Econometrics II

Lecture 11: Matched DID and Synthetic Controls

David Schönholzer

Stockholm University

May 7, 2024

What if Parallel Trends Assumption (PTA) Is Unlikely to Hold?

- Review of DID so far: if PTA holds, then can identify treatment effects:
  - Static and dynamic: lectures 9 and 10
  - Homogeneous or heterogeneous: first parts of lectures versus second parts
- But what if PTA unlikely to hold?
  - Briefly discussed some solutions, e.g. more fixed effects
  - Today: what to do if evidence against PTA or a priori unlikely
  - These methods are typically dynamic and robust to heterogeneity
  - Also allow us to estimate effects when only few units are treated
- Many practical applications!

Plan for Today

1. Matched Difference-in-Differences
   - Matching with Panel Data
   - Example
2. Synthetic Control Methods
   - Basic Idea
   - Examples
3. Appendix
   - Synthetic Controls with Many Treated Units
   - Matrix Completion Methods
     - Setup
     - Examples

Table of Contents

1 Matched Difference-in-Differences
Matching with Panel Data
Example

2 Synthetic Control Methods
Basic Idea
Examples

3 Appendix
Synthetic Controls with Many Treated Units
Matrix Completion Methods
Setup
Examples

Reminder: Counterfactuals and Within-Cell Comparisons

- **Covariate imbalance**: control group not a good counterfactual

$$\mathbb{E}\left[Y_i | D_i = 0\right] = \mathbb{E}\left[Y_i(0) | D_i = 0\right] \neq \mathbb{E}\left[Y_i(0) | D_i = 1\right]$$

  - Potential solution: control for $\mathbf{X}_i$ such that

$$\mathbb{E}\left[Y_i(0) | D_i = 0, \mathbf{X}_i = x\right] = \mathbb{E}\left[Y_i(0) | D_i = 1, \mathbf{X}_i = x\right]$$

  and compare treated and control only within cells defined by x

- **PTA violation**: control group *trend* not a good counterfactual

$$\mathbb{E}\left[Y_{i1} - Y_{i0} | D_i = 0\right] = \mathbb{E}\left[Y_{i1}(0) - Y_{i0}(0) | D_i = 0\right] \neq \mathbb{E}\left[Y_{i1}(0) - Y_{i0}(0) | D_i = 1\right]$$

  - Potential solution: control for $\mathbf{X}_i$ such that

$$\mathbb{E}\left[Y_{i1}(0) - Y_{i0}(0) | D_i = 0, \mathbf{X}_i = x\right] = \mathbb{E}\left[Y_{i1}(0) - Y_{i0}(0) | D_i = 1, \mathbf{X}_i = x\right]$$

  and compare treated and control *trends* only within cells defined by x

Conditional Parallel Trends

- Consider the 2×2 DID with $t \in \{0, 1\}$ with pre-treatment covariate vector $\mathbf{X}_i$
- We now make the **conditional parallel trends assumption** (CPTA):

$$\mathbb{E}\left[Y_{i1}(0) - Y_{i0}(0) | D_i = 0, \mathbf{X}_i\right] = \mathbb{E}\left[Y_{i1}(0) - Y_{i0}(0) | D_i = 1, \mathbf{X}_i\right] \quad \text{a.s.}$$

- We need **sufficient overlap**: $0 < \Pr\left(D_i = 1 | \mathbf{X}_i\right) < 1$ almost surely
- We can then identify the ATT at a given $\mathbf{X}_i = x$ as

$$\tau(x) = \underbrace{\mathbb{E}\left[Y_{i1} - Y_{i0} | D_i = 1, \mathbf{X}_i = x\right]}_{\Delta \text{ over time in } D_i=1, \mathbf{X}_i=x} - \underbrace{\mathbb{E}\left[Y_{i1} - Y_{i0} | D_i = 0, \mathbf{X}_i = x\right]}_{\Delta \text{ over time in } D_i=0, \mathbf{X}_i=x}$$

i.e. the DID within sub-population x identifies the ATT for x

- Can then identify unconditional τ by averaging:

$$\tau = \mathbb{E}\left[Y_{i1}(1) - Y_{i1}(0) | D_i = 1\right] = \mathbb{E}\left[\tau(\mathbf{X}_i) | D_i = 1\right]$$

OLS Approach to CPTA

- May consider running

$$Y_{it} = \alpha_i + \gamma_t + \beta D_i P_t + (\mathbf{X}_i P_t)' \delta + \varepsilon_{it}$$

- However, $\beta \neq \tau$ unless the ATT is homogeneous w.r.t. $\mathbf{X}_i$
- Intuition: This OLS setup implicitly models CEF of $Y_{i1} - Y_{i0}$ as follows:
  - CEF depends on $\mathbf{X}_i$ with a constant slope δ, regardless of D_i
  - If ATT varies e.g. by age then change in CEF may depend on D_i

- See Abadie (2005) for details on this issue
- Luckily, there are semi-/non-parametric solutions to this problem

Regression Adjustment to DID with CPTA

- Heckman et al (1997) exploit:

$$\begin{aligned}\tau &= \mathbb{E}\left[\tau(\mathbf{X}_i)|D_i = 1\right] \\ &= \mathbb{E}\left[Y_{i1} - Y_{i0}|D_i = 1\right] - \mathbb{E}\left[\mathbb{E}\left[Y_{i1} - Y_{i0}|D_i = 0, \mathbf{X}_i\right]|D_i = 1\right]\end{aligned}$$

  where the first line follows from the LIE
- Estimate CEF for controls by x and average using treated $\mathbf{X}_i$-distribution:

$$\widehat{\tau}_{\text{RA}} = \frac{1}{N_1}\sum_{i:D_i=1}\left\{(Y_{i1} - Y_{i0}) - \widehat{\mathbb{E}}\left[Y_{i1} - Y_{i0}|D_i = 0, \mathbf{X}_i\right]\right\}$$

