Causal Inference

Lecture #3

Juraj Medzihorsky

SGIA & RMC
Durham University

6 February 2024

Synthetic Control

Comparative politics and the **synthetic control** method

A Abadie, A Diamond… - American Journal of …, 2015 - Wiley Online Library

… Moreover, in contrast to regression analysis techniques, the **synthetic control** method makes … of interest and the **synthetic control**. We apply the **synthetic control** method to estimate the …

Save Cite Cited by 2228 Related articles All 21 versions

Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program

A Abadie, A Diamond, J Hainmueller - Journal of the American …, 2010 - Taylor & Francis

… the application of **synthetic control** methods to comparative … relative to a comparable **synthetic control** region. We estimate … units, the potential applicability of **synthetic control** methods to …

Save Cite Cited by 5337 Related articles All 43 versions

Assessing economic liberalization episodes: A **synthetic control** approach

A Billmeier, T Nannicini - Review of Economics and Statistics, 2013 - direct.mit.edu

… Within the **synthetic control** framework, we ask whether liberalizing the economy in year T … —the **synthetic control**. The comparison economies that form the **synthetic control** unit are …

Save Cite Cited by 536 Related articles All 8 versions

Synth: An R package for **synthetic control** methods in comparative case studies

A Abadie, A Diamond, J Hainmueller - Journal of Statistical …, 2011 - papers.ssrn.com

… The weights that define the **synthetic control** unit are chosen such that the **synthetic control** … The goal of the **synthetic control** method is to construct a **synthetic control** group that yields a …

Save Cite Cited by 528 Related articles All 17 versions

Synthetic Control: Motivation

- Sometimes, like difference-in-difference designs, we have panel data on multiple units and an exogenously imposed “treatment” on some of those units, but...

Synthetic Control: Motivation

- Sometimes, like difference-in-difference designs, we have panel data on multiple units and an exogenously imposed "treatment" on some of those units, but...
- both the treatment and the outcomes that we want to study are observed at a high level of aggregation (e.g., cities, provinces, countries), and...

Synthetic Control: Motivation

- Sometimes, like difference-in-difference designs, we have panel data on multiple units and an exogenously imposed “treatment” on some of those units, but...
- both the treatment and the outcomes that we want to study are observed at a high level of aggregation (e.g., cities, provinces, countries), and...
- that level of aggregation leads to two major challenges:

Synthetic Control: Motivation

- Sometimes, like difference-in-difference designs, we have panel data on multiple units and an exogenously imposed “treatment” on some of those units, but...
- both the treatment and the outcomes that we want to study are observed at a high level of aggregation (e.g., cities, provinces, countries), and...
- that level of aggregation leads to two major challenges:
    1. Only a small number of cases (often just 1 unit) experience the treatment, and

Synthetic Control: Motivation

- Sometimes, like difference-in-difference designs, we have panel data on multiple units and an exogenously imposed “treatment” on some of those units, but...
- both the treatment and the outcomes that we want to study are observed at a high level of aggregation (e.g., cities, provinces, countries), and...
- that level of aggregation leads to two major challenges:
    1. Only a small number of cases (often just 1 unit) experience the treatment, and
    2. We have reason to believe that the parallel trends assumption of the difference-in-differences framework is unreasonable

Synthetic Control: Motivation

- Sometimes, like difference-in-difference designs, we have panel data on multiple units and an exogenously imposed "treatment" on some of those units, but...
- both the treatment and the outcomes that we want to study are observed at a high level of aggregation (e.g., cities, provinces, countries), and...
- that level of aggregation leads to two major challenges:
    1. Only a small number of cases (often just 1 unit) experience the treatment, and
    2. We have reason to believe that the parallel trends assumption of the difference-in-differences framework is unreasonable

- Now we find ourselves in the world of what is often referred to as a "comparative case study"

Synthetic Control: Motivation

Comparative case studies have a long history in social science

Synthetic Control: Motivation

Comparative case studies have a long history in social science

- Particularly strong in political science

Synthetic Control: Motivation

Comparative case studies have a long history in social science

- Particularly strong in political science
- **Qualitative case studies**
  - Identify two or more instances of a phenomenon
  - Inductive reasoning about similarities and differences between cases that shed light on causal effects

Synthetic Control: Motivation

Comparative case studies have a long history in social science

- Particularly strong in political science

- **Qualitative case studies**
  - Identify two or more instances of a phenomenon
  - Inductive reasoning about similarities and differences between cases that shed light on causal effects
- **Quantitative case studies**
  - Identify an exogenous event that puts a small number of aggregate units into one group and a small number into another group
  - Employ causal inference techniques within the potential outcomes framework

Synthetic Control: Motivation

Researchers performing comparative case studies often (but not always)

