4) and affected people (25,000) comes from the EM-DAT database.

### **4.3.3 Results**

Placebo testing confirms two central results. First, Kauai’s population trajectory took a hit following the hurricane, and kept growing slowly in the following years, with a permanent loss of 12% of its population. Second, private sector employment also declined significantly, with a permanent loss of around 15% of the total county employment.

## **5 Applying SCM to evaluate environmental policy**

This section is dedicated to a literature review focused on policy evaluation related to emissions, international agreements and event studies. Analogously to the previous section, I will focus on strategy, data sources and results for each work.

### **5.1 Cole, Elliott, Liu (2020) on the impact of lockdowns on Wuhan air pollution and health**

This paper by Cole et al. applies a combination of machine learning to clean weather data and a variation of SCM called augmented SCM to study the effect of reduced air pollution during lockdown on health. Improvements in air quality and associated health benefits are critical points in elaborating climate change related policy and emissions regulation.

### 5.1.1 Strategy

The city of Wuhan is one of 17 large cities in its province of Hubei (China), and it implemented a lockdown at least two weeks in advance with respect to other cities.

The authors adopt a machine learning approach that allows to remove the effect of weather conditions on concentrations of pollutants. Weather normalization has two advantages: first, it allows for accounting of local meteorological conditions since they affect pollution concentration, and second, it reduces volatility in the outcome variable, which is a prerequisite for successful application of the SCM.

Traditional econometric approaches are unsuitable for this task because pollutant concentration depends on local weather conditions. Instead, according to the authors, random forest machine learning techniques are the better option, applied by combining the work from Grange et al. (2018) and Vu et al. (2019).

Once weather is normalized, the augmented SCM (ASCM) is used to compute synthetic values for pollutant concentrations in Wuhan, had the lockdown not taken place. The advantage of ASCM over regular SCM is that it extends the method to cases where the pre-treatment fit would not be good enough to proceed. The Ben-Michael et al. (2019) approach allows for negative weights and improves the pre-intervention fit compared to SCM alone and penalizes for potential extrapolation.

The analysis spans for 30 days pre-treatment, and for around 12 days starting from January 21st, when a clear drop in emission manifests itself. The authors exclude anticipation effects as backdating the treatment does not affect results.

The impacts on mortality are computed considering only  $\text{NO}_2$  and the related health damages literature. The outcome is a significant reduction in mortality.

### 5.1.2 Data

The Qingyue Open Environmental Data Center provides data on sulphur dioxide ( $\text{SO}_2$ ), nitrogen dioxide ( $\text{NO}_2$ ), carbon monoxide ( $\text{CO}$ ) and particulate matter ( $\text{PM}_{10}$ ) between January 2013 and February 2020 for thirty Chinese cities. Information on meteorological data (temperature, relative humidity, wind direction, wind speed and pressure) comes from the “worldmet” R package (<https://www.ncdc.noaa.gov/isd>).

### 5.1.3 Results

Concentrations of  $\text{NO}_2$  fell significantly during the period, while the fall in concentration of  $\text{PM}_{10}$  was not statistically significant. Concentrations of  $\text{SO}_2$  did not fall, likely due to coal reliance for domestic heating during winter, as well as  $\text{CO}$ . The findings on reduced mortality show a rough estimate of 496 prevented deaths in Wuhan, 3368 prevented deaths in Hubei, and 10,822 in China.

Overall, this paper sheds light on how measures such as pedestrian zones, congestion charging and urban planning could benefit public health. It also shows how SCM may be combined with other methods to estimate the effect of climate-impacting policies.

## 5.2 Steinkraus (2018) on the green paradox

Steinkraus applies the SCM to the empirical study of the green paradox by analysing the coal production of four US states (Kentucky, Montana, Pennsylvania, Virginia) that enacted a greenhouse gas plan. The green paradox refers to the potential unwanted worsening of global warming as a response to climate policy intended to curb it. This might happen when plans do not take supply-side behavior into account, leading to a reduction in demand for fossil fuels that does not affect the supply, or generating unwanted behavior such as pushing for faster extraction before the policy becomes effective.