- $\widehat{\mathbb{E}}\left[Y_{i1} - Y_{i0}|D_i = 0, \mathbf{X}_i\right]$ is estimated CEF evaluated for treated $\mathbf{X}_i$
- How do we construct estimates for it?
  - Many semi-/non-parametric solutions
  - Often used: matching

DID Regression Adjustment Through Matching

- Goal: find control units with similar (or identical) $\mathbf{X}_i$ as treated
- Example of matching estimation procedure:
  1. For each treated i, choose unit $m(i)$ such that $m(i) = \arg\min_{m:D_i=0} \|\mathbf{X}_i - \mathbf{X}_m\|$
  2. Estimate $\widehat{\mathbb{E}}\left[Y_{i1} - Y_{i0} | D_i = 0, \mathbf{X}_i\right]$ using $\widehat{\mathbb{E}}\left[Y_{m(i),1} - Y_{m(i),0}\right]$
  3. Assign G_i to $m(i)$ and estimate e.g. matched event study as

$$Y_{it} = \alpha_i + \gamma_t + \underbrace{\sum_{\ell \in \mathcal{L}} 1\left[t = g + \ell\right] \delta_\ell}_{\text{avg. trend in treated and matched controls}} + \underbrace{\sum_{\ell \in \mathcal{L}} D_i P_t^\ell \beta_\ell}_{\text{deviation in treated}} + \varepsilon_{it}$$

- Intuitively appealing, often with very compelling results, plus het-robust!
- However, inference has complications (often ignored), Abadie Imbens (2006)

Table of Contents

1. Matched Difference-in-Differences
   - Matching with Panel Data
   - Example
2. Synthetic Control Methods
   - Basic Idea
   - Examples
3. Appendix
   - Synthetic Controls with Many Treated Units
   - Matrix Completion Methods
     - Setup
     - Examples

Goldschmidt and Schmieder (2017)

- Interested in effect of outsourcing on daily wages
- Use matched administrative worker-firm data from Germany
- Outsourcing: workers with same job and site but switch to subsidiary firm
- Concern: control group should capture time and life cycle effects

- Matched DID approach:
  1. Let D_i be whether a worker is ever affected by an outsourcing event
  2. Donor pool of control workers: same industry/occupation but $D_i = 0$
  3. $\mathbf{X}_i$: tenure, establishment size, wages at $\ell = -2$ and -3
  4. Run probit of D_i on $\mathbf{X}_i$ to and compute fitted values $\widehat{e}(\mathbf{X}_i)$
  5. Match each i to $m(i) = \arg\min_{m:D_i=0} |\widehat{e}(\mathbf{X}_i) - \widehat{e}(\mathbf{X}_m)|$

Match Evaluation

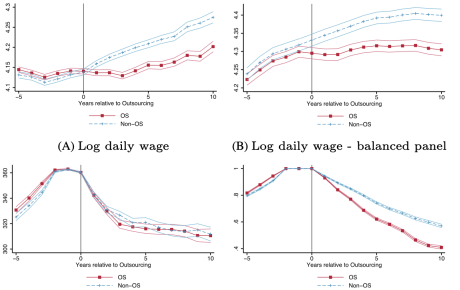
TABLE I

CHARACTERISTICS OF OUTSOURCED AND NONOUTSOURCED WORKERS

	Outsourced at $t = -1$	Matched non-OS at $t = -1$	FCSL at BSF/temp	FCSL not at BSF/temp
Avg establishment daily wage in euro	78.83	77.42	53.65	74.49
	(20.16)	(20.32)	(19.59)	(17.94)
Establishment effect*	0.03	0.03	−0.14	0.02
	(0.14)	(0.15)	(0.18)	(0.15)
Establishment size	1,120.63	1,107.55	265.41	1,683.45
	(2,416.86)	(3,207.42)	(385.18)	(5,204.99)
Real daily wage in euro	69.93	69.96	51.07	63.71
	(29.47)	(30.73)	(24.80)	(25.36)
Age in years	42.29	43.63	40.25	41.87
	(7.98)	(9.75)	(8.49)	(8.43)
Female	0.45	0.46	0.40	0.40
Years of education	10.16	10.23	9.93	10.06
	(1.17)	(1.34)	(1.06)	(0.89)
College degree	0.02	0.03	0.01	0.01
Living in West Germany	0.86	0.88	0.85	0.94
Working full-time	0.78	0.76	0.70	0.78
Tenure in years	8.58	8.51	3.91	6.16
	(5.80)	(6.32)	(3.83)	(5.29)
Food occupation	0.21	0.21	0.05	0.14
Cleaning occupation	0.11	0.11	0.41	0.24
Security occupation	0.03	0.03	0.11	0.08
Logistics occupation	0.34	0.34	0.42	0.53
Observations	21,195	21,195	6,412,854	35,201,181

Notes. Mean of each variable with standard deviation in parentheses. Columns (1) and (2) include on-site outsourced and matched nonoutsourced workers age 25–55 with at least 2 years of tenure in year before outsourcing. Statistics calculated in year before outsourcing. Columns (3) and (4) include workers in food, cleaning, security, and logistics occupations who are age 25–55 and employed at an establishment with 50 or more workers. Column (3) includes these workers who are employed at business service firms (BSF) or temp firms, and column (4) includes these workers who are not employed at BSF or temp firms. All columns exclude East Germany prior to 1997.