- Do not have the type of longitudinal data that we have when we say that we have "panel" data
  - The same quantitative measurements on the same units at regular time intervals

Synthetic Control: Motivation

Researchers performing comparative case studies often (but not always)

- Do not have the type of longitudinal data that we have when we say that we have "panel" data
  - The same quantitative measurements on the same units at regular time intervals
- Consider causality via alternative modes of inference, such as process tracing

Synthetic Control: Motivation

Researchers performing comparative case studies often (but not always)

- Do not have the type of longitudinal data that we have when we say that we have "panel" data
    - The same quantitative measurements on the same units at regular time intervals
- Consider causality via alternative modes of inference, such as process tracing

But we want to perform inference by rooting our chosen case study firmly within the potential outcomes framework

Synthetic Control: Motivation

Researchers performing comparative case studies often (but not always)

- Do not have the type of longitudinal data that we have when we say that we have "panel" data
    - The same quantitative measurements on the same units at regular time intervals
- Consider causality via alternative modes of inference, such as process tracing

But we want to perform inference by rooting our chosen case study firmly within the potential outcomes framework

What can we do?

Synthetic Control: Motivation

The pioneering work of Abadie and Gardeazabal (2003) provides one potential path forward if we are willing to make some assumptions

The pioneering work of Abadie and Gardeazabal (2003) provides one potential path forward if we are willing to make some assumptions

The basic idea

- If we are confident that the cases are simply not comparable, and therefore that the parallel trends assumption of difference in differences will not hold...

Synthetic Control: Motivation

The pioneering work of Abadie and Gardeazabal (2003) provides one potential path forward if we are willing to make some assumptions

The basic idea

- If we are confident that the cases are simply not comparable, and therefore that the parallel trends assumption of difference in differences will not hold...
- Why not try to combine information across units that did not receive the treatment to construct a more reasonable counterfactual for the treated unit(s)?

Synthetic Control: Motivation

The pioneering work of Abadie and Gardeazabal (2003) provides one potential path forward if we are willing to make some assumptions

The basic idea

- If we are confident that the cases are simply not comparable, and therefore that the parallel trends assumption of difference in differences will not hold...
- Why not try to combine information across units that did not receive the treatment to construct a more reasonable counterfactual for the treated unit(s)?
- And if the number of cases is too small for standard statistical inference, simulate to get measures of uncertainty

Synthetic Control Method: Implementation

- Let $X_1 = (Z_1, \bar{Y}_1^{K_1}, \ldots, \bar{Y}_1^{K_M})'$ be a $(k \times 1)$ vector of pre-intervention characteristics.
    - The vector X_1 contains all relevant available predictors of the outcome variable
- Similarly, X_0 is a $(k \times J)$ matrix which contains the same variables for the unaffected units

- Let $X_1 = (Z_1, \bar{Y}_1^{K_1}, \ldots, \bar{Y}_1^{K_M})'$ be a $(k \times 1)$ vector of pre-intervention characteristics.
  - The vector X_1 contains all relevant available predictors of the outcome variable
- Similarly, X_0 is a $(k \times J)$ matrix which contains the same variables for the unaffected units
- The vector W^* is chosen to minimize the following expression:

$$\sum_{m=1}^{k} v_m (X_{1m} - X_{0m} W)^2,$$

where

  - $w_j \geq 0$ for $j = 2, \ldots, J + 1$ (only positive or zero weight can be applied to a case in the donor pool)
  - v_m is a weight that captures the relative importance of the mth variable (covariate or pretreatment outcome) in measuring the distance between treated and control units

Synthetic Control Method: Implementation

With these two components in mind:

1. W: the vector of weights, $w_2, \ldots, w_{J+1}$ applied to each unit in the donor pool
2. V: the diagonal matrix whose diagonal is $v_1, \ldots, v_k$, the weights applied to each covariate

there are a variety of ways one could choose the synthetic control, $W^*(V)$

With these two components in mind:

1. W: the vector of weights, $w_2, \ldots, w_{J+1}$ applied to each unit in the donor pool
2. V: the diagonal matrix whose diagonal is $v_1, \ldots, v_k$, the weights applied to each covariate

there are a variety of ways one could choose the synthetic control, $W^*(V)$

The most common method in the literature for choosing V seems to be the choice that minimizes the mean-squared prediction error for the pre-intervention periods:

$$\sum_{t=1}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j^*(V) Y_{jt} \right)^2$$

The Application: California's Proposition 99

In 1988, California first passed comprehensive tobacco control legislation:

- increased cigarette tax by 25 cents/pack
- earmarked tax revenues to health and anti-smoking budgets
- funded anti-smoking media campaigns
- spurred clean-air ordinances throughout the state
- produced more than $100 million per year in anti-tobacco projects

