An Introduction to Modern Panel Data Methods: Differences-in-Differences and Synthetic Control

Diego Ciccia

Kellogg School of Management, Northwestern University

Introduction

- This set of lectures revolves around three causal inference methods for panel data
  - Differences-in-Differences (DID)
  - Synthetic Control (SC)
  - Synthetic Differences-in-Differences (SDID)
- The main focus is on SDID, hence DID and SC only serve as a stepping stone
- These methods work within similar frameworks
  - I will try to strike the *similarities-in-similarities*
  - I will discuss DID in a way that is conducive to SC and SDID
- Objective: give you a user-oriented overview of DID/SC/SDID
  - Broad enough to understand what the packages are doing
  - Not *too* simplified, to give you solid grounds to read the papers/books if needed

Table of Contents

Sources

- Clément de Chaisemartin and Xavier D'Haultfoeuille, *Causal Inference with Differences-in-Differences: Credible Answers to Hard Questions*, Online version
- Kirill Borusyak, *Applied Econometrics Notes, UC Berkeley*, Github
- Ivan Canay, *480-3 Econometrics, Northwestern University*, Slides
- (New book!) Damian Clarke. 2026. *Applied Microeconometrics*. MIT Press

Static potential outcome framework, I

- In impact evaluation, intervention units are assigned to some treatment D
  - Upskilling-reskilling training program, local/regional policies, ...
- Later on, data on some outcome Y is collected
- Treatment D arguably has some effect on Y, which is unknown to the researchers
- A way to formalize treatment effects is via the **potential outcome framework**
  - Neyman proposed this framework in his master's thesis in 1923
- The potential outcome for a unit assigned to treatment d is $Y(d)$
  - If $D \in \{0, 1\}$, $Y(1)$ outcome had the unit been treated, $Y(0)$ untreated
- The (individual) treatment effect (TE) is

$$TE \equiv Y(1) - Y(0)$$

Static potential outcome framework, II

- People generally *respond* differently to treatment
  - Some people may learn a lot from a job training program, other way less
  - As such, TE is often formalized as a random variable
  - We are interested is some characteristics of its distribution, e.g., mean, variance (!)
- Since policies are generally thought to be scaled-up to the whole population, we are often interested in *average* treatment effect

$$ATE \equiv \mathbb{E}[Y(1) - Y(0)]$$

- Unless we have experimental data, most of the time we can at most identify the average treatment effect among people *who were treated*

$$ATT \equiv \mathbb{E}[Y(1) - Y(0)|D = 1]$$

Static potential outcome framework, III

- **Fundamental problem of causal inference**: either $Y(1)$ or $Y(0)$ is observed
- Our data generally consists of (Y, D), where

$$\begin{aligned} Y = Y(D) &= Y(1)D + Y(0)(1 - D) \\ &= Y(0) + D \cdot TE \end{aligned}$$

- Data (Y, D) does not reveal $TE = Y(1) - Y(0)$
  - $Y = Y(1)$ if $D = 1$, we observe treated potential outcomes among treated units
  - $Y = Y(0)$ if $D = 0$, we observe untreated potential outcomes among untreated units
  - No one has $D = 1$ and $D = 0$ simultaneously
  - $Y(0)$ if $D = 1$ and $Y(1)$ if $D = 0$ are unknown
- **Goal**: under which assumptions, characteristics of TE are **identified** by the data?
  - Identified: expressed as a function of characteristics of **observable** data $(\mathbf{Y}, \mathbf{D})$

The meaning of ATT

- In DID/SC/SDID, the target parameter is ATT
  - ATT conditions on the subpopulation that was treated
  - Outside of randomized trials, this subpopulation may self-select into treatment
- Another useful interpretation of ATT:

$$\mathbb{E}[Y - Y(0)] = ATT \cdot P(D = 1)$$

  - Keep realized outcomes as they are $\rightarrow$ status-quo policy
  - Consider a policy where no one is treated $\rightarrow$ counterfactual policy
  - The *sign* of ATT tells us whether *on average* units are better off keeping their current treatment status vs being all untreated

Panel data

- So far, I have described the potential outcome framework for some undefined population, abstracting from a particular characterization of the data
- DID/SC/SDID require access to **panel** data:
  - Units $\{1, ..., G\}$ are each observed in periods $\{1, ..., T\}$
- As a result, the effective data at our disposal is

$$\{(Y_{g,t}, D_{g,t})\}_{(g,t)\in\{1,...,G\}\times\{1,...,T\}}$$

- Double-indexing allows us to rewrite the collection of outcomes and treatments in a convenient *matrix* form

Treatment design, I

- Each unit g can either be treated ($D_{g,t} = 1$) or untreated ($D_{g,t} = 0$) at time t
- Across all t, unit g has treatment vector $\mathbf{D}_g = (D_{g,1}, ..., D_{g,T})$
- The **design** of a study is the collection of treatment vectors

$$\mathbf{D} = (\mathbf{D}_1, ..., \mathbf{D}_G) = \begin{pmatrix} (g,t) & 1 & 2 & ... & T \\ 1 & D_{1,1} & D_{1,2} & ... & D_{1,T} \\ 2 & D_{2,1} & D_{2,2} & ... & D_{2,T} \\ \vdots & \vdots & \ddots & ... & \vdots \\ G & D_{G,1} & D_{G,2} & ... & D_{G,T} \end{pmatrix}$$