### 5.2.1 Strategy

A counterfactual of coal production is computed for each treated state and used to estimate the causal impact on coal production of the plan coming into effect. The donor pool is comprised of 16 states (Alabama, Georgia, Idaho, Indiana, Kansas, Louisiana, Mississippi, Nebraska, North Dakota, Ohio, Oklahoma, South Dakota, Tennessee, Texas, West Virginia and Wyoming) that did not announce or enact any plan for greenhouse gas emissions.

In light of Kaul et al. (2015), the author includes only coal production data at two points in time (1990 and the pre-treatment year).

### 5.2.2 Data

The data used in this study consists of total tons of coal produced between 1983 and 2014, which was obtained from the US Energy Information Administration (EIA). Other predictor variables include state-level energy production, state-level energy consumption, renewable energy production (all obtained from the EIA State Energy Data System), population density (from

US census data), recoverable reserves (from Höök and Aleklett (2009)), GDP (from the Bureau of Economic Analysis), and foreign distribution of coal production (as reported by EIA Annual Coal Distribution Report). The study also used data from the Seamless Digital Chart of the World and the Global Administrative Areas Project to compute the average distance of the states from the coast, in order to control for trade openness.

### **5.2.3 Results**

Findings show that while Pennsylvania, Virginia and Kentucky have not experienced an increase in coal production, Montana actually experienced a positive and significant jump in production as a response. However, this might be due to the state being the most likely to experience the legislation as an exogenous shock for political reasons and lack of previous interventions on the topic. Montana also has a longer border that allowed for export to unaffected states.

Due to mixed results, the author concludes that state by state reactions are very case dependent and point to the need to consider the supply side when constructing legislation.

## **5.3 Andersson (2019) on the study of carbon taxes and CO<sub>2</sub> emissions**

Andersson explores the causality of the link between carbon taxes and emission by analysing the introduction of a carbon tax and a value added tax on transport fuel in 1991 Sweden. The carbon tax mainly affects the transport sector, which is Sweden's main source of CO<sub>2</sub> emissions. For this reason, the analysis focuses on this sector.

### **5.3.1 Strategy**

The first part of the paper is dedicated to constructing a counterfactual Sweden using the SCM, with a donor pool of 14 countries. Since Sweden was one of the first countries in the world to implement such taxation, with the explicit aim of preventing climate change, it represents an ideal candidate for the method. The second half of the paper computes tax and price elasticities of the demand for gas fuel in order to distinguish between the carbon tax and the value added tax effects.

The author applies the method as in Abadie et al. (2010) via joint optimization of  $W$  and  $V$  in order to minimize the MSPE of the outcome variable over the pre-treatment period.



### 5.3.2 Data

The World Bank provided panel data on per capita CO<sub>2</sub> emissions for the transport sector of 25 OECD countries, by source (road, rail, domestic navigation, domestic aviation). Data on transport emissions was gathered from national statistical agencies and includes all transport fuels in order to capture changes in demand and substitutions between fuels (mainly between gasoline and diesel). The dataset covers the period from 1960 to 2005, when the EU implemented the European Union Emissions Trading System. Thus, there are 30 years before treatment and 16 years post treatment. Of the 25 aforementioned countries, those that enacted carbon taxes that touched the transport sector, made large changes in fuel taxation, or are known for fuel tourism were excluded from the dataset.

Data on predictors GDP per capita, number of motor vehicles, per capita gasoline consumption and percentage of population residing in urban areas comes from the World Bank and the works of Feenstra, Inklaar and Timmer (2013) and Dargay, Gately and Sommer (2007). Additional predictors are lagged values of CO<sub>2</sub> emissions for 1970, 1980, 1989.

### 5.3.3 Results

An analysis of differences between Sweden and the other OECD countries shows that until 1980 growth of emissions was similar, but between 1980 and 1989 emissions in Sweden grew twice as fast. This would have led to distorted results if a method such as difference-in-differences (DiD) were applied instead of SCM due to a violation of the parallel trend assumption.

Comparing real Sweden with synthetic Sweden shows that the average emission reduction for the 1990-2005 period is of 10.9% (2.5M metric tons) for the transport sector. Of this, 3.6% alone can be attributed to the carbon tax.

