émicos/Cursos
Dic-
ta-
dos/PUCP/QLab/Clase/”pd1enc.dfu

émicos/Cursos
Dic-
ta-
dos/PUCP/QLab/Clase/”rerunfilecheck.cfg

Synthetic Control Methods: Theory and application

Juan Manuel del Pozo Segura

The University of Sussex

j.del-pozo@sussex.ac.uk

PUCP Q Lab - Jan. 2020

Outline

1. Motivation

2. Formal aspects
    - Potential outcomes, econometric model and its bias
    - Comparison to regression and estimation of the SC
    - Inference, falsification and robustness
    - Limitations and recommendations for practitioners

SCM as a causal estimator

- As we have seen so far in the course, the most important questions in economics usually involve analysing **causal questions**.
  - In some cases, these causal questions can be "easily" answered by an RCT
  - Unfortunately, in most cases, causal questions are not and *cannot* be answered via this golden standard. E.g.
    - Smoking and cancer
    - Immigration and wages

  Consequently, we need to rely on **quasi experiments**
- Depending on the nature of the data and on the phenomenon under analysis, we can use different **identification strategies**
  - This class is concerned with 1 of those: Abadie's **Synthetic Control Method (SCM)**
  - *"The synthetic control approach developed by Abadie et al. [2010, 2015] and Abadie and Gardeazabal [2003] is arguably the most important innovation in the policy evaluation literature in the last 15 years."*
    Susan Athey and Guido Imbens (2017)

The type of interventions of interest here

- We are interested in **comparative case studies** where
  1. The treatment happens on **aggregated entities** (such as states)
  2. We have to compare the evolution of aggregate outcomes
     - for units affected by a particular occurrence of the event or intervention of interest
     - for some control group of unaffected units
- E.g.
  - Card (1990): impact of the 1980 Mariel Boatlift using other cities in the southern United States as comparison group.
  - Card and Krueger (1994): evolution of employment in fast food restaurants in NJ and its PN around the time of an increase in NJ's minimum wage.
  - Abadie and Gardeazabal (2003): evolution in GDP in Basque Country and other Spanish regions due to terrorist conflict in the former
- These studies are feasible because
  1. many policy interventions of interest in the social sciences take place at an *aggregate* level
  2. widespread availability of data for units affected by the event of interest and a set of unaffected units
     - Many times we have data on a sample of disaggregated units
     - In case we do not, we only require aggregate data

Why should we bother learning a method other than DiD?

- We now know that we can analyse case studies like these via **DiD** when we have data both
  1. for the period before and after the quasi-experiment
  2. for a (set of) treated an control units
- Advantages:
  - Simple, powerful and no strict need of panel: works also with even with RCS
  - Vast literature on how to correctly analyse inference issues (**Cameron & Miller 2015** and references there, Bertrand, M. et al. 2002, Brewer, M. et al. 2017)
  - Modifications such as 3D, 4D that relax the PTA assumption for identification
- Disadvantages
  1. What units we choose as controls?
  2. With small G and G_1 (as usually is) we have small power and we need to work hard on inference (Cameron & Miller 2015, Mackinnon, J. & Webb, M. 2016 and see also Hansen, B. (2007) for the policy autocorrelation problem)
  3. Strong reliance on the PTA assumption for identification (Lee 2016, Angrist & Pischke 2008, Imbens & Wooldridge 2015)
- These are, precisely, the problems that SCM solves! However, there is no free lunch and it also has some limitations. Let's see each of these in turn.

What units we choose as controls?

- One of the seminal papers for DiD comes from Card (1990)
  - This studies the impact of the 1980 Mariel Boatlift when approximately 125,000 Cubans emigrated to Florida over this 6 month period of time.
  - Card saw this as an **exogenous shift** in the labor supply curve.
    - treated area: Miami
    - control group: Atlanta, Los Angeles, Houston and Tampa-St. Pbrg. *The choice of these cities is in a footnote in the paper*: “similar based on
    1. demographics
    2. economic conditions”

  Card estimated a **simple DD** model and found no effect
- However, Angrist & Krueger (1999) show the risk of inference from analyzing an event with a small number of treatment and control units.
  - Analyze as Card (1990) a “non-existent" Mariel Boatlift in *1994*
    - In 1994 Castro again announced that Cubans who wanted to leave, could leave.
    - So, a big inflow of refugees to Miami was about to take place but *did not.*
  - They show that between 1993 (pre-non-shock) and 1995 (post-non-shock) unemployment rate for Black workers
    - in Miami increased by 3.6 pp
    - in control group of cities in Card (1990) decreased by 2.7 pp.
  - Hence, *we would estimate a* **fake treatment effect** *of 6.3* pp

The problem with small G ...

- Cameron & Miller (2015) (and the last chapter of Lee 2016) greatly discuss the inference issues around FE, as well as what to cluster over.
- In the **standard setting** when $G \to \infty$, we can use
  - the Liang & Zeger CRVE $\hat{V}_{clu}\left[\hat{\beta}\right] = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\sum_{g=1}^{G}\boldsymbol{X}_g'\hat{\boldsymbol{u}}_g\hat{\boldsymbol{u}}_g'\boldsymbol{X}_g\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$ (, `vce(cluster clustervar)`) which
    - use T_{G-1} critical values
    - replace $\hat{\boldsymbol{u}}_g$ by $\hat{\boldsymbol{u}}_g \times \sqrt{\frac{G}{G-1}\frac{N-1}{N-K}}$

    even after including the regions FEs dummies ◂ DiD Clusters (G)
  - pairs cluster bootstrap
- However, an important part of this paper is devoted to the problem of small G (e.g.: Peru, with $G = 24$). In this case, the asymptotics in L&Z CRVU do not work. We need to painstakingly take other routes
  - Bell & McCaffrey (2002) CR2VE and CR3VE bias corrections
  - Wild cluster bootstrap method (Rademacher and/or Webb weights)
  - Donald & Lang (2007) T_{G-L} distribution
  - Imbens & Kolesar (2012) effective DoF and Carter et al. (2017) effective G

...and to put matters worse: the problem with small G_1

- MacKinnon & Webb (2016) emphasize another usually overlooked aspect in DiD inference with large G: L&Z CRVE assumes
  - A1. $G \rightarrow \infty$
  - A2. The within-cluster error correlations are the same for all $g \in G$
  - A3. Each $g \in G$ contains an equal number of observations ($N_g, \forall g$)
- So, *even with large G, L&Z can fail* because the 'rule of $G = 42$' no longer holds when assumption 3 relaxed. They find that in the **DiD case** with dichotomous regressors that
  - N_g matters and in cases with large imbalances, the wild bootstrap works reasonably well
  - But unfortunately, all the methods fail *badly* when G_1 is small or (in some cases) large