*The establishment effects are the predicted fixed effects from the AKM model described in Section IV.A. The establishment effects are normalized to be equal to 0 in the sample of all workers from 1979 to 2009 (the period we use for the AKM model).



(A) Log daily wage (B) Log daily wage - balanced panel

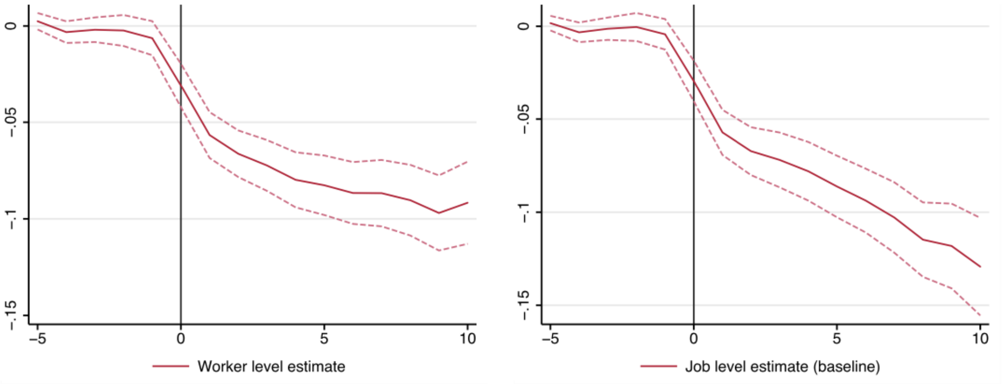
(C) Days worked per year (D) Probability of working at outsourced job

FIGURE IV

Employment Outcomes of Outsourced and Nonoutsourced Workers before and after On-site Outsourcing

The figures follow two group of workers: the first is a group of workers who are outsourced between year $t = -1$ and $t = 0$ (the first year at the new establishment), while the second group is a control group of nonoutsourced workers. The control group was chosen by finding workers employed in the same industry and occupation with similar tenure and establishment size in the year prior to outsourcing, who have similar wages two and three years prior to outsourcing as the outsourced workers. The figures show average characteristics of the workers in the two groups before and after the outsourcing event. Panels A, C, and D show data from the unbalanced panels of workers in the outsourced and control group. Panel B restricts the data to a balanced panel of individuals observed in each year from 5 years before to 10 years after the outsourcing event.

Goldschmidt and Schmieder (2017): Event Studies on Daily Wages



(A) All worker observations before and after outsourcing

(B) Sample restricted to observations remaining at the same job

Table of Contents

1 Matched Difference-in-Differences
Matching with Panel Data
Example

2 Synthetic Control Methods
Basic Idea
Examples

3 Appendix
Synthetic Controls with Many Treated Units
Matrix Completion Methods
Setup
Examples

Return to NJ and PA

- Recall that we said

$$\begin{aligned}\text{Treatment effect} &= (\bar{Y}_{\text{NJ},1} - \bar{Y}_{\text{PA},1}) - (\bar{Y}_{\text{NJ},0} - \bar{Y}_{\text{PA},0}) \\ &= (\bar{Y}_{\text{NJ},1} - \bar{Y}_{\text{NJ},0}) - (\bar{Y}_{\text{PA},1} - \bar{Y}_{\text{PA},0})\end{aligned}$$

- We showed that DID uses

$$\widehat{\tau}_{\text{DID}} = \underbrace{(\bar{Y}_{\text{NJ},1} - \bar{Y}_{\text{NJ},0})}_{\Delta \text{ over time in } D_i = 1} - \underbrace{(\bar{Y}_{\text{PA},1} - \bar{Y}_{\text{PA},0})}_{\Delta \text{ over time in } D_i = 0}$$

- Alternatively, could do

$$\widehat{\tau}_{\text{SCM}} = \underbrace{(\bar{Y}_{\text{NJ},1} - \bar{Y}_{\text{PA},1})}_{\Delta \text{ over group in } t = 1} - \underbrace{(\bar{Y}_{\text{NJ},0} - \bar{Y}_{\text{PA},0})}_{\Delta \text{ over group in } t = 0}$$

- This is approach of *synthetic control method* (SCM)

A Simple $2 \times T$ Case

- Consider some $2 \times T$ block structure with $g = T$, e.g.

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

- Imagine only one unit $i = N$ in treatment group
- We are interested in

$$\begin{aligned} \tau &= \mathbb{E}\left[Y_{iT}(1) - Y_{iT}(0) \mid D_i = 1\right] \\ &= Y_{NT}(1) - Y_{NT}(0) = Y_{NT} - \underbrace{Y_{NT}(0)}_{\text{model}} \end{aligned}$$

- Imagine cross-sectional OLS with all $i = 1, ..., N - 1$:

$$Y_{iT} = \phi_0 + \sum_{t=1}^{T-1} \phi_t Y_{it} + \varepsilon_{iT}$$

which Athey et al. (2018) call *horizontal regression*

Horizontal Regression

- Why is it called horizontal regression?

$$\begin{bmatrix} Y_{i1} & \cdots & Y_{iT-1} & \leftarrow Y_{iT} \\ Y_{N1} & \cdots & Y_{NT-1} & Y_{NT} \end{bmatrix}$$

  i.e. regressing post-control on pre-control

- We can then construct the fitted value

$$\widehat{Y}_{NT} = \widehat{\phi}_0 + \sum_{t=1}^{T-1} \widehat{\phi}_t Y_{Nt}$$

  as our estimate of $Y_{NT}(0)$, so that $\widehat{\tau} = Y_{NT} - \widehat{Y}_{NT}$

- This is very close what DID using PT does! It sets:

$$\widehat{\phi}_0 = \bar{Y}_{1,\text{pre}} - \bar{Y}_{0,\text{pre}}$$
$$\widehat{\phi}_t = \frac{1}{T-1}$$