## 5.4 Maamoun (2019) on empirical evidence of the success of the Kyoto Protocol

Maamoun studies the empirical impact of the Kyoto Protocol (KP) via generalized SCM (GSCM) on ratifying countries' emissions. The KP rested on the principle of "common but differentiated responsibilities" and posed legally binding targets to its Annex B countries (industrialized countries) which were mandated to reduce their emissions by 5.2% on average with respect to base year 1990 during the first commitment period 2008-2012. Annex B includes 36 countries out of a total 192 ratifying countries, and

between those 34 represent the treated group, while 2 are excluded for lack of proper data (Monaco and Liechtenstein).

The KP was greatly criticized and empirical evidence using regression methods and DiD, DiD with matching to solve the lack of parallel trend and game theory models yield different conclusions on its effectiveness in reducing greenhouse gas and CO<sub>2</sub> emissions. Furthermore, regression methods are affected by selection bias due to the voluntary nature of the KP agreement, and other endogeneity concerns. GSCM offers the possibility to apply SCM to the case of multiple treated units to compute a more accurate counterfactual for the treated group.

#### **5.4.1 Strategy**

GSCM improves upon regular SCM by combining it with interactive fixed effects model (IFE). It allows for multiple treated unit and different treatment periods. It also provides standard errors and confidence intervals, which improves interpretability. Overall, GSCM is preferable to SCM in this context, and preferable to DiD when the parallel trend assumption might not be respected. For a detailed introduction of GSCM see Xu (2017).

The main limitations of GSCM are that it is biased if the pre-treatment period has less than 10 observations and the control group is less than 40 units large. In this case, the method can be applied since the control group is 119 countries and the pre-treatment period spans more than 10 years. A second limitation is that complex data generating processes should not be used, for instance there should not be structural breaks, or a dynamic relation between treatment and outcome variable, and it is not suited for different treatment intensities. Finally, differently from regular SCM, specification plays an important role.

The treated group is composed of the 34 Annex B countries for which necessary data is available. The computed counterfactual would represent the business as usual (BAU) scenario.

#### **5.4.2 Data**

The dataset covers the 1995-2012 period, with the treatment period starting in 2005 (enforcement period) rather than 2008 (commitment period). The output variable is greenhouse gas (GHG) excluding land use and forestry, gathered from the World Resources Institute's (WRI) CAIT Climate Data Explorer. Three control variables from the World Bank's World Development Indicators are included: real GDP per capita, squared real GDP per capita and population size.

### 5.4.3 Results

According to results, the KP successfully reduced GHG emissions for the industrialized countries by an estimated 6.8% with respect to the BAU counterfactual scenario. Removing EU countries from the treated group as a robustness exercise shows that the KP is still effective at reducing emissions, although significance decreases slightly.

The KP assigns “Hot Air” amounts of emissions that could exceed the BAU levels without violating the protocol. This measure was greatly criticized due to concerns that it would remove the incentive to decrease emissions. If we remove the Former Soviet Union countries from the analysis, who were allocated more emissions than their BAU levels, the magnitude of reduction reaches 17%. This is in line with criticism that the extra emissions allowances did not push for a decrease in emissions at best, and incentivised an increase at worst.

To alleviate the selection bias for Annex B countries who would be the ones legally bound by the agreement, the author uses the US as a control group for industrialized countries. Results are still significant and of similar magnitude.

## 5.5 He et al. (2022) on Free Trade Zones in China

This study uses SCM to analyse the effect of China’s free trade zones (FTZ) policy launched in 2013 in Shanghai and in 2015 in Tianjin, Guangdong and Fujian to understand its impact on CO<sub>2</sub> emissions. FTZs are areas where policies are implemented to promote trade openness, such as free entry and exit of goods, zero-tariff areas, acceleration of government functions, incentives to apply innovative management models, and promotion of investments. They are specifically targeted at the service and finance sectors (banking, insurance, capital markets). FTZs can reduce emissions if they improve efficiency in allocation of environment resources and boost eco-friendly technologies. They might increase emissions if they attract high emission goods production.

The FTZ policy also included environmental quality suggestions to improve energy efficiency and reduce GHG and other pollutants emissions.