PTA assumption for identification

- The potential outcome for a *non treated observation* in a given state s and time t is
$$E\left(Y_i^0|s,t\right) = \alpha_s + \tau_t$$
in this case, Y will be determined by the sum of
1. a time-invariant state fixed effect, γ_s idiosyncratic to the state
2. a time effect τ_t that is common across all states
- Under the **conditional independence assumption**, the **average treatment effect** is $E\left(Y_i^1 - Y_i^0|s,t\right) = \beta_d$. So, the observed outcome can be written as
$$Y_i = \underbrace{\alpha_s + \tau_t}_{E\left(Y_i^0|s,t\right)} + \beta_d D_i + \varepsilon_{ist}$$
$\alpha_s + \tau_t$ represents the PTA; its fulfillment makes the DiD valid ◂ DiD PTA
- Given this, β_d will give us the causal effect if we include, along with D_{st}, dummies for states s and periods t. This *will work* because these latter will wipe out the FEs α_s and τ_t in this linear setting,
- However, since the validity of this depends on the PTA and this in turn depends on the fact that FEs enter additively, our 2-way FE will not yield the causal effect if the FEs are in fact, e.g., interacted.

What does this mean for DiD estimator?

- DiD is *still* a very useful and powerful tool for analysis. However, the circumstances where it is weak are staples of economic quasi-experiment. In many cases
  1. The treatment happens on **aggregated entities** (such as states) on a population of interest with small G
  2. In several important cases, the treatment occurs on $G_1 = 1$ or $G_1 =$ very low
  3. In an attempt to escape the small G problem we take all the non treated units as control group
  4. How do we know that the additivity assumption for the FEs actually holds?
- Following Abadie et al. (2010), a seminal paper:
  - There is uncertainty about the ability of the subjectively-chosen control group to reproduce the counterfactual outcome trajectory that the affected units would have experienced in the absence of the intervention or event of interest.
  - This uncertainty is *not* reflected by the standard errors constructed with traditional inferential techniques for comparative case studies

How does SCM help in overcoming this?

- In this seminal paper, Abadie et al. advocate the use of data-driven procedures to construct suitable comparison groups to reduce discretion in the choice of the comparison control units
  - It is *difficult to find 1 unexposed unit* that approximates nest the unit(s) exposed to the event of interest.
  - So, the core the idea behind the SCM is to take as **counterfactual** a **Synthetic Control** (**SC**), a combination of units which provides a *better* comparison for the unit exposed to the intervention than any single unit alone.
- E.g.
  - Abadie and Gardeazabal (2003):they use a combination of 2 Spanish regions to approximate the economic growth that the Basque Country would have experienced in the absence of terrorism.
  - Peri and Yasenov (2019): use a combination of cities in USA to approximate the evolution that the Miami labor market would have experienced in the absence of the Mariel Boatlift.

How does SCM help in overcoming this?

- SCM has 3 attractive features relative to traditional regression methods
  1. **transparency:** Because the SC is a weighted average of the available control units, the SCM makes explicit:
     - the *relative contribution* of each control unit to the counterfactual of interest
     - the *similarities* (or lack thereof) between the unit affected by the intervention of interest and the SC, in terms of

     1.1 preintervention outcomes

     1.2 predictors of postintervention outcomes
  2. safeguard against **extrapolation**:
     - the weights that make up the SC, the counterfactual, sum up to 1
     - so by using as counterfactual a convex hull of control group units, it is based on where data *actually is*
  3. construction of the SC does **not require access to the post-treatment outcomes**.
- Importantly, the model extends the traditional linear difference-in-differences framework. This allows that the effects of unobserved variables on the outcome vary with time
- However, we need to know when this method can actually be implemented

Outline

1. Motivation

2. Formal aspects
    - Potential outcomes, econometric model and its bias
    - Comparison to regression and estimation of the SC
    - Inference, falsification and robustness
    - Limitations and recommendations for practitioners

A motivating model

- In this section we follow Abadie et al. (2010), and we will supplement it with Abadie et al. (2015) and Abadie (2020)
- To simplify the exposition, we assume that
  - there are $J+1$ aggregated units (we could aggregate the data from the regions exposed to the intervention)
    - **only 1 unit or region is subject to the intervention of interest**
    - there are J unexposed units, called the **donor pool (DP)**
  - there are T periods: $\underbrace{1,2,...,T_0}_{\text{pre-treatment}}; \underbrace{T_0+1,T_0+2,...,T}_{\text{post-treatment}}$
- In terms of the potential outcomes

1. Y_{jt}^N is the *outcome* that would be observed for
   1.1 $j=2,...,J+1$ in $t=1,...,T$, i.e. at all periods for the untreated units
   1.2 $j=1$ in $t=1,...,T_0$, i.e. where there is not intervention for the treated unit
2. Y_{jt}^I is the *outcome* that would be observed for $j=1$ in $t=T_0+1,..,T$, i.e. where there is exposure to the intervention for the treated unit

We assume the intervention has *no effect* on the outcome before T_0+1

- We still hold the **SUTVA**: Y_i, $i=2,...,J+1$ are not affected by intervention in $i=1$

The counterfactual and the treatment effect

- The **effect of the intervention** for unit $j = 1$ at *every year t* is given by

  $$\boxed{\alpha_{1t} = Y_{1t}^{I} - Y_{1t}^{N}}$$

  Since $Y_{1t} = Y_{1t}^{I}$ for $t > T_0$, then for $t > T_0$

  $$\alpha_{1t} = Y_{1t}^{I} - Y_{1t}^{N} = Y_{1t} - \boxed{Y_{1t}^{N}}$$

  Because Y_{1t} is observed, the challenge is estimating the **counterfactual** Y_{1t}^{N}. It is a counterfactual because

  - At $t > T_0$ the treatment for $j = 1$ *already happened* and then we observe Y_{1t}^{I}
  - So, at $t > T_0$ we estimate Y_{1t}^{N} for $j = 1$ based on the other units. This is what SCM focuses on!

- Since $\alpha_{1t} = Y_{1t} - Y_{1t}^{N} \Leftrightarrow Y_{1t} = Y_{1t}^{N} + \alpha_{1t}$, we can define D_{it} as a dummie:

  - 1 if $j = 1$ and is exposed to the intervention at year t $(t = T_0 + 1, ..., T)$
  - 0 otherwise

  So, the observed outcome for unit i at any year t is

  $$\boxed{Y_{jt} = Y_{jt}^{N} + \alpha_{jt} D_{jt}}$$

The underlying model (the big strength!)