Vertical Regression

- Consider instead a *vertical regression*:

$$\begin{bmatrix} Y_{i1} & \cdots & Y_{iT-1} & Y_{iT} \\ \uparrow & \uparrow & \uparrow & \\ Y_{N1} & \cdots & Y_{NT-1} & Y_{NT} \end{bmatrix}$$

  i.e. regressing pre-treatment on pre-control
- Construct the fitted value

$$\widehat{Y}_{NT} = \widehat{\lambda}_0 + \sum_{i=1}^{N-1} \widehat{\lambda}_i Y_{iT}$$

- This is almost what SCM does, with restrictions:
  1. $\widehat{\lambda}_0 = 0$
  2. $\widehat{\lambda}_i \geq 0$ for all $i \leq N-1$
  3. $\sum_{i=1}^{N-1} \widehat{\lambda}_i = 1$
- Unlike DID, PT does *not* have to hold unconditionally for SCM

Implementation

- Let

$$\begin{bmatrix} \mathbf{Y}_{\text{c,pre}} & \mathbf{Y}_{\text{c},T} \\ \mathbf{Y}'_{\text{t,pre}} & Y_{\text{t},T} \end{bmatrix}$$

  where

  - $\mathbf{Y}_{\text{c,pre}} = (\mathbf{Y}_{\text{c},1}, \dots, \mathbf{Y}_{\text{c},T-1})$ is $(N-1) \times (T-1)$ pre-control matrix
  - $\mathbf{Y}_{\text{t,pre}} = (Y_{N1}, \dots, Y_{NT-1})'$ is $(T-1) \times 1$ pre-treatment vector
  - $\mathbf{Y}_{\text{c},T} = (Y_{1T}, \dots, Y_{N-1,T})'$ is $(N-1) \times 1$ post-control vector

- SCM is typically implemented as CMD:

$$\min_{\boldsymbol{\lambda}=(\lambda_1,\dots,\lambda_{N-1})} \left(\mathbf{Y}_{\text{t,pre}} - \mathbf{Y}'_{\text{c,pre}}\boldsymbol{\lambda}\right)' \mathbf{W} \left(\mathbf{Y}_{\text{t,pre}} - \mathbf{Y}'_{\text{c,pre}}\boldsymbol{\lambda}\right)$$

  with **W** a symmetric weighting matrix

- Yields weights $\left(\widehat{\lambda}_1, \dots, \widehat{\lambda}_{N-1}\right)$ that form *synthetic control group*

Table of Contents

1 Matched Difference-in-Differences
- Matching with Panel Data
- Example

2 Synthetic Control Methods
- Basic Idea
- Examples

3 Appendix
- Synthetic Controls with Many Treated Units
- Matrix Completion Methods
  - Setup
  - Examples

Abadie & Gardeazabal (2003)

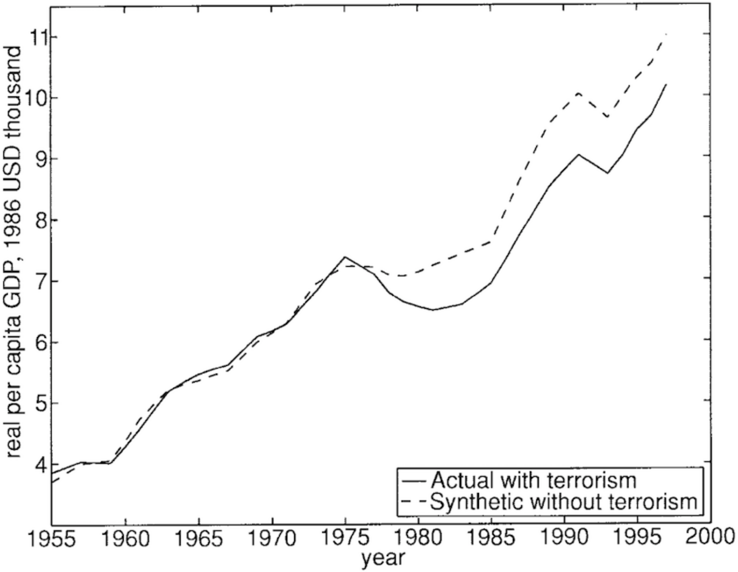
- Question: what is economic effect of terrorism?

- Authors study the Basque country (treatment unit)
- Want to use other Spanish regions as control units
  - But no parallel pre-trends in average of others
  - Construct weighted average using SCM — birth of SCM

- What are the weights?
  - $\widehat{\lambda}_i = 0$ for most regions
  - $\widehat{\lambda}_{\text{Catalonia}} = 0.85$ and $\widehat{\lambda}_{\text{Madrid}} = 0.15$

- So does this synthetic control group "work"?

Actual and Synthetic Control Trend



Andersson (2019): OECD Average (Left); Synthetic Sweden (Right)

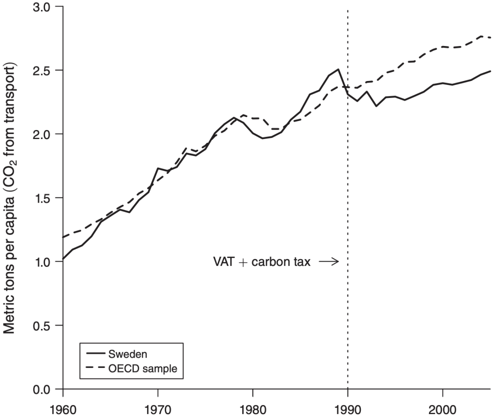


FIGURE 3. PATH PLOT OF PER CAPITA CO_2 EMISSIONS FROM TRANSPORT DURING 1960–2005: SWEDEN VERSUS THE OECD AVERAGE OF MY 14 DONOR COUNTRIES