### 5.5.1 Strategy

The authors use the creation of the first FTZ in Shanghai as the start of treatment and build a synthetic Shanghai to measure the impact of the policy on CO<sub>2</sub> emissions. They apply the SCM as in Abadie et al. (2010) to

construct a synthetic counterfactual unit based on CO<sub>2</sub> emissions from a donor pool of provinces where FTZs were not established.

The same analysis is done for the provinces of Tianjin, Fujian and Guangdong which established FTZs in 2015. Eventually, the provinces of Liaoning, Zhejiang, Henan, Hubei, Chongqing, Sichuan and Shaanxi implemented FTZ in 2017. When computing synthetic Shanghai, these provinces are excluded from the donor pool to avoid anticipatory effects on their emissions.

Note that the authors do not mention whether they controlled for other emission-affecting policies that may have been in effect in the donor pool provinces during the training period. If such policies were present, the provinces should have been excluded from the donor pool.

### 5.5.2 Data

It is notable that the authors do not employ data on CO<sub>2</sub> emissions from the Intergovernmental Panel on Climate Change (IPCC), notwithstanding this being the standard practice in the literature. Instead, they choose to use the official figures released by the China Emissions Accounts and Datasets (CEAD) stating that it has “a more comprehensive accounting range”, without clarifying details on the discrepancies between the accounting choices of the two sources. CEAD provides overall lower figures for emissions than those from the IPCC, with an approximate 8% discrepancy in 2012, but it does provide data at the provincial level, which is not easily retrievable from the IPCC dataset.

The dataset covers 30 provinces for the period 1997-2017. From the China Statistical Yearbook and the China Energy Statistical Yearbook, the authors gather data on output per capita, industrial structure, development level of service industry, industrial structure upgrading, energy efficiency, population density, technology market turnover and total fossil energy consumption.

### 5.5.3 Results

Looking at the synthetic Shanghai computed with the exclusion of the 2015 and 2017 FTZ provinces, they find a 10% reduction in CO<sub>2</sub> emissions from treatment. The same analysis is done for the provinces of Tianjin and Fujian, whose synthetic counterpart suggests a reduction in emissions post treatment, and for Guangdong, for which the discrepancy is not as marked nor consistent. The difference in results may be due to Shanghai being in many ways the cultural, technological and productive center of China, therefore it may be more responsive than other provinces to policy of this kind.

## 6 Discussion and conclusion

The synthetic control method has emerged as a valuable tool for causal inference in a variety of fields, particularly in settings where traditional experimental designs are not feasible. The method provides a flexible and intuitive framework for estimating treatment effects, allowing researchers to assess the impact of policy interventions, natural disasters, and other events of interest. The method has been successfully applied to a wide range of research questions, from the evaluation of health interventions to the analysis of economic policies.

SCM has several advantages over other methods of causal inference. First, it allows for the estimation of treatment effects in settings where traditional experimental designs are not feasible. Second, the method can incorporate a large number of covariates, allowing for a more comprehensive analysis of treatment effects. Third, the method is relatively easy to implement and interpret, making it accessible to researchers with varying levels of statistical expertise. Fourth, the method provides a flexible and intuitive framework for estimating treatment effects, allowing researchers to assess the impact of policy interventions, natural disasters, and other events of interest. Overall, SCM represents an important contribution to the field of causal inference, providing researchers with a powerful and accessible tool for estimating treatment effects.

Despite its widespread use in applied research, SCM has been subject to some criticisms and limitations. One of the main critiques relates to the choice of control units, which can significantly affect the results of the analysis. In particular, the method assumes that the control units have a similar pre-intervention trend as the treated unit, which may not always be the case. Moreover, the choice of control units can be subjective and dependent on the researcher’s prior beliefs, leading to potential bias in the estimation of treatment effects.

There are several other methods that can be used for similar purposes. One such method is the difference-in-differences (DiD) approach. In fact, many of the reviewed paper present DiD results alongside SCM for comparison. Nevertheless, SCM remains a valid alternative when we are unable to verify the parallel trend assumption or when there is uncertainty in its regard.