- Suppose that the counterfactual Y_{jt}^{N} in $Y_{jt} = Y_{jt}^{N} + \alpha_{jt} D_{jt}$ can be estimated using a **linear factor model**, a generalization of DiD/FE models

$$\boxed{Y_{jt}^{N} = \delta_t + \underbrace{(\boldsymbol{\theta}_t)_{(1\times r)} (\boldsymbol{Z}_j)_{r\times 1}}_{\text{this can change every year}} + \underbrace{(\boldsymbol{\lambda}_t)_{(1\times F)} (\boldsymbol{\mu}_j)_{(F\times 1)}}_{\text{unobserved}} + \varepsilon_{jt}}$$

  where
  - δ_t is an common factor with constant factor loadings across units (i.e. time FEs in DiD)
  - $\boldsymbol{Z}_j$ is a vector of observed covariates (not affected by the intervention) and $\boldsymbol{\theta}_t$ is a vector of unknown parameters (i.e. different effects of Z_i for different years)
  - ε_{jt} are unobserved transitory shocks at the region level with $E(\varepsilon_{it}) = 0$
  - $\boldsymbol{\lambda}_t$ is a vector of unobserved **common factors** that change in time and $\boldsymbol{\mu}_j$ is a vector of unknown **factor loadings**. This is where the flexibility comes from!
- This is the framework Abadie et al. (2010) use to study the **bias properties of synthetic controls estimators**.

What does the factor model actually allows for?

- The key to understand the advantage of this model lies in $(\boldsymbol{\lambda}_t)_{(1\times F)}(\boldsymbol{\mu}_j)_{(F\times 1)}$. Following Bai (2009), we can express this as

$$(\boldsymbol{\lambda}_t)_{(1\times F)}\left(\boldsymbol{\mu}_j\right)_{(F\times 1)} = \lambda_{1t}\mu_{1j} + \lambda_{2t}\mu_{2j} + \ldots + \lambda_{Ft}\mu_{Fj}$$

- On the one hand, if $F = 2$, so that $\boldsymbol{\lambda}_t = \begin{bmatrix} 1 & \tau_t \end{bmatrix}$ and $\boldsymbol{\mu}_j = \begin{bmatrix} 1 & \alpha_j \end{bmatrix}'$ then $(\boldsymbol{\lambda}_t)_{(1\times 2)}\left(\boldsymbol{\mu}_j\right)_{(2\times 1)} = \alpha_i + \tau_t$ and so we are in the canonical DiD model

$$Y_{jt}^N = \delta_t + (\boldsymbol{\theta}_t)_{(1\times r)}\left(\boldsymbol{Z}_j\right)_{r\times 1} + \alpha_j + \tau_t + \varepsilon_{it}$$

  and *only* if this is the case, then these FEs can be eliminated by taking time differences. I.e. the canonical FE model

  1. allows for the presence of unobserved confounders
  2. *restricts* the effect of those confounders to be constant in time

- On the other hand, in the factor model the FEs can enter in the model interactively and not only additively. Hence, if this is the case, these FEs *cannot* be eliminated by taking time differences. I.e. the factor model

  1. allows for the presence of unobserved confounders
  2. allows the effects of confounding unobserved characteristics to *vary with time*

How to create the SC based on the factor model

- How good is this flexible model for estimating the counterfactual? Remember that SCM estimates this as a SC based on $\sum_{j=2}^{J+1} w_j Y_{jt}^N$, a weighted average of units in the DP which *best* approximates the outcome in $j = 1$
- Hence, we need a vector of weights $\boldsymbol{W}$ which will allow creating the SC

$$\boldsymbol{W} = (w_2, ..., w_{J+1})' \text{ s.t.} 1)\ w_j \geq 0 \text{ for } j = 2, ..., J+1$$
$$2)\ w_2 + \cdots + w_{J+1} = 1$$

Note that

- Once we get these $w_j; j = 2, ..., J+1$ we create the SC calculating using $j = 2, ..., J+1$ in the DP.

$$\begin{aligned}
\sum_{j=2}^{J+1} w_j Y_{jt}^N &= \sum_{j=2}^{J+1} w_j \left(\delta_t + \boldsymbol{\theta}_t \boldsymbol{Z}_j + \boldsymbol{\lambda}_t \boldsymbol{\mu}_j + \varepsilon_{it}\right) \\
&= \sum_{j=2}^{J+1} w_j \delta_t + \sum_{j=2}^{J+1} w_j \boldsymbol{\theta}_t \boldsymbol{Z}_j + \sum_{j=2}^{J+1} w_j \boldsymbol{\lambda}_t \boldsymbol{\mu}_j + \sum_{j=2}^{J+1} w_j \varepsilon_{jt} \\
&= \delta_t \underbrace{\sum_{j=2}^{J+1} w_j}_{=1} + \boldsymbol{\theta}_t \sum_{j=2}^{J+1} w_j \boldsymbol{Z}_j + \boldsymbol{\lambda}_t \sum_{j=2}^{J+1} w_j \boldsymbol{\mu}_j + \sum_{j=2}^{J+1} w_j \varepsilon_{jt} = \delta_t + \boldsymbol{\theta}_t \sum_{j=2}^{J+1} w_j \boldsymbol{Z}_j + \boldsymbol{\lambda}_t \sum_{j=2}^{J+1} w_j \boldsymbol{\mu}_j + \sum_{j=2}^{J+1} w_j \varepsilon_{jt}
\end{aligned}$$

- The 2 restrictions assure **sparse SC**, i.e. made of small number of *j*s in the DP

The bias of the estimator for the counterfactual

- To derive the **bias** Abadie calculates first

$$\underbrace{Y_{1t}^N}_{\text{counterfactual}} - \underbrace{\sum_{j=2}^{J+1} w_j Y_{jt}^N}_{\text{estimator}} = (\delta_t + \boldsymbol{\theta}_t \boldsymbol{Z}_1 + \boldsymbol{\lambda}_t \boldsymbol{\mu}_1 + \varepsilon_{1t}) - \left(\delta_t + \boldsymbol{\theta}_t \sum_{j=2}^{J+1} w_j \boldsymbol{Z}_j + \boldsymbol{\lambda}_t \sum_{j=2}^{J+1} w_j \boldsymbol{\mu}_j + \sum_{j=2}^{J+1} w_j \varepsilon_{jt}\right)$$

$$= \boldsymbol{\theta}_t \left(\boldsymbol{Z}_1 - \sum_{j=2}^{J+1} w_j \boldsymbol{Z}_j\right) + \boldsymbol{\lambda}_t \left(\boldsymbol{\mu}_1 - \sum_{j=2}^{J+1} w_j \boldsymbol{\mu}_j\right) + \left(\underbrace{(w_2 + ... + w_{J+1})}_{=1} \varepsilon_{1t} - \sum_{j=2}^{J+1} w_j \varepsilon_{jt}\right)$$

$$= \boldsymbol{\theta}_t \left(\boldsymbol{Z}_1 - \sum_{j=2}^{J+1} w_j \boldsymbol{Z}_j\right) + \boldsymbol{\lambda}_t \left(\boldsymbol{\mu}_1 - \sum_{j=2}^{J+1} w_j \boldsymbol{\mu}_j\right) + \sum_{j=2}^{J+1} w_j \left(\varepsilon_{1t} - \varepsilon_{jt}\right)$$