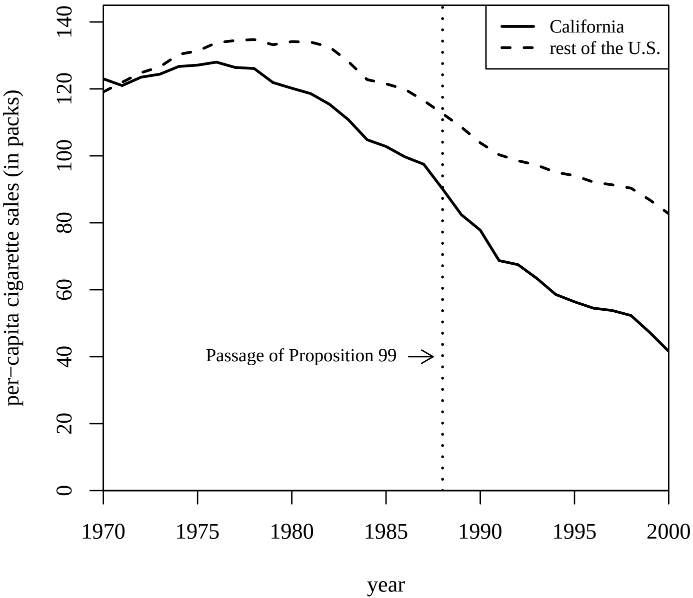
The Application: California's Proposition 99

In 1988, California first passed comprehensive tobacco control legislation:

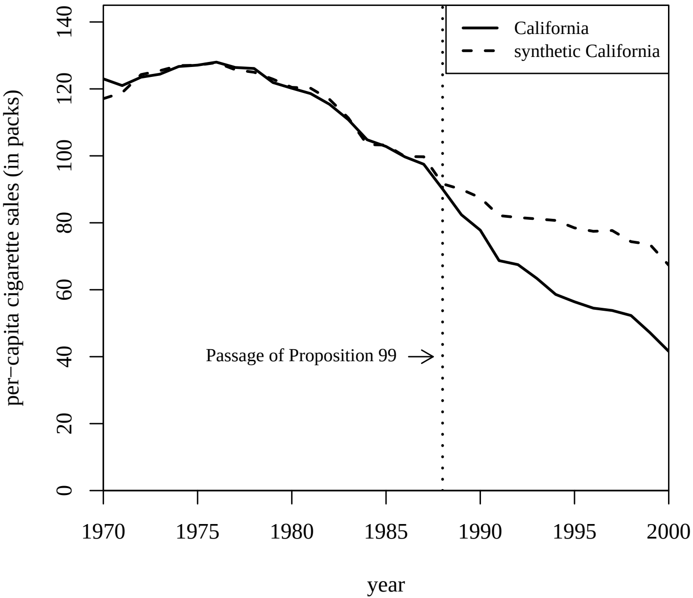
- increased cigarette tax by 25 cents/pack
- earmarked tax revenues to health and anti-smoking budgets
- funded anti-smoking media campaigns
- spurred clean-air ordinances throughout the state
- produced more than $100 million per year in anti-tobacco projects

Other states that subsequently passed control programs are excluded from donor pool of controls (AK, AZ, FL, HA, MA, MD, MI, NJ, NY, OR, WA, DC)

Cigarette Consumption: CA and the Rest of the U.S.



Cigarette Consumption: CA and synthetic CA

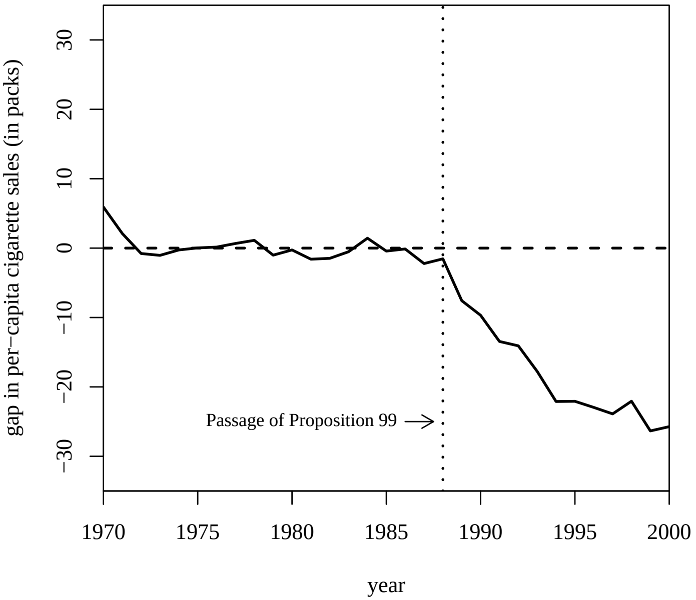


Predictor Means: Actual vs. Synthetic California

Variables	California Real	California Synthetic	Average of 38 control states
Ln(GDP per capita)	10.08	9.86	9.86
Percent aged 15-24	17.40	17.40	17.29
Retail price	89.42	89.41	87.27
Beer consumption per capita	24.28	24.20	23.75
Cigarette sales per capita 1988	90.10	91.62	114.20
Cigarette sales per capita 1980	120.20	120.43	136.58
Cigarette sales per capita 1975	127.10	126.99	132.81

Note: All variables except lagged cigarette sales are averaged for the 1980-1988 period (beer consumption is averaged 1984-1988).