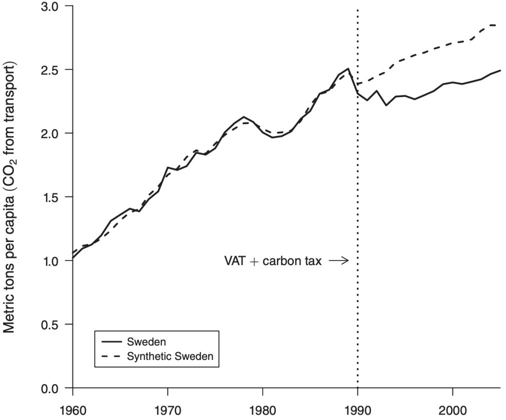


FIGURE 4. PATH PLOT OF PER CAPITA CO_2 EMISSIONS FROM TRANSPORT DURING 1960–2005: SWEDEN VERSUS SYNTHETIC SWEDEN

Comparison and Country Weights

TABLE 1—CO_2 EMISSIONS FROM TRANSPORT PREDICTOR MEANS BEFORE TAX REFORM

Variables	Sweden	Synth. Sweden	OECD sample
GDP per capita	20,121.5	20,121.2	21,277.8
Motor vehicles (per 1,000 people)	405.6	406.2	517.5
Gasoline consumption per capita	456.2	406.8	678.9
Urban population	83.1	83.1	74.1
CO_2 from transport per capita 1989	2.5	2.5	3.5
CO_2 from transport per capita 1980	2.0	2.0	3.2
CO_2 from transport per capita 1970	1.7	1.7	2.8

Notes: All variables except lagged CO_2 are averaged for the period 1980–1989. GDP per capita is purchasing power parity (PPP)—adjusted and measured in 2005 US dollars. Gasoline consumption is measured in kilograms of oil equivalent. Urban population is measured as percentage of total population. CO_2 emissions are measured in metric tons. The last column reports the population-weighted averages of the 14 OECD countries in the donor pool.

TABLE 2—COUNTRY WEIGHTS IN SYNTHETIC SWEDEN

Country	Weight	Country	Weight
Australia	0.001	Japan	0
Belgium	0.195	New Zealand	0.177
Canada	0	Poland	0.001
Denmark	0.384	Portugal	0
France	0	Spain	0
Greece	0.090	Switzerland	0.061
Iceland	0.001	United States	0.088

Note: With the synthetic control method, extrapolation is not allowed so all weights are between $0 \leq w_j \leq 1$ and $\sum w_j = 1$.

Treatment Effect Estimate

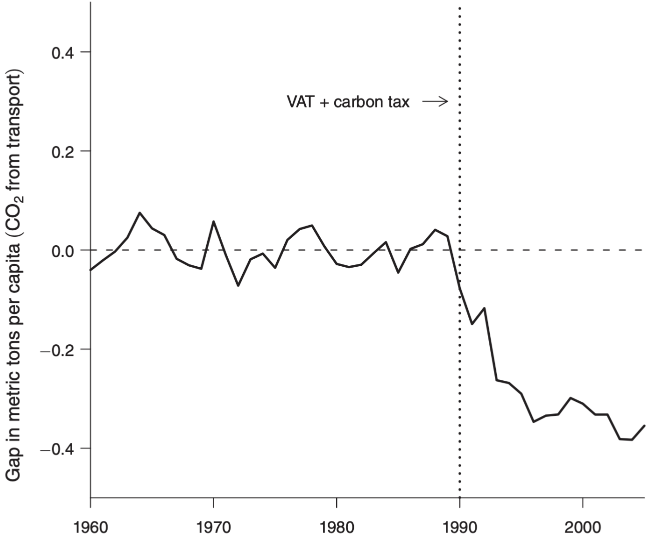


FIGURE 5. GAP IN PER CAPITA CO_2 EMISSIONS FROM TRANSPORT BETWEEN SWEDEN AND SYNTHETIC SWEDEN

Placebo in Time

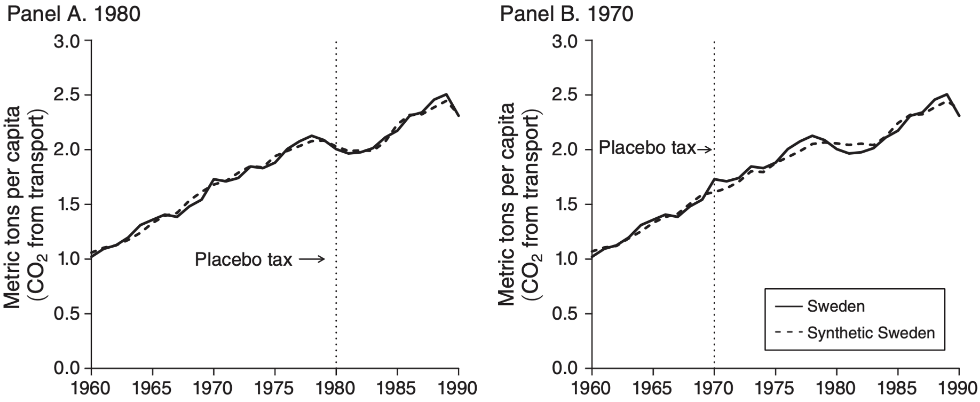


Figure 6. Placebo In-Time Tests

Notes: In panel A, the placebo tax is introduced in 1980, ten years prior to the actual policy changes. In panel B, the placebo tax is introduced in 1970.