- After reducing this expression, we need to assume the existence of **optimal weights** $\left(w_2^*, ..., w_{J+1}^*\right)$ such that for the pre-intervention period

$$\underbrace{\sum_{j=2}^{J+1} w_j^* Y_{jt}}_{\text{donors}} = w_2^* Y_{2t} + w_3^* Y_{3t} + ... + w_{J+1}^* Y_{J+1,t} = \underbrace{Y_{1t}}_{\text{treated}} ; t = \underbrace{1, ..., T_0}_{\text{pre-intervention}}$$

$$\underbrace{\sum_{j=2}^{J+1} w_j^* \boldsymbol{Z}_j}_{\text{donors}} = w_2^* \boldsymbol{Z}_2 + w_3^* \boldsymbol{Z}_3 + ... + w_{J+1}^* \boldsymbol{Z}_{J+1} = \underbrace{\boldsymbol{Z}_1}_{\text{treated}} ; t = \underbrace{1, ..., T_0}_{\text{pre-intervention}}$$

which allows us to write

$$\underbrace{Y_{1t}^N}_{} - \sum_{j=2}^{J+1} w_j^* Y_{jt}^N = \sum_{j=2}^{J+1} w_j^* \left(\varepsilon_{1t} - \varepsilon_{jt}\right) - \boldsymbol{\lambda}_t \left(\boldsymbol{\lambda}^{P\prime} \boldsymbol{\lambda}^P\right)^{-1} \boldsymbol{\lambda}^{P\prime} \boldsymbol{\varepsilon}_1^P + \boldsymbol{\lambda}_t \left(\boldsymbol{\lambda}^{P\prime} \boldsymbol{\lambda}^P\right)^{-1} \boldsymbol{\lambda}^{P\prime} \sum_{j=2}^{J+1} w_j^* \boldsymbol{\varepsilon}_j^P$$

The bias of the estimator for the counterfactual

- So, when $w_j = w_j^*, j = 2, ..., J+1$, the **bias** for $t > T_0$ is

$$E\left[\underbrace{Y_{1t}^N}_{\text{counterfactual}} - \underbrace{\sum_{j=2}^{J+1} w_j^* Y_{jt}^N}_{\text{estimator}}\right] = \underbrace{E\left(\sum_{j=2}^{J+1} w_j^* \left(\varepsilon_{1t} - \varepsilon_{jt}\right)\right)}_{=E(R_{3t})=0} - \underbrace{E\left(\boldsymbol{\lambda}_t \left(\boldsymbol{\lambda}^{P\prime}\boldsymbol{\lambda}^P\right)^{-1} \boldsymbol{\lambda}^{P\prime} \boldsymbol{\varepsilon}_1^P\right)}_{=E(R_{2t})=0} + \underbrace{E\left(\boldsymbol{\lambda}_t \left(\boldsymbol{\lambda}^{P\prime}\boldsymbol{\lambda}^P\right)^{-1} \boldsymbol{\lambda}^{P\prime} \sum_{j=2}^{J+1} w_j^* \boldsymbol{\varepsilon}_j^P\right)}_{=\boxed{E(R_{1t}) \neq 0}}$$

- Applying the Cauchy–Schwarz, the Hölder and the Rosenthal inequality, this bias is **bounded by**, i.e. *is at most*

$$\boxed{E\left[\underbrace{Y_{1t}^N}_{\text{counterfactual}} - \underbrace{\sum_{j=2}^{J+1} w_j^* Y_{jt}^N}_{\text{estimator}}\right] = E(R_{1t}) \leq \left[C(P)^{\frac{1}{p}} \left(\frac{\overline{\lambda}^2 F}{\xi}\right) \boxed{J^{\frac{1}{p}}}\right] \times max \left[\frac{\boxed{\overline{m}_p^{\frac{1}{p}}}}{\boxed{T_0^{1-\frac{1}{p}}}}, \frac{\overline{\sigma}}{\boxed{T_0^{\frac{1}{2}}}}\right]}$$

where $\overline{m}_p = \left[\max_{j=2,...,J+1} \left(\frac{1}{T_0} \sum_{t=1}^{T_0} E|\varepsilon_{jt}|^p\right)\right]$

- So, if T_0 is large relative to the scale of ε_{jt}, then the bias $E\left[Y_{1t}^N - \sum_{j=2}^{J+1} w_j^* Y_{jt}^N\right] \approx 0$ and so an unbiased estimator of α_{1t} is

$$\boxed{\hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}; t \in \{T_0 + 1, ..., T\}}$$

What we need for this nice result to happen

- For this result to hold, we required the existence of the optimal weights $\left(w_2^*, ..., w_{J+1}^*\right)$ that create a **perfect fit** in terms of Y and covariates $\boldsymbol{Z}$ in the pre-intervention period: $\sum_{j=2}^{J+1} w_j^* Y_{jt} = Y_{1t}$ and $\sum_{j=2}^{J+1} w_j^* \boldsymbol{Z}_j = \boldsymbol{Z}_1$ for $t = 1, ..., T_0$. This happens ***only if***

  $\left(Y_{11}, ..., Y_{1T_0}, \boldsymbol{Z}_1\right)$ **belongs** to the convex hull of $\left\{\left(Y_{21}, ..., Y_{2T_0}, \boldsymbol{Z}_2\right), ..., \left(Y_{J+1,1}, ..., Y_{J+1,T_0}, \boldsymbol{Z}_{J+1}\right)\right\}$

- However, in practice
  - it is *often the case* that $\sum_{j=2}^{J+1} w_j^* Y_{jt} \approx Y_{1t}$ and $\sum_{j=2}^{J+1} w_j^* \boldsymbol{Z}_j \approx \boldsymbol{Z}_1$ but still the bias bound *kicks in*
  - in some cases
    - $\sum_{j=2}^{J+1} w_j^* Y_{jt} << Y_{1t}$ or $\sum_{j=2}^{J+1} w_j^* Y_{jt} >> Y_{1t}$, and
    - $\sum_{j=2}^{J+1} w_j^* \boldsymbol{Z}_j << \boldsymbol{Z}_1$ or $\sum_{j=2}^{J+1} w_j^* \boldsymbol{Z}_j >> \boldsymbol{Z}_1$

    and so there is a *large bias* in the estimation so that *it is not recommended using the SCM*

- Hence, it is important to calculate the magnitude of $Y_{1t} - \sum_{j=2}^{J+1} w_j^* \boldsymbol{Z}_j = \boldsymbol{Z}_1$ between $j = 1$ and the SC to decide if the characteristics of the treated unit are sufficiently matched by the SC