Smoking Gap Between CA and synthetic CA



Positive State Weights in the Synthetic CA

State	Weight
Colorado	0.164
Connecticut	0.069
Montana	0.199
Nevada	0.234
Utah	0.334

Inference

- Iteratively apply the synthetic method to each state in the "donor pool" and obtain a distribution of placebo effects

Inference

- Iteratively apply the synthetic method to each state in the “donor pool” and obtain a distribution of placebo effects
- Compare the gap for California to the distribution of the placebo gaps

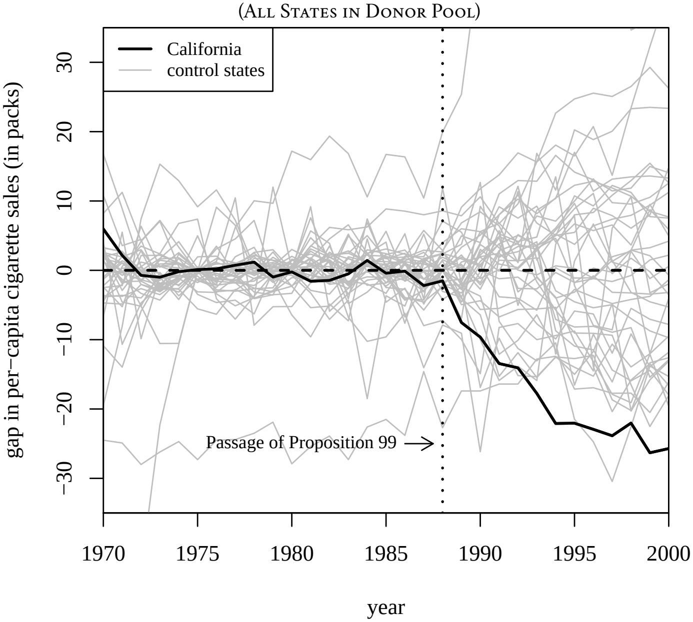
Inference

- Iteratively apply the synthetic method to each state in the "donor pool" and obtain a distribution of placebo effects
- Compare the gap for California to the distribution of the placebo gaps
- Question is whether the effect estimated by the synthetic control for the unit affected by the intervention is large relative to the effect estimated for a unit chosen at random

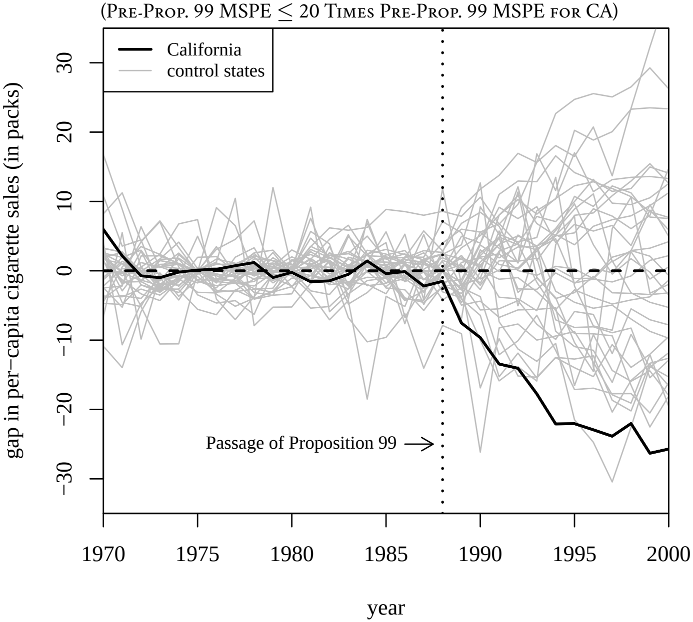
Inference

- Iteratively apply the synthetic method to each state in the "donor pool" and obtain a distribution of placebo effects
- Compare the gap for California to the distribution of the placebo gaps
- Question is whether the effect estimated by the synthetic control for the unit affected by the intervention is large relative to the effect estimated for a unit chosen at random
- This produces valid inference regardless of the number of available comparison units, time periods, and whether the data are individual or aggregate