Placebo Across Countries and Inference

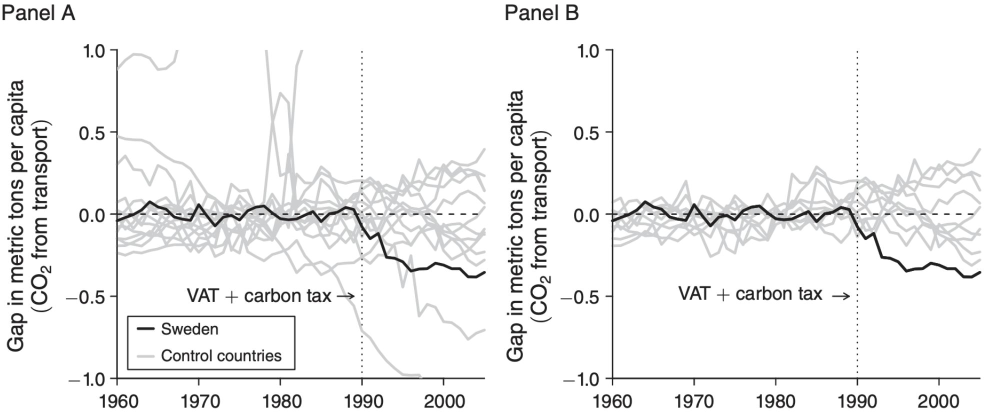


FIGURE 7. PERMUTATION TEST: PER CAPITA CO_2 EMISSIONS GAP IN SWEDEN AND PLACEBO GAPS FOR THE CONTROL COUNTRIES

Notes: Panel A shows per capita CO_2 emissions gap in Sweden and placebo gaps in all 14 OECD control countries. Panel B shows per capita gap in Sweden and placebo gaps in nine OECD control countries (countries with a pretreatment MSPE 20 times higher than Sweden's are excluded).

Table of Contents

1 Matched Difference-in-Differences
Matching with Panel Data
Example

2 Synthetic Control Methods
Basic Idea
Examples

3 Appendix
Synthetic Controls with Many Treated Units
Matrix Completion Methods
Setup
Examples

Multiple Treated Units

- Ben-Michael et al (2022) develop SCM for staggered rollout: multiple units
- Same setup as before, but N_0 units never-treated and $N_1 = N - N_0$ treated
- Treatment times given by G_i with $G_i = \infty$ for never-treated
- Continue to assume **no anticipation**: $Y_{it}(g) = Y_{it}(\infty)$ for $g > t$, implying:

$$Y_{it} = 1\left[t < G_i\right] Y_{it}(\infty) + 1\left[t \geq G_i\right] Y_{it}(G_i)$$

- **Target parameter**: unit-specific dynamic treatment effect for each $i : G_i < \infty$:

$$\tau_{i\ell} = Y_{ig+\ell}(G_i) - Y_{ig+\ell}(\infty)$$

  and by no anticipation we have $\tau_{i\ell} = 0$ for any $\ell < 0$
- Can use these to compute **dynamic ATT estimates**:

$$\tau_\ell = \frac{1}{N_1} \sum_{i:G_i<\infty} \tau_{i\ell}$$

SCM Estimator

- Define donor set $\mathcal{D}_{i\ell}$: all never treated or not-yet treated $j \neq i$
- The SCM weights for treated unit i are estimated via

$$\min_{\lambda_i \in \Lambda_i} \underbrace{\frac{1}{G_i - 1} \sum_{\ell < 0, i: G_i < \infty} \left(Y_{ig+\ell} - \sum_{j=1}^{N} \lambda_{ji} Y_{jg+\ell} \right)^2}_{\text{SCM objective}} + \underbrace{\rho \sum_{i=1}^{N} \lambda_{ji}^2}_{\text{Regularization}}$$

with Λ_i such that $\lambda_{ji} = 0$ when $j \notin \mathcal{D}_{i\ell}$

- Regularization tunes objective towards including all units
- Missing potential outcome is then estimated as

$$\widehat{Y}_{ig+\ell}(\infty) = \sum_{j=1}^{N} \lambda_{ji} Y_{jg+\ell}$$

- And treatment effect estimates are $\widehat{\tau}_{i\ell} = Y_{ig+\ell} - \widehat{Y}_{ig+\ell}(\infty)$

Mean Squared Placebo Treatment Effect

- Let ℓ_U be number of pre-treatment estimates
- Then $\widehat{\boldsymbol{\tau}}_i^{\text{pre}} = (\widehat{\tau}_{i,\ell_U}, ..., \widehat{\tau}_{i,-1}) \in \mathbb{R}^{\ell_U}$ is vector of placebo effects
- Hence the SCM objective is the mean squared placebo effect:

$$q_i(\widehat{\boldsymbol{\lambda}}_i)^2 \equiv \frac{1}{G_i - 1} \left\| \widehat{\boldsymbol{\tau}}_i^{\text{pre}} \right\|_2^2 = \frac{1}{G_i - 1} \sum_{\ell<0, i: G_i < \infty} \left(Y_{ig+\ell} - \sum_{j=1}^{N} \lambda_{ji} Y_{jg+\ell} \right)^2$$