◂ Example of good pre-treat. covariate balance

How the method can control for unmeasured factors

- Abadie (2020) states that, under the factor model: A SC that reproduces Z_1 *and* $\boldsymbol{\mu}_1$ would provide an *unbiased* estimator of the treatment effect for the treated ◂ Linear factor model
  - However, $\boldsymbol{\mu}_1$ is not observed so it cannot be matched directly in the data
  - Hence, a SC *can* reproduce the values of Z_1 but *not those of* $\boldsymbol{\mu}_1$. I.e. we can have a close match for pre-treatment outcomes for pre-treatment outcomes but still the SC does not match the values of μ_1
- Hence
  1. The ability of SC to reproduce the trajectory of the Y_1 over an extended period of time provides an indication of low bias ◂ Example of good pre-treatment fit
  2. Large T_0 cannot drive down the bias if the fit is bad. In fact, in practice, a SC may *not perfectly fit* the characteristics of the treated units

J and the bias

- Under a factor model for Y_{it}^{N} , larger J
  - Makes it easier to fit pre-treatment outcomes even when there are substantial discrepancies in $\boldsymbol{\mu}$s between the $j = 1$ and the SC
  - However, the bias bound depends positively on J because a large number of units in the DP may create or exacerbate the bias of the estimator, especially if the $\boldsymbol{\mu}_j$ in the DP greatly differ from $\boldsymbol{\mu}_1$ ◂ Bias bound
- A practical implication of this is that *each of the units in the DP have to be chosen judiciously* to provide a reasonable control for the treated unit. So that units in the DP should have
  - similar values of the observed attributes Z_i relative to the treated unit
  - similar values of *the unobserved attributes* $\boldsymbol{\mu}_j$ relative to the treated unit

Outline

1. Motivation

2. Formal aspects
   - Potential outcomes, econometric model and its bias
   - Comparison to regression and estimation of the SC
   - Inference, falsification and robustness
   - Limitations and recommendations for practitioners

The matrices with pre-intervention characteristics

- $\boldsymbol{W} = (w_2, ..., w_{J+1})^{'}$ s.t. $w_j \geq 0$ for $j = 2, ..., J+1$ and $w_2 + \cdots + w_{J+1} = 1$
  - As before, each possible W represents a **SC**, i.e. a *weighted average of the units in the DP*
  - The 2 constraints imply that the SC is a **convex combination** of untreated units so to avoid **extrapolation**
- Let
  - $\boldsymbol{Z}_j$ is a $(r \times 1)$ vector of **observed covariates** for j
  - $\left(\overline{Y}_j^{K_1}, ..., \overline{Y}_j^{K_M}\right)$ is a $(M \times 1)$ vector of **observed values** for Y for j

so

- $\boldsymbol{X}_1 = \left(\begin{array}{c} \boldsymbol{Z}_1 \\ \hline \overline{Y}_1^{K_1} \\ ... \\ \overline{Y}_1^{K_M} \end{array}\right)_{k \times 1 = (r+M) \times 1}$ a vector of ***preintervention*** characteristics for $j = 1$
- $\boldsymbol{X}_0 = \left(\begin{array}{cccc} \boldsymbol{Z}_2 & \boldsymbol{Z}_3 & ... & \boldsymbol{Z}_{J+1} \\ \hline \overline{Y}_2^{K_1} & \overline{Y}_3^{K_1} & ... & \overline{Y}_{J+1}^{K_1} \\ ... & ... & & ... \\ \overline{Y}_2^{K_M} & \overline{Y}_3^{K_M} & ... & \overline{Y}_{J+1}^{K_M} \end{array}\right)_{k \times J = (r+M) \times J}$ is a matrix of ***preintervention*** characteristics for js in the DP

◂ Example of X matrices

Comparison to regression-based counterfactuals

- SCM constructs a SC as a linear combination of units in DP under 2 restrictions. However, *also* **regression** creates weighted counterfactuals!

$$\left(\hat{\beta}'\right)_{T_1\times k}(X_1)_{k\times 1} = \left(\left(X_0X_0'\right)^{-1}X_0Y_0'\right)'X_1 = Y_0\left[X_0'\left(X_0X_0'\right)^{-1}X_1\right] = (Y_0)_{T_1\times J}(W^{reg})_{J\times 1}$$

  where

  - X_0 is a $k\times J$ matrix with *preintervention* characteristics for the units in DP
  - X_1 is a $k\times 1$ vector with *preintervention* characteristics for $j=1$

- The sum of the weights of this R-B counterfactual is given by

$$\left(\iota'\right)_{1\times J}(W^{reg})_{J\times 1} = \iota' X_0'\left(X_0X_0'\right)^{-1}X_1 = \left(\left(X_0X_0'\right)^{-1}X_0\iota\right)'X_1$$

- Assume that, as usual, regression includes an intercept so the first row of X_0 is a vector of 1s. Then, because $\left(X_0X_0'\right)^{-1}X_{0\iota}$ can be seen as the $\hat{\beta}$s of the regression ι (a vector of 1s) on X_0, the only non-0 coefficient is the intercept (=1). So, *mechanically*,

$$\left(\left(X_0X_0'\right)^{-1}X_0\iota\right)_{K\times 1} = \begin{pmatrix} 1 & 0 & \ldots & 0 \end{pmatrix}$$

Comparison to regression-based counterfactuals

- The latter implies that the sum of weights is

$$\left(\boldsymbol{\iota}'\right)_{1\times J}(W^{reg})_{J\times 1} = \boldsymbol{\iota}' X_0' \left(X_0 X_0'\right)^{-1} X_1 = \left(\left(X_0 X_0'\right)^{-1} X_0 \boldsymbol{\iota}\right)' X_1 = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} X_1 = 1$$

  so, the regression-based counterfactuals are created using
  - weights that sum to 1
  - *not restricted to be between 0 and 1*: may take negative values
- As a result, regression-based counterf. can lead to extrapolation outside the support of the comparison units. This implies that
  - regression weights **extrapolate** to produce a perfect fit *even if* the Xs of $j=1$ cannot be approximated by a weighted average of the Xs of js in the DP
  - Technically, even if X_1 is far from the convex hull of the columns of X_0, regression weights extrapolate to produce

$$X_0 W_{reg} = X_0 X_0' \left(X_0 X_0'\right)^{-1} X_1 = X_1$$

- So, usually,
  - regression-based counterfactual relies on **extrapolation**
  - Instead, the SCM-based counterfactual
    1. closely fits the values of the characteristics of the units
    2. does not extrapolate outside of the support of the data