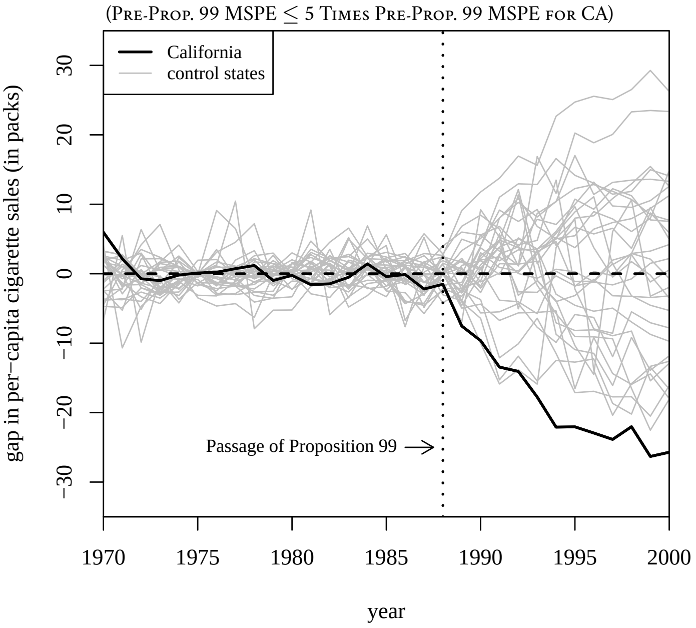
Smoking Gap for CA and 38 control states



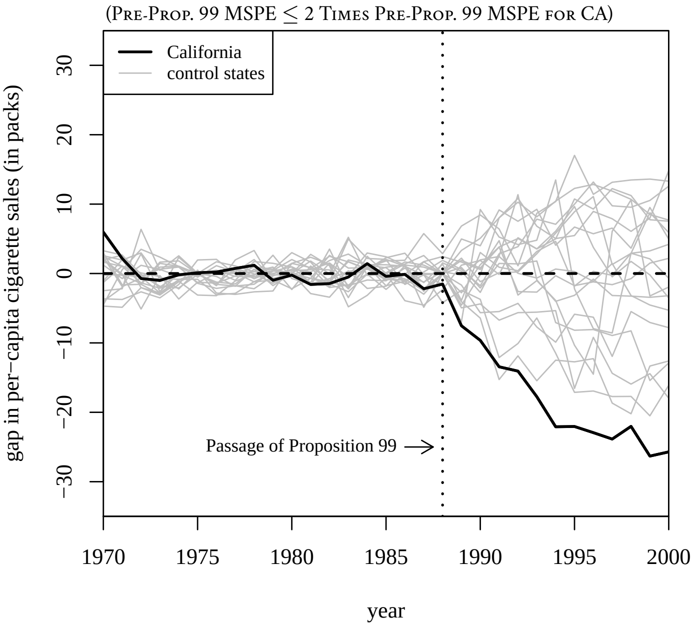
Smoking Gap for CA and 34 control states



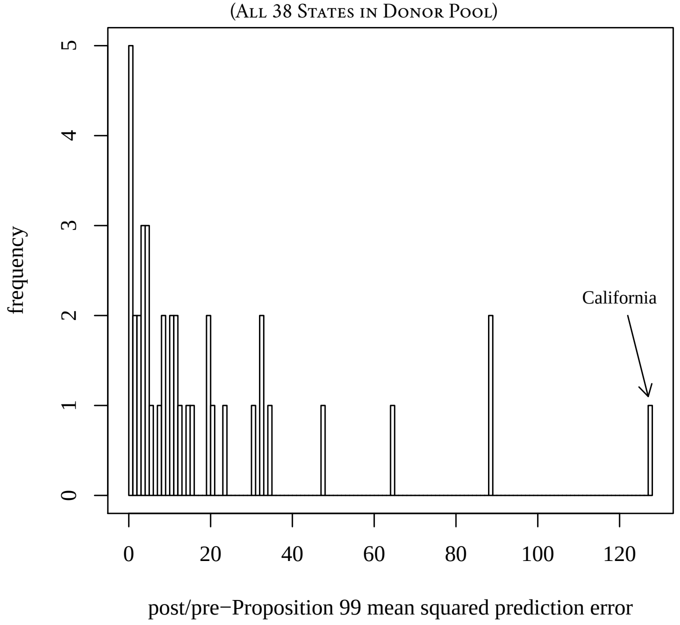
Smoking Gap for CA and 29 control states