- Want to pick weights $\widehat{\boldsymbol{\lambda}}_i$ to minimize this error

Combining Multiple Treated Units

- We now want $\boldsymbol{\lambda}_i$ for all i with $G_i < \infty$
- Collect these into $N \times N_1$ matrix $\boldsymbol{\Lambda} = [\boldsymbol{\lambda}_{N_0+1}, ..., \boldsymbol{\lambda}_N]$
- Can write $\widehat{\tau}_\ell$ in two equivalent ways:

$$\widehat{\tau}_\ell = \frac{1}{N_1} \sum_{i:G_i<\infty} \widehat{\tau}_{i\ell}$$

$$\text{(average of unit-specific SCM-estimates)} = \frac{1}{N_1} \sum_{i:G_i<\infty} \left[Y_{ig+\ell} - \sum_{j=1}^{N} \widehat{\lambda}_{ji} Y_{jg+\ell} \right]$$

$$\text{(SCM-estimate for average treated unit)} = \frac{1}{N_1} \sum_{i:G_i<\infty} Y_{ig+\ell} - \sum_{i:G_i<\infty} \sum_{j=1}^{N} \frac{\widehat{\lambda}_{ji}}{N_j} Y_{jg+\ell}$$

Two Goodness-of-Fit Measures

- Depending on interpretation of $\widehat{\tau}_\ell$, can evaluate pre-fit via:
  1. Root mean square of pre-treatment fits across treated units:

$$q^{\text{sep}}(\widehat{\boldsymbol{\Lambda}}) = \sqrt{\frac{1}{N_1}\sum_{i:G_i<\infty} q_i^2(\boldsymbol{\lambda}_i)} = \sqrt{\frac{1}{N_1}\sum_{i:G_i<\infty} \frac{1}{G_i-1}\left\|\widehat{\boldsymbol{\tau}}_i^{\text{pre}}\right\|_2^2}$$

  2. Root mean square of pre-treatment fit of average unit:

$$q^{\text{pool}}(\widehat{\boldsymbol{\Lambda}}) = \frac{1}{\sqrt{\ell_U}}\left\|\frac{1}{N_1}\sum_{i:G_i<\infty}\widehat{\boldsymbol{\tau}}_i^{\text{pre}}\right\|_2$$

- So two different ways to pick weights! Separately or pooled

Lower Bound of Error in Pre-Treatment Fit

- Focus on effect of first period of treatment τ_0
- Can show that lower bound of estimation error is weighted average:

Theorem (Estimation error as weighted average)

Under assumptions on the DGP it can be shown that the error is bounded by:

$$|\hat{\tau}_0 - \tau_0| \leq \rho^{pool} q^{pool}(\widehat{\boldsymbol{\Lambda}}) + \rho^{sep} q^{sep}(\widehat{\boldsymbol{\Lambda}}) + \zeta$$

where ρ^{pool} and ρ^{sep} depend on treatment structure and data, and ζ is noise

$\rightarrow$ Optimal SCM for staggered rollout is partially pooled!

Table of Contents

1 Matched Difference-in-Differences
Matching with Panel Data
Example

2 Synthetic Control Methods
Basic Idea
Examples

3 Appendix
Synthetic Controls with Many Treated Units
Matrix Completion Methods
Setup
Examples

Matrix Completion Methods

- Goal: estimate causal effect of $D_i \in \{0, 1\}$ in panel data
- We have seen two approaches:
  1. DID: estimate stable pattern over t across i
  2. Synthetic control: estimate stable pattern over i across t
- New approach: Matrix completion (Athey et al. 2018)
  - Find stable pattern over (i, t)
  - Data-driven mix of patterns across units and periods
  - Combines machine learning/computer science with metrics

Setup: Control Counterfactual $\mathbf{Y}$

- Consider $N \times T$ matrix $\mathbf{Y}$ of outcomes (wide format)
- Only observe Y_{it} for some i and t
  - Let $(i, t) \in \mathcal{M}$ be missing entries
  - $(i, t) \in \mathcal{O}$ are observed entries
- Potential outcomes: observe only $Y_{it}(0)$ or $Y_{it}(1)$
- Causal inference approach:
  - Impute $\mathbf{Y}(0) \equiv \mathbf{Y}$ matrix for i with $D_i = 1$
  - Use those to construct $\mathbb{E}\left[Y_{it}(1) - Y_{it}(0) | D_i = 1\right]$

Data Structure

1. Complete treatment matrix:

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & 1 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 1 \\ 0 & 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 0 & \cdots & 1 \end{bmatrix}$$

   where $D_{it} = 1$ if $(i, t) \in \mathcal{M}$ and zero otherwise

2. Incomplete counterfactual outcome matrix:

$$\mathbf{Y} = \begin{bmatrix} Y_{11} & Y_{12} & ? & \cdots & Y_{iT} \\ ? & ? & Y_{23} & \cdots & ? \\ Y_{31} & ? & Y_{33} & \cdots & ? \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{N1} & ? & Y_{N3} & \cdots & ? \end{bmatrix}$$

Patterns of Missing Data

- Block structure: subset treated (i.e. missing) for all $t \geq e$

$$\mathbf{Y} = \begin{bmatrix} \circ & \circ & \circ & \circ & \cdots & \circ \\ \circ & \circ & \circ & \circ & \cdots & \circ \\ \circ & \circ & \circ & \circ & \cdots & \circ \\ \circ & \circ & \bullet & \bullet & \cdots & \bullet \\ \circ & \circ & \bullet & \bullet & \cdots & \bullet \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \circ & \circ & \bullet & \bullet & \cdots & \bullet \end{bmatrix}$$

where $\bullet$ denotes $(i, t) \in \mathcal{M}$ and $\circ$ mean $(i, t) \in \mathcal{O}$

Special Block Structures

- Two special cases of block structures:

$$\mathbf{Y} = \begin{bmatrix} \circ & \circ & \cdots & \circ & \circ \\ \circ & \circ & \cdots & \circ & \circ \\ \circ & \circ & \cdots & \circ & \bullet \\ \circ & \circ & \cdots & \circ & \bullet \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \circ & \circ & \cdots & \circ & \bullet \end{bmatrix} \quad \text{and } \mathbf{Y} = \begin{bmatrix} \circ & \circ & \circ & \cdots & \circ & \circ \\ \circ & \circ & \circ & \cdots & \circ & \circ \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \circ & \circ & \circ & \cdots & \circ & \circ \\ \circ & \circ & \bullet & \cdots & \bullet & \bullet \end{bmatrix}$$