Regression discontinuity (RD) is another method that can be used to estimate treatment effects in settings where the treatment is assigned based on a threshold. RD compares outcomes on either side of the threshold, assuming that units close to the threshold are similar in all other respects except for the treatment. While RD can be a powerful tool for causal inference, it

requires a precise definition of the threshold and may not be applicable in settings where the treatment is not assigned based on a clear cutoff.

Outside of the realm of parametric methods, a notable competitor for SCM is found in machine learning (ML) applications. When choosing between SCM and ML algorithms for causal inference, it is important to consider the specific characteristics of the data and research question. ML algorithms are non-parametric methods that can handle complex relationships between variables and accommodate time-varying treatment effects and multiple treated units. However, they may suffer from overfitting if the model is too complex and may be more difficult to interpret. On the other hand, SCM is a parametric method that requires less strict assumptions about the data and may be more appropriate for research questions that require a clear causal interpretation of the treatment effect or when estimating a few key parameters. ML algorithms may be more appropriate when the research question requires more flexibility and complexity in the modeling or when the data is more complex and requires a different approach.

Despite its limitations, SCM has continued to be refined and extended in recent years, with new developments addressing some of its shortcomings. For example, extensions of the method have been proposed to accommodate multiple treated units or time-varying treatment effects, while other research has focused on improving the selection of control units or assessing the sensitivity of the results to alternative specifications. Also, the work of Klößner and colleagues successfully addresses the aforementioned numerical instability issues.

In this paper, I introduced SCM as a statistical technique for estimating causal effects in observational studies. I discussed the properties, advantages, and contextual and data requirements of the method. Furthermore, I explored the challenges that may hinder the effectiveness of the method and discussed how to address them.

In the second part of the paper, I reviewed the literature on the applications of the synthetic control method to the analysis of extreme natural events and the evaluation of environmental policies. I highlighted some of the key findings and insights from these studies, which demonstrate the potential of the synthetic control method for policy evaluation and decision-making. For policymakers, the method can help in making evidence-based decisions, particularly in situations where experimental designs are not feasible. It can also provide a more accurate and reliable estimation of the causal effect of a policy, which can help in allocating resources more effectively.

Overall, this analysis suggests that SCM can be a valuable tool for causal inference and policy evaluation in situations where traditional experimental designs are not feasible. However, as with any statistical method, its

effectiveness depends on the quality and availability of data, as well as the appropriateness of the assumptions and model specifications.

That said, there are still opportunities for improvement. One potential avenue for development is the incorporation of machine learning techniques, such as deep learning and neural networks, into the synthetic control framework. These techniques may offer advantages in capturing complex interactions among variables and improving prediction accuracy. An example of successful combination between these methods is Cole et al. (2020), as discussed in section 5.1. Another possible avenue is the use of multiple treatment groups, where the synthetic control group is created for each treatment group, rather than a single control group. This approach may provide more robust and reliable estimates of causal effects in situations where there are multiple policies or interventions being evaluated simultaneously. Finally, the development of more efficient algorithms for the solution of joint optimization problems is a focal point to consider. Overall, the continued development and refinement of SCM have the potential to further expand its usefulness and applicability.

## 7 References

1. Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal of the American statistical Association* 105.490 (2010): 493-505.
2. Abadie, Alberto, and Javier Gardeazabal. "The economic costs of conflict: A case study of the Basque Country." *American economic review* 93.1 (2003): 113-132.
3. Abadie, Alberto, and J  r  my L'hour. "A penalized synthetic control estimator for disaggregated data." *Journal of the American Statistical Association* 116.536 (2021): 1817-1834.
4. Abadie, Alberto. "Using synthetic controls: Feasibility, data requirements, and methodological aspects." *Journal of Economic Literature* 59.2 (2021): 391-425.
5. Acemoglu, Daron, et al. "The value of connections in turbulent times: Evidence from the United States." *Journal of Financial Economics* 121.2 (2016): 368-391.
6. Acemoglu, Daron, Simon Johnson, and James A. Robinson. "Institutions as a fundamental cause of long-run growth." *Handbook of economic growth* 1 (2005): 385-472.
7. Andersson, Julius J. "Carbon taxes and CO 2 emissions: Sweden as a case study." *American Economic Journal: Economic Policy* 11.4 (2019): 1-30.
8. Athey, Susan, and Guido W. Imbens. "The state of applied econometrics: Causality and policy evaluation." *Journal of Economic perspectives* 31.2 (2017): 3-32.
9. Barone, Guglielmo, and Sauro Mocetti. "Natural disasters, growth and institutions: a tale of two earthquakes." *Journal of Urban Economics* 84 (2014): 52-66.
10. Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. "The augmented synthetic control method." *Journal of the American Statistical Association* 116.536 (2021): 1789-1803.
11. Cavallo, Eduardo, et al. "Catastrophic natural disasters and economic growth." *Review of Economics and Statistics* 95.5 (2013): 1549-1561.