Smoking Gap for CA and 19 control states



Ratio Post-Prop. 99 MSPE to Pre-Prop. 99 MSPE



Synthetic Control Method: A Motivating Model

- Suppose that we observe $J + 1$ units in periods $1, 2, \ldots, T$

Synthetic Control Method: A Motivating Model

- Suppose that we observe $J+1$ units in periods $1, 2, \ldots, T$
- Region "one" is exposed to the intervention during periods $T_0 + 1, \ldots, T$

Synthetic Control Method: A Motivating Model

- Suppose that we observe $J + 1$ units in periods $1, 2, \ldots, T$
- Region "one" is exposed to the intervention during periods $T_0 + 1, \ldots, T$
- Let Y_{it}^N be the outcome that would be observed for unit i at time t in the absence of the intervention

Synthetic Control Method: A Motivating Model

- Suppose that we observe $J+1$ units in periods $1, 2, \ldots, T$
- Region "one" is exposed to the intervention during periods $T_0+1, \ldots, T$
- Let Y_{it}^N be the outcome that would be observed for unit i at time t in the absence of the intervention
- Let Y_{it}^I be the outcome that would be observed for unit i at time t if unit i is exposed to the intervention in periods T_0+1 to T

Synthetic Control Method: A Motivating Model

- Suppose that we observe $J+1$ units in periods $1, 2, \ldots, T$
- Region "one" is exposed to the intervention during periods $T_0 + 1, \ldots, T$
- Let Y_{it}^N be the outcome that would be observed for unit i at time t in the absence of the intervention
- Let Y_{it}^I be the outcome that would be observed for unit i at time t if unit i is exposed to the intervention in periods $T_0 + 1$ to T
- We aim to estimate the effect of the intervention on the treated unit $(\alpha_{1T_0+1}, \ldots \alpha_{1T})$, where $\alpha_{1t} = Y_{1t}^I - Y_{1t}^N = Y_{1t} - Y_{1t}^N$ for $t > T_0$.

Synthetic Control Method: A Motivating Model

- We can always write $Y_{it}^N = \delta_t + \nu_{it}$, where
  - δ_t is an unknown common factor, constant across units

Synthetic Control Method: A Motivating Model

- We can always write $Y_{it}^N = \delta_t + \nu_{it}$, where
  - δ_t is an unknown common factor, constant across units
  - ν_{it} has mean zero for each t

Synthetic Control Method: A Motivating Model

- We can always write $Y_{it}^N = \delta_t + \nu_{it}$, where
  - δ_t is an unknown common factor, constant across units
  - ν_{it} has mean zero for each t
- Suppose that ν_{it} is given by the following model:

$$\nu_{it} = Z_i\theta_t + \lambda_t\mu_i + \varepsilon_{it},$$

  - Z_i is a $(1 \times r)$ vector of observed covariates

Synthetic Control Method: A Motivating Model

- We can always write $Y_{it}^N = \delta_t + \nu_{it}$, where
  - δ_t is an unknown common factor, constant across units
  - ν_{it} has mean zero for each t
- Suppose that ν_{it} is given by the following model:

$$\nu_{it} = Z_i\theta_t + \lambda_t\mu_i + \varepsilon_{it},$$

  - Z_i is a $(1 \times r)$ vector of observed covariates
  - θ_t is a $(r \times 1)$ vector of unknown parameters

Synthetic Control Method: A Motivating Model

- We can always write $Y_{it}^{N} = \delta_t + \nu_{it}$, where
  - δ_t is an unknown common factor, constant across units
  - ν_{it} has mean zero for each t
- Suppose that ν_{it} is given by the following model:

$$\nu_{it} = Z_i \theta_t + \lambda_t \mu_i + \varepsilon_{it},$$

  - Z_i is a $(1 \times r)$ vector of observed covariates
  - θ_t is a $(r \times 1)$ vector of unknown parameters
  - λ_t is a $(1 \times F)$ vector of unknown common factors, constant across units

Synthetic Control Method: A Motivating Model

- We can always write $Y_{it}^N = \delta_t + \nu_{it}$, where
  - δ_t is an unknown common factor, constant across units
  - ν_{it} has mean zero for each t
- Suppose that ν_{it} is given by the following model:

$$\nu_{it} = Z_i \theta_t + \lambda_t \mu_i + \varepsilon_{it},$$

  - Z_i is a $(1 \times r)$ vector of observed covariates
  - θ_t is a $(r \times 1)$ vector of unknown parameters
  - λ_t is a $(1 \times F)$ vector of unknown common factors, constant across units
  - μ_i is a $(F \times 1)$ vector of unit-specific unknown factors