The minimization process to find W^*

- Given this, the vector $\boldsymbol{W}^*$ is chosen to minimize the MSE of the SC

$$\boldsymbol{W}^* = \underset{W}{\text{argmin}} \|\boldsymbol{X}_1 - \boldsymbol{X}_0 \boldsymbol{W}\|_V \equiv (\boldsymbol{X}_1 - \boldsymbol{X}_0 \boldsymbol{W})' \boldsymbol{V} (\boldsymbol{X}_1 - \boldsymbol{X}_0 \boldsymbol{W}) = \sum_{m=1}^{k} v_m \left(\boldsymbol{X}_{1m} - \sum_{j=2}^{J+1} w_j \boldsymbol{X}_{jm} \right)^2$$

$$\text{s.t.} \, w_2 \geq 0, ..., w_{J+1} \geq 0$$

$$w_2 + ... + w_{J+1} = 1$$

where $\boldsymbol{V} = [v_m], m = 1, ..., k$ is a $k \times k = (r+M) \times (r+M)$ **diagonal** symmetric and positive semidefinite matrix. Each v_m assigns a weight to every covariate m ($\boldsymbol{Z}$ or $\overline{Y}_2^{K_1} ... \overline{Y}_2^{K_M}$) to indicate its importance in forming the SC

- Since each potential choice of $V = (v_1, ..., v_k)$ produces a synthetic control $\boldsymbol{W}^* \equiv \boldsymbol{W}^*(\boldsymbol{V})$, we need to take care on finding a reasonable V^*

The minimization process to find V^*

- The choice of Vcan be subjective (not recommended) or data-driven

1. **minimization**: Abadie & Gardeazabal (2003) suggest V such that the SC (defined by the$W^*(V)$ we just found) approximates the trajectory of Y_1 *only* in the *preintervention* periods $t = 1, ..., T_0$

   $$\boldsymbol{V}^* = \underset{V}{\operatorname{argmin}} (\boldsymbol{Y}_1 - \boldsymbol{Y}_0 \boldsymbol{W}^*(\boldsymbol{V}))^{'} (\boldsymbol{Y}_1 - \boldsymbol{Y}_0 \boldsymbol{W}^*(\boldsymbol{V})) = \sum_{t=1}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j^*(V) Y_{jt} \right)^2$$

   where V is the set of all non-negative diagonal $(K \times K)$ matrices. We use this to derive again the W^* matrix

2. **regression-based**: we for every t, we regress Y_j on $\sum_{k=1}^{r+M} \beta_k Z_{kj}$ and then we apply Kaul et al. (2018) formula

3. **cross validation**: if T_0 *is large enough* we can divide pre-intervention periods into

   3.1 Initial **training period**: given a V, we can compute using only data from this period $W^*(V)$

   3.2 Subsequent **validation period**: V minimizes MSPE produced by the weights $W^*(V)$ during the validation period.

Outline

1. Motivation

2. Formal aspects
   - Potential outcomes, econometric model and its bias
   - Comparison to regression and estimation of the SC
   - Inference, falsification and robustness
   - Limitations and recommendations for practitioners

What does inference mean in this aggregate case

- The SEs commonly reported in regression-based comparative case studies reflect only the unavailability of aggregate data
- However, despite that aggregate data is used for estimation in SCM, there is still uncertainty about the value of the parameters of interest.
  - Still there is uncertainty from ignorance about the ability of the control group to reproduce the counterfactual
  - So, using individual micro data increases the total amount of uncertainty if the outcome of interest is an aggregate.
- Large sample inferential techniques are *not well suited to comparative case studies* when the number of units in the comparison group is small.
  - Abadie proposes **exact inferential techniques**, akin to **permutation tests**, in which the distribution of a test statistic is computed under random permutations assigning units to intervention and nonintervention groups
  - This inferential exercise is *exact* in the sense that regardless of the J, T or whether the data are individual or aggregate, it is always possible to calculate the exact distribution of the estimated effect of the placebo interventions.
- Under the H_0 of no intervention effect, the estimated effect for $j = 1$ is not expected to be abnormal relative to the **placebo effects distribution**.

Permutations tests (in-space placebos)

- Take a j in the DP and compute ◂ Example of permutations test
  1. the W^*matrix taking j as the treated and shifting $j=1$ to the DP and, based on $\hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}$, calculate
  $$\hat{\alpha}_{jt} = Y_{jt} - \sum_{k=1}^{J+1} w_k^* Y_{kt} \text{ with } k = 1,\dots,J+1 \text{ without the current} j$$
  2. the RMSPE for the *pre-treatment period*:
  $$RMSPE_j^{pre} = \sqrt{\left(\frac{1}{T_0}\sum_{t=1}^{T_0}\left(Y_{jt} - \sum_{k=1}^{J+1} w_k^* Y_{kt}\right)^2\right)} \text{ with } k = 1,\dots,J+1 \text{ without the current} j$$

  measures *lack of fit* between path of Y_{jt} and its SC in *pre-intervention period*

  Repeating this $\forall j$ in the DP provides the **distribution** of estimated effects
- Based on this, we assess *graphically* if the effect estimated for $j=1$ is large relative to the distrib. of the placebo effects (for the js in the DP).
- But since series with poor fit in pre-treatment period *do not provide good information* to measure the significance of the effect for $j=1$, we need
  1. to create a graphic with the effects of all js ($j=1$ and the placebos)
  2. to create graphics with the effects of *only those js with* $\frac{RMSPE_j^{pre}}{RMSPE_{\boxed{1}}^{pre}} < cutoff$

P-values

- Abadie et al. (2010) also shows how to construct **exact p-values** which obviates choosing a cut-off for the exclusion of ill-fitting placebos.
  - This measures the *qua*lity of the fit of the SC a j in the post-treatment relative to the quality of the fit in the pre-treatment period
  - Consists on looking at the distribution of **ratios of post and pre treatment period MSPEs** because a *large postintervention* RMSPE is *not* indicative of a large effect of the intervention if the preintervention RMSPE is also large, i.e. if the SC does *not* closely reproduce Y_{jt} *in the pre-treatment period*
- To do so ◂ Example of p-values

1. Take a j in the DP and compute
   1.1 the W^*matrix taking j as the treated and putting $j = 1$ into the DP
   1.2 the ratio of the post-to-pre-treatment RMSPE, with

$$RMSPE_j^{post} = \sqrt{\left(\frac{1}{T - T_0} \sum_{t=T_0+t}^{T} \left(Y_{jt} - \sum_{k=1}^{J+1} w_k^* Y_{kt} \right)^2 \right)} \text{ with } k = 1, ..., J+1 \text{ without the current} j$$

2. Repeat this $\forall j$ in the DP to get the distribution of estimated effects
3. Sort the ratios in descending order from greatest to highest
4. Calculate the **treatment unit's ratio** in the distribution as

$$p = \frac{RANK}{TOTAL\ OF\ OBSERVATIONS}$$

Falsification: back casting (in-time placebos)

- As usual, an important part of the applied work is concerned with *validating* our identification strategy. One way to do this concerns the timing of the intervention (Abadie et al. 2015)
- Suppose, that the SCM estimates a sizable effect for a certain intervention of interest.
    - This result would *not* be valid if the method *also* estimated large effects when applied to dates $t < T_0$ when the intervention did not occur (similar to the “pre-program test" in Heckman and Hotz (1989)
    - We refer to these falsification exercises as “**in-time placebos**”. We expect there are no estimated effects prior to the intervention, i.e. that the treatment appear *only around* the $T_0 + 1$
- These tests are feasible if there are available data for a sufficiently large T when no structural shocks to the outcome variable occurred

Robustness: model specification and leave-one-out

- In order to assess the robustness of our results,

1. we can include additional predictors of the outcome to construct the synthetic control. We expect the results to not change much regardless of which and how many predictor variables we included.
2. **leave-one-out**: testing the sensitivity of results to changes in js that provide $w > 0$ in W^*. E.g.
   - Say that the SC is estimated as a weighted average of $j = 2, 3, 4$ out of $J = 10$ (i.e. received a positive weight)
   - So, we can *iteratively* re-estimate the baseline model to construct a SC omitting *everytime* 1 of the js in $j = 2, 3, 4$.