12. Coffman, Makena, and Ilan Noy. "Hurricane Iniki: measuring the long-term economic impact of a natural disaster using synthetic control." *Environment and Development Economics* 17.2 (2012): 187-205.
13. Cole, Matthew A., Robert JR Elliott, and Bowen Liu. "The impact of the Wuhan Covid-19 lockdown on air pollution and health: a machine learning and augmented synthetic control approach." *Environmental and Resource Economics* 76.4 (2020): 553-580.
14. Dargay, Joyce, Dermot Gately, and Martin Sommer. "Vehicle ownership and income growth, worldwide: 1960-2030." *The energy journal* 28.4 (2007).
15. Donohue, John J., Abhay Aneja, and Kyle D. Weber. "Right-to-carry laws and violent crime: A comprehensive assessment using panel data and a state-level synthetic control analysis." *Journal of Empirical Legal Studies* 16.2 (2019): 198-247.
16. Doudchenko, Nikolay, and Guido W. Imbens. "Balancing, regression, difference-in-differences and synthetic control methods: A synthesis." No. w22791. National Bureau of Economic Research, 2016.
17. Dube, Arindrajit, and Ben Zipperer. "Pooling multiple case studies using synthetic controls: An application to minimum wage policies." (2015).
18. Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. "The next generation of the Penn World Table." *American economic review* 105.10 (2015): 3150-3182.
19. Grange, Stuart K., et al. "Random forest meteorological normalisation models for Swiss PM 10 trend analysis." *Atmospheric Chemistry and Physics* 18.9 (2018): 6223-6239.
20. He, Lingyun, et al. "Free Trade Zone Policy and Carbon Dioxide Emissions: a Synthetic Control Group Approach." *Polish Journal of Environmental Studies* 31.4 (2022).
21. Höök, Mikael, Robert Hirsch, and Kjell Aleklett. "Giant oil field decline rates and their influence on world oil production." *Energy policy* 37.6 (2009): 2262-2272.
22. Kaul, Ashok, et al. "Synthetic control methods: Never use all pre-intervention outcomes together with covariates." (2015).

23. Klößner, Stefan, and Gregor Pfeifer. "Outside the box: Using synthetic control methods as a forecasting technique." *Applied Economics Letters* 25.9 (2018): 615-618.
24. Klößner, Stefan, et al. "Comparative politics and the synthetic control method revisited: A note on Abadie et al.(2015)." *Swiss journal of economics and statistics* 154 (2018): 1-11.
25. Luo, Kevin, and Tomoko Kinugasa. "Do natural disasters influence long-term savings?: Assessing the impact of the 2008 Sichuan earthquake on household saving rates using synthetic control." *China: An International Journal* 18.3 (2020): 59-81.
26. Maamoun, Nada. "The Kyoto protocol: Empirical evidence of a hidden success." *Journal of Environmental Economics and Management* 95 (2019): 227-256.
27. Malo, Pekka, et al. "Computing Synthetic Controls Using Bilevel Optimization." (2020).
28. Steinkraus, Arne. "A synthetic control assessment of the green paradox: The role of climate action plans." *German Economic Review* 20.4 (2019): e545-e570.
29. Vu, Tuan V., et al. "Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique." *Atmospheric Chemistry and Physics* 19.17 (2019): 11303-11314.
30. Xu, Yiqing. "Generalized synthetic control method: Causal inference with interactive fixed effects models." *Political Analysis* 25.1 (2017): 57-76.