Synthetic Control Method: A Motivating Model

- We can always write $Y_{it}^N = \delta_t + \nu_{it}$, where
  - δ_t is an unknown common factor, constant across units
  - ν_{it} has mean zero for each t
- Suppose that ν_{it} is given by the following model:

$$\nu_{it} = Z_i\theta_t + \lambda_t\mu_i + \varepsilon_{it},$$

  - Z_i is a $(1 \times r)$ vector of observed covariates
  - θ_t is a $(r \times 1)$ vector of unknown parameters
  - λ_t is a $(1 \times F)$ vector of unknown common factors, constant across units
  - μ_i is a $(F \times 1)$ vector of unit-specific unknown factors
  - ε_{it} are unobserved transitory shocks

Synthetic Control Method: A Motivating Model

- We can always write $Y_{it}^N = \delta_t + \nu_{it}$, where
  - δ_t is an unknown common factor, constant across units
  - ν_{it} has mean zero for each t
- Suppose that ν_{it} is given by the following model:

$$\nu_{it} = Z_i\theta_t + \lambda_t\mu_i + \varepsilon_{it},$$

  - Z_i is a $(1 \times r)$ vector of observed covariates
  - θ_t is a $(r \times 1)$ vector of unknown parameters
  - λ_t is a $(1 \times F)$ vector of unknown common factors, constant across units
  - μ_i is a $(F \times 1)$ vector of unit-specific unknown factors
  - ε_{it} are unobserved transitory shocks
  - Specification allows heterogeneous responses to multiple unobserved factors

Synthetic Control Method: A Motivating Model

- We can always write $Y_{it}^N = \delta_t + \nu_{it}$, where
  - δ_t is an unknown common factor, constant across units
  - ν_{it} has mean zero for each t
- Suppose that ν_{it} is given by the following model:

$$\nu_{it} = Z_i\theta_t + \lambda_t\mu_i + \varepsilon_{it},$$

  - Z_i is a $(1 \times r)$ vector of observed covariates
  - θ_t is a $(r \times 1)$ vector of unknown parameters
  - λ_t is a $(1 \times F)$ vector of unknown common factors, constant across units
  - μ_i is a $(F \times 1)$ vector of unit-specific unknown factors
  - ε_{it} are unobserved transitory shocks
  - Specification allows heterogeneous responses to multiple unobserved factors
- In contrast, the Difference-in-Differences (or Fixed-Effects) model restricts λ_t to be constant

Synthetic Control Method: A Motivating Model

- We can always write $Y_{it}^N = \delta_t + \nu_{it}$, where
  - δ_t is an unknown common factor, constant across units
  - ν_{it} has mean zero for each t
- Suppose that ν_{it} is given by the following model:

$$\nu_{it} = Z_i \theta_t + \lambda_t \mu_i + \varepsilon_{it},$$

  - Z_i is a $(1 \times r)$ vector of observed covariates
  - θ_t is a $(r \times 1)$ vector of unknown parameters
  - λ_t is a $(1 \times F)$ vector of unknown common factors, constant across units
  - μ_i is a $(F \times 1)$ vector of unit-specific unknown factors
  - ε_{it} are unobserved transitory shocks
  - Specification allows heterogeneous responses to multiple unobserved factors
- In contrast, the Difference-in-Differences (or Fixed-Effects) model restricts λ_t to be constant
- But: Note the linearity in Z_i assumption

Synthetic Control Method: A Motivating Model

- We can always write $Y_{it}^N = \delta_t + \nu_{it}$, where
  - $\delta_t \Rightarrow$ Time FE
  - $\nu_{it} \Rightarrow$ Zero mean in each time period
- Suppose that ν_{it} is given by the following model:

$$\nu_{it} = Z_i\theta_t + \lambda_t\mu_i + \varepsilon_{it},$$

  - Z_i is a $(1 \times r) \Rightarrow$ Exogenous covariates
  - $\theta_t \Rightarrow$ Coefficients, constant across units
  - $\lambda_t \Rightarrow$ DiD assumes this to be time-constant
  - μ_i is a $(F \times 1) \Rightarrow \lambda_t \cdot \mu_i =$ Time-varying unit FE
  - $\varepsilon_{it} \Rightarrow$ Usual i.i.d. error term
  - Specification allows heterogeneous responses to multiple unobserved factors
- In contrast, the Difference-in-Differences (or Fixed-Effects) model restricts λ_t to be constant
- But: Note the linearity in Z_i assumption

- The vector Z_i may contain pre- and post-intervention values of time-varying variables, as long as they are not affected by the intervention

Synthetic Control Method: A Motivating Model

- The vector Z_i may contain pre- and post-intervention values of time-varying variables, as long as they are not affected by the intervention
- For example, if $T = 2$, $T_0 = 1$, $Z_i = (Z_{i1}, Z_{i2})$,

$$\theta_1 = \begin{pmatrix} \beta \\ 0 \end{pmatrix} \quad \text{and} \quad \theta_2 = \begin{pmatrix} 0 \\ \beta \end{pmatrix},$$

then $Z_i\theta_t = Z_{it}\beta$

Synthetic Control Method: A Motivating Model

- Let $W = (w_2, \ldots, w_{J+1})'$ with $w_j \geq 0$ for $j = 2, \ldots, J+1$ and $w_2 + \cdots + w_{J+1} = 1$. Each value of W represents a potential synthetic control