   By excluding each of these j we sacrifice some goodness of fit but *still* this allows us to evaluate to what extent our results are driven by any particular control country.

Outline

1. Motivation

2. Formal aspects
   - Potential outcomes, econometric model and its bias
   - Comparison to regression and estimation of the SC
   - Inference, falsification and robustness
   - Limitations and recommendations for practitioners

Warnings

- The SCM facilitates comparative case studies when *no single untreated unit* provides a *good comparison* for the unit affected by the treatment or event of interest. This is often the case
  - When the treatment affects large aggregates like regions or countries
  - This results that a limited number of untreated units are available.
- Should be excluded from the DP those units
  1. affected by the intervention of interest or of similar nature
  2. that suffered large idiosyncratic shocks in Y if these shocks would have NOT affected $j = 1$ if the treatment did not happen
  3. with different characteristics $\boldsymbol{X}$ to the treated unit (i.e. $j = 1$ and the DP should behave similarly over extended periods of time *prior to the intervention*) so to avoid
     - 3.1 interpolation biases
     - 3.2 **overfitting**, i.e. when the characteristics of $j = 1$ are artificially matched by combining idiosyncratic variations in the sample of js in the DP

Specifying the variables in X

- Abadie et al. 2020 mentions that the credibility of a SC depends on its ability to track the pre-intervention trajectory of Y_1. So how do we choose what to include in in X_1 and X_0? We have some leeway
  - We can match the entire trajectory of Y_1 by including only the average of the Y_j in the DP in the pre-treatment period for
  - This is because the co-movement of Y across the js is *exactly what synthetic controls are designed to exploit*,
- The advantage from such *a* **summary** *of Y* in the pre-intervention period (*as opposed to including all annual values of Y as predictors*) to calculate the SC resides in a higher sparsity of the resulting SC
  - Sparse SC (that is, synthetic controls made of a small number of comparison units) are easy to interpret and evaluate
- Does this mean that we should only include pre-intervention outcomes and ignore the information of other predictors, Z_j? No!
  - In $Y_{jt}^N = \delta_t + \boldsymbol{\theta}_t Z_j + \boldsymbol{\lambda}_t \boldsymbol{\mu}_j + \varepsilon_{jt}$ covariates excluded from Z_j are mechanically absorbed into $\boldsymbol{\mu}_j$, which increases the number of components of $\boldsymbol{\mu}_j$ and, therefore, the **bound on the bias**

Requirements (1)

1. *Aggregate data on predictors and outcomes*
   - Specifically
     1.1 outcomes and predictors of the outcome for the unit or units exposed to the intervention of interest
     1.2 outcomes and predictors of the outcome for a set of comparison units
   - Sometimes, when aggregate data do not exist we can employ *aggregates* of micro data
2. *Sufficient T_0*: bias of the SC estimator is *bounded* by a function that is inversely proportional to T_0, during which the SC closely tracks the trajectory of the outcome variable for the affected unit
   - With a small T_0 close or even perfect fit of the predictor values for the treated unit may be spuriously attained, and so the resulting SC may fail to reproduce the trajectory of the outcome for $j = 1$ in the absence of the intervention.
   - The severity of this problem *can be diminished if X includes good predictors of post-intervention values of Y_{jt}* other than pre-intervention values of Y
   - However, a caveat is the possibility of **structural breaks**

Requirements (2)

1. **Sufficient post-intervention information**.
   - The evaluation data must include *outcome measures that*
     1.1 *are affected by the intervention*
     1.2 *are relevant for the policy decision that is the object of the study*.
   - This may be problematic if
     1.1 the effect of an intervention is expected to arise *gradually* over time
     1.2 no forward looking measures of the outcome are available
   - Extensive postintervention information allows a *more complete picture of the effects of the intervention*, in time and across the various outcomes of interest.

It is not recommended using this method when the pretreatment fit is poor or T_0 is small

What is cluster sampling? (Wooldridge 2010, Chap 20)

- Most of the quasi experiments can be thought (see Abadie et al. 2017) in terms of **Cluster sampling**: *clusters or groups, rather than individuals, randomly drawn from a large population of clusters*
- E.g. in evaluating the impact an immigration shock on individual wages in Peru, one might sample departments from the entire country (as opposed to randomly drawing individuals from the population of Peru).
  - these departments *constitute the clusters*
  - the workers *within* the departments are the individual units.
- The cluster sampling scheme generally implies that
  1. the *outcomes of units within a cluster are correlated through unobserved* "**cluster effects**."
  2. some covariates, such as GDP, will be *perfectly correlated* because workers in the same region have the same GDP. Other covariates, such as education level, are likely to have substantial correlation but will vary within a region.
- So
  1. We have many clusters that can be assumed to be independent of each other
  2. Observations within a cluster are correlated
  3. Cluster samples are naturally unbalanced even without sample selection

PTA (Lee 2016)

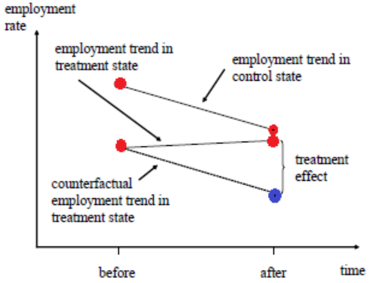

employment rate
employment trend in treatment state
employment trend in control state
treatment effect
counterfactual employment trend in treatment state
before
after
time

Figure 5.2.1: Causal effects in the differences-in-differences model

Source: A&P (2008)

- Remember, in the RCS case with 2 areas and 2 periods, where $S_i = 1$[individual i sampled at $t = 1$] is the treatment period and $Q_i = 1$[individual i is in the treatment group] is the treatment qualification
  - $Y^0_{t=0} = \beta_1 + \beta_q Q + W^{'}_2 \beta_w + u_2$
  - $Y^0_{t=1} = Y^0_{t=0} + \underline{\beta_\tau} = \beta_1 + \underline{\beta_\tau} + \beta_q Q + W^{'}_3 \beta_w + u_3$ (time effect β_τ added)
  - $Y^1_{t=1} = Y^0_{t=1} + \underline{\beta_d} = \beta_1 + \underline{\beta_\tau} + \underline{\beta_d} + \beta_q Q + W^{'}_3 \beta_w + u_3$ (treatment effect β_d added)