- Left: ATT estimable by DID
- Right: ATT estimable by synthetic control approach

Staggered Adoption

- Staggered adoption

$$\mathbf{Y} = \begin{bmatrix} \circ & \circ & \circ & \circ & \cdots & \circ \\ \circ & \circ & \circ & \circ & \cdots & \bullet \\ \circ & \circ & \circ & \bullet & \cdots & \bullet \\ \circ & \circ & \bullet & \bullet & \cdots & \bullet \\ \circ & \circ & \bullet & \bullet & \cdots & \bullet \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \circ & \bullet & \bullet & \bullet & \cdots & \bullet \end{bmatrix}$$

- ATT estimable by event study

Thin and Fat Matrices

- Thin and fat matrices:

$$\mathbf{Y} = \begin{bmatrix} \circ & \circ & \circ \\ \circ & \circ & \circ \\ \circ & \circ & \bullet \\ \circ & \circ & \bullet \\ \vdots & \vdots & \vdots \\ \circ & \circ & \bullet \end{bmatrix} \quad \text{and } \mathbf{Y} = \begin{bmatrix} \circ & \circ & \circ & \cdots & \circ & \circ \\ \circ & \circ & \circ & \cdots & \circ & \circ \\ \circ & \circ & \bullet & \cdots & \bullet & \bullet \end{bmatrix}$$

- Thin matrix: e.g. classical DID with $T = 3$
- Fat matrix: e.g. synthetic control for single treated unit

Nuclear Norm Matrix Completion Estimator

- Idea: data-driven mix of horizontal and vertical prediction
- Model counterfactual as $\mathbf{Y} = \mathbf{L}^* + \varepsilon$ with $\mathbb{E}\left[\varepsilon | \mathbf{L}^*\right] = 0$
- How about just minimizing sum of squared errors?

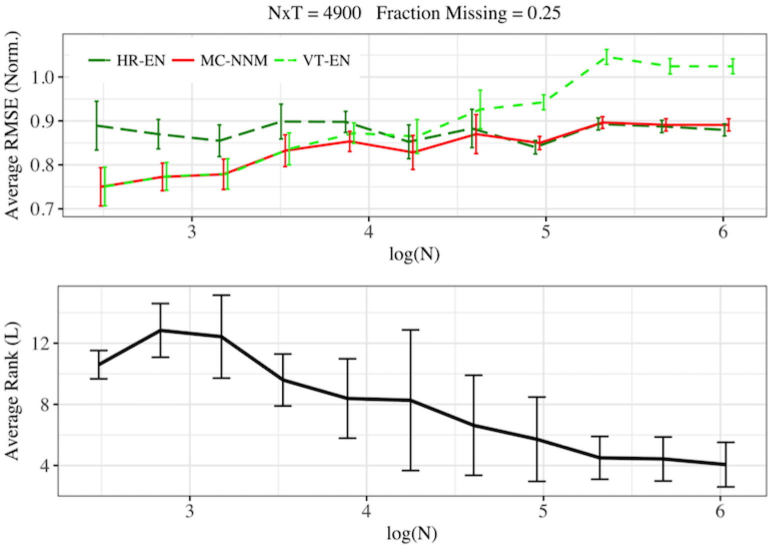
$$\min_{\mathbf{L}} \frac{1}{|\mathcal{O}|} \sum_{(i,t)\in\mathcal{O}} \left(Y_{it} - L_{it}\right)^2$$

- Does not work! Does not depend on L_{it} for $(i, t) \in \mathcal{M}$
- Instead:

$$\widehat{\mathbf{L}} = \arg\min_{\mathbf{L}} \left\{ \frac{1}{|\mathcal{O}|} \sum_{(i,t)\in\mathcal{O}} \left(Y_{it} - L_{it}\right)^2 + \lambda \left\|\mathbf{L}\right\|_* \right\}$$

  with penalty factor λ and nuclear matrix norm $\|\cdot\|_*$

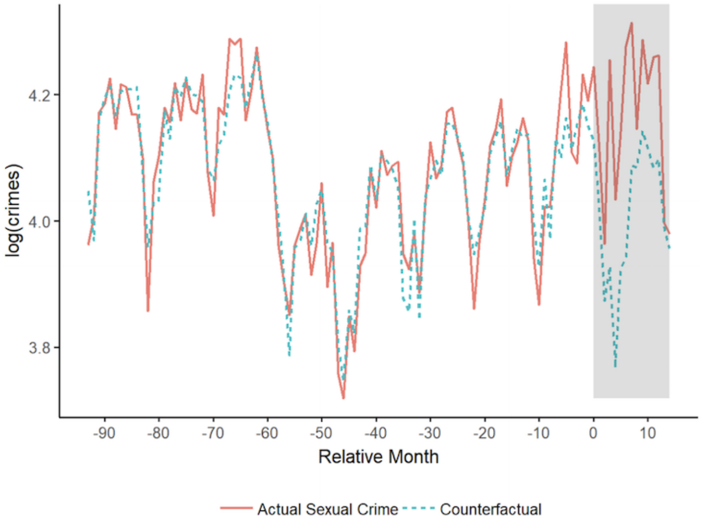
Athey et al. (2018): Impute Missing Stock Data



Levy and Mattsson (2020): Impute Counterfactual Sex Crimes

Figure 5: Matrix Completion Results

(a) Counterfactual Versus Actual Outcomes



Implied Treatment Effect of #MeToo