Synthetic Control Method: A Motivating Model

- Let $W = (w_2, \ldots, w_{J+1})'$ with $w_j \geq 0$ for $j = 2, \ldots, J+1$ and $w_2 + \cdots + w_{J+1} = 1$. Each value of W represents a potential synthetic control
- Let $\bar{Y}_i^{K_1}, \ldots, \bar{Y}_i^{K_M}$ be M linear functions of pre-intervention outcomes

Synthetic Control Method: A Motivating Model

- Let $W = (w_2, \ldots, w_{J+1})'$ with $w_j \geq 0$ for $j = 2, \ldots, J+1$ and $w_2 + \cdots + w_{J+1} = 1$. Each value of W represents a potential synthetic control
- Let $\bar{Y}_i^{K_1}, \ldots, \bar{Y}_i^{K_M}$ be M linear functions of pre-intervention outcomes
- Suppose that we can choose W^* such that:

$$\sum_{j=2}^{J+1} w_j^* Z_j = Z_1, \; \sum_{j=2}^{J+1} w_j^* \bar{Y}_j^{K_1} = \bar{Y}_1^{K_1}, \; \cdots \;, \; \sum_{j=2}^{J+1} w_j^* \bar{Y}_j^{K_M} = \bar{Y}_1^{K_M}$$

Synthetic Control Method: A Motivating Model

- Let $W = (w_2, \dots, w_{J+1})'$ with $w_j \geq 0$ for $j = 2, \dots, J+1$ and $w_2 + \cdots + w_{J+1} = 1$. Each value of W represents a potential synthetic control
- Let $\bar{Y}_i^{K_1}, \dots, \bar{Y}_i^{K_M}$ be M linear functions of pre-intervention outcomes
- Suppose that we can choose W^* such that:

$$\sum_{j=2}^{J+1} w_j^* Z_j = Z_1, \ \sum_{j=2}^{J+1} w_j^* \bar{Y}_j^{K_1} = \bar{Y}_1^{K_1}, \ \cdots \ , \ \sum_{j=2}^{J+1} w_j^* \bar{Y}_j^{K_M} = \bar{Y}_1^{K_M}$$

- Then, for $t \in \{T_0 + 1, \dots, T\}$, an unbiased estimator of α_{1t} is given by:

$$\widehat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}$$

A Design-Based Perspective on Synthetic Control Methods

Lea Bottmer* Guido Imbens† Jann Spiess‡ Merrill Warnick§

First version: November 2020
Current version: January 2021

Abstract

Since their introduction in Abadie and Gardeazabal (2003), Synthetic Control (SC) methods have quickly become one of the leading methods for estimating causal effects in observational studies with panel data. Formal discussions often motivate SC methods by the assumption that the potential outcomes were generated by a factor model. Here we study SC methods from a design-based perspective, assuming a model for the selection of the treated unit(s), *e.g.*, random selection as guaranteed in a randomized experiment. We show that SC methods offer benefits even in settings with randomized assignment, and that the design perspective offers new insights into SC methods for observational data. A first insight is that the standard SC estimator is not unbiased under random assignment. We propose a simple modification of the SC estimator that guarantees unbiasedness in this setting and derive its exact, randomization-based, finite sample variance. We also propose an unbiased estimator for this variance. We show in settings with real data that under random assignment this Modified Unbiased Synthetic Control (MUSC) estimator can have a root mean-squared error (RMSE) that is substantially lower than that of the difference-in-means estimator. We show that such an improvement is weakly guaranteed if the treated period is similar to the other periods, for example, if the treated period was randomly selected. The improvement is most likely to be substantial if the number of pre-treatment periods is large relative to the number of control units.

Keywords: Randomization, Synthetic Controls, Panel Data, Causal Effects

SCM: Summary

- Panel data is useful in that it allows us to get to conditional independence via selection on unobservables
- Difference-in-differences designs allow us to make standard statistical inferences when we are willing to make the parallel trends assumption
- Synthetic control methods can be seen as an extension of the difference-in-differences logic for situations in which
  - Both treatment and control are measured at a high level of aggregation
  - There is a small number of, and usually only 1, treated unit
  - We have a small N with respect to potential control units
- Synthetic control methods use measures on the non-treated donor pool units to construct counterfactuals for treated units based on a weighted average
- Methods of inference for synthetic control are non-standard, but will almost surely involve some type of resampling/simulation methods

?

Thank you!

Appendix