PTA (Lee 2016)

- Then

$$\boxed{Y_i = (1-S_i)\,Y_{i,t=0} + S_i\,Y_{i,t=1} = (1-S_i)\,\underbrace{Y^0_{i,t=0}}_{=Y_{i,t=0}} + S_i\underbrace{\left[(1-Q_i)\,Y^0_{i,t=1} + Q_i\,Y^1_{i,t=1}\right]}_{=Y_{i,t=1}}}$$

after plugging and operating

$$Y_i = \beta_1 + \beta_\tau S_i + \beta_q Q_i + \boxed{\beta_d}\,\underbrace{S_i \times Q_i}_{=D_i} + W_i\beta_w + u_i$$

where $D_i = \begin{cases} 1 & \text{if } S_i = 1 \text{ and } Q_i = 1 \\ 0 & \text{otherwise} \end{cases}$, $(1-S_i)\,W_2' + S_i W_3' \equiv W_i$ and $(1-S_i)\,U_2 + S_i U_3 \equiv U_i$.

- In the case with more areas and periods, e.g. if treatment occurs in $t=1$ and $t=2$, we can add a dummy for each of these and
  - 1 interaction for the treatment $S_i \times Q_i$ (as seen) only
  - 1 interactions of the treatment qualification with a particular year
    - 1[individual i sampled at $t=1$] $\times\, Q_i = 1$ [individual i is in ANY of the treatment groups]
    - 1[individual i sampled at $t=2$] $\times\, Q_i = 1$ [individual i is in ANY of the treatment groups]

▸ Back

Example of good covariate pre-treat. balance

- it is important to calculate the magnitude of $Y_{1t} - \sum_{j=2}^{J+1} w_j^* \boldsymbol{Z}_j = \boldsymbol{Z}_1$ between $j = 1$ and the SC to have an idea of the size of bias of the SC we could face

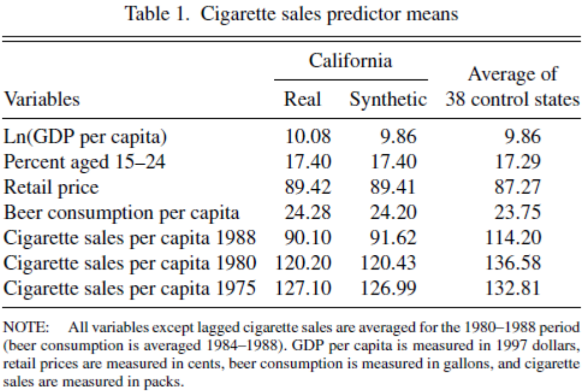


Table 1. Cigarette sales predictor means

	California		Average of
Variables	Real	Synthetic	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). GDP per capita is measured in 1997 dollars, retail prices are measured in cents, beer consumption is measured in gallons, and cigarette sales are measured in packs.

Source: Abadie et al. (2010)

Back

Example of good pre-treat. fit

- The ability of SC to reproduce the trajectory of the Y_1 over an extended period of time provides an indication of low bias
- An example of a good fit, with $T_0 = 1988$, is

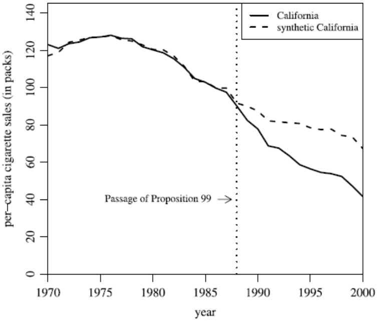


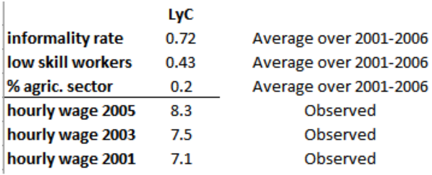
Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

Source: Abadie et al. (2010)

▸ Back

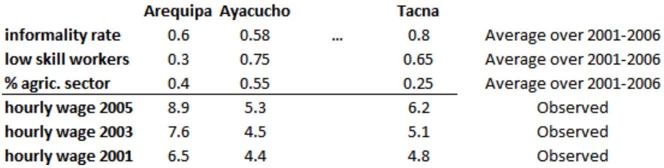
Example of X matrices

- For example, if $Y$is hourly wage and we have yearly data 2001-2010 abd the treatment happens from 2007($= T_0 + 1$) onwards in Lyc only
  - X_1 has $r = 3$ and $M = 3$ rows and 1 column



	LyC	
informality rate	0.72	Average over 2001-2006
low skill workers	0.43	Average over 2001-2006
% agric. sector	0.2	Average over 2001-2006
hourly wage 2005	8.3	Observed
hourly wage 2003	7.5	Observed
hourly wage 2001	7.1	Observed

  - X_0 has $r = 3$ and $M = 3$ rows and 23 columns



	Arequipa	Ayacucho		Tacna	
informality rate	0.6	0.58	...	0.8	Average over 2001-2006
low skill workers	0.3	0.75		0.65	Average over 2001-2006
% agric. sector	0.4	0.55		0.25	Average over 2001-2006
hourly wage 2005	8.9	5.3		6.2	Observed
hourly wage 2003	7.6	4.5		5.1	Observed
hourly wage 2001	6.5	4.4		4.8	Observed

▸ Back

Example of permutation test

- The solid line is the estimated TE for California and the gray line is the estimated TE for the other states in each placebo run
- If our TE is significative, we expect that after $T_0 + 1$ the black line is more extreme compared to the rest

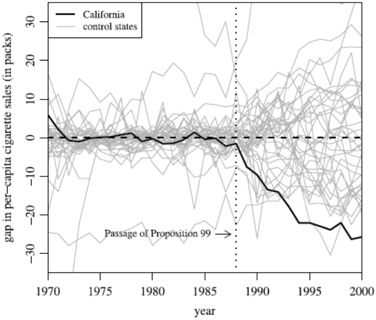
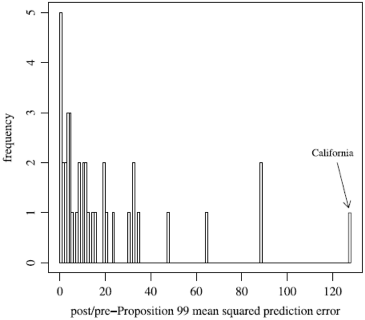


Figure 4. Per-capita cigarette sales gaps in California and placebo gaps in all 38 control states.

Source: Abadie et al. (2010)

Example of p-value in SCM

- Each bar is the ratio of MSPE for Califonia and for each placebo run
- If our TE is significative, we expect that the ratio for California very extreme



Source: Abadie et al. (2010)

▸ Back