- $\mathbf{D}$ is $G \times T$ matrix whose (g,t) entry is treatment status $D_{g,t}$

Treatment design, II

- Basic design:
  - Treated units get treated at the **same** time and never go back to untreated
  - Ex. units 1 and 2 get treated at 3, while unit 3 is a **never-treated** unit

$$\mathbf{D} = \begin{pmatrix} (g,t) & 1 & 2 & 3 & 4 \\ 1 & 0 & 0 & 1 & 1 \\ 2 & 0 & 0 & 1 & 1 \\ 3 & 0 & 0 & 0 & 0 \end{pmatrix}$$

- Staggered design:
  - Treated units get treated at a **different** time and never go back to untreated
  - Ex. units 1 and 2 are in different **cohorts**

$$\mathbf{D} = \begin{pmatrix} (g,t) & 1 & 2 & 3 & 4 \\ 1 & 0 & 0 & 1 & 1 \\ 2 & 0 & 0 & 0 & 1 \\ 3 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Treatment design, III

- For now, I only consider basic designs
  - I will cover staggered designs in the context of SDID
  - One can also have more general settings (e.g., on-and-off)
- Basic designs clearly *partition* units and periods in two subgroups each:
  - Units in $\{1, ..., G\}$ either belong to G_1 (treated group) or G_0 (untreated)
  - Periods in $\{1, ..., T\}$ either belong to T_1 (post) or T_0 (pre)
- I use $|S|$ to refer to how many units are in set S
  - This is generally referred to as the *cardinality* of set S
  - Why? It makes the notation for the estimators nicer, also avoids tedious algebra
- With this notation, we have
  - $|G_1|$ units that are treated for $|T_1|$ periods
  - Treatment starts at period $|T_0| + 1$
  - $|G_0|$ units that are never-treated

Model-based perspective on inference

- A non-trivial point in this type of studies is: **what is random in the data?**
  - Units sampled from a larger population? Treatment assignment?
  - The answer to this question affects the way we draw inference on estimators
  - Section 2.4 of *Credible Answers to Hard Questions* really elucidates on this point
- The framework that is generally used in DID/SC/SDID is **model-based**
  - Treatment design $\mathbf{D}$ is fixed (non-stochastic)
  - Randomness comes from shocks that affect units' outcomes *on top of their treatment*
- This framework is generally enforced by **conditioning on the design**
  - Every expectation $\mathbb{E}[.]$ is implicitly $\mathbb{E}[.|\mathbf{D}]$
  - Since (G_0, G_1, T_0, T_1) are functions of $\mathbf{D}$, they can be *taken out* of the expectations
  - What remains random conditioning on $\mathbf{D}$ is *potential outcomes*

Potential outcomes in panel data, I

- At most, potential outcome of unit g at time t depends on g, t and the *design*

$$Y_{g,t} = Y_{g,t}(\mathbf{D})$$

- Dependence of potential outcomes on (g, t) mirrors the model-based framework
  - (g, t)-level shocks affect potential outcomes *even* conditioning on treatment
- We make a few assumptions:
  - **No interference**: Potential outcomes of unit g only depend on unit g's treatment

$$Y_{g,t} = Y_{g,t}(\mathbf{D}_g) = Y_{g,t}(D_{g,1}, ..., D_{g,T})$$

  - **No anticipation**: Potential outcomes at time t do not depend on future treatment

$$Y_{g,t} = Y_{g,t}(D_{g,1}, ..., D_{g,t})$$

Potential outcomes in panel data, II

- In SC/SDID, the following assumption is also maintained
  - **No dynamics**: Potential outcomes at time t do not depend on past treatment

$$Y_{g,t} = Y_{g,t}(D_{g,t})$$

- No dynamics rules out compounded effects from persistence of treatment status
  - For instance, it asssumes that just-laid-off workers are on the same treatment trajectory as workers who have been out of the workforce for a long time
- This assumption is often relaxed in some DID applications (e.g., event-study)
  - I keep it to understand the differences in DID/SC/SDID on an equal footing

Potential outcomes in panel data, III

- Under no dynamics, one can collect all observed outcomes in matrix form
- Say we have design

$$\mathbf{D} = \begin{pmatrix} (g,t) & 1 & 2 & 3 \\ 1 & 0 & 1 & 1 \\ 2 & 0 & 1 & 1 \\ 3 & 0 & 0 & 0 \end{pmatrix}$$

- The observed outcomes are

$$\mathbf{Y} = \begin{pmatrix} Y_{1,1}(0) & Y_{1,2}(1) & Y_{1,3}(1) \\ Y_{2,1}(0) & Y_{2,2}(1) & Y_{2,3}(1) \\ Y_{3,1}(0) & Y_{3,2}(0) & Y_{3,3}(0) \end{pmatrix}$$

- We do not observe $Y_{g,t}(0)$ for entries in red

Potential outcomes in panel data, IV

- By just adding and subtracting, we get

$$\mathbf{Y} = \begin{pmatrix} Y_{1,1}(0) & Y_{1,2}(1) & Y_{1,3}(1) \\ Y_{2,1}(0) & Y_{2,2}(1) & Y_{2,3}(1) \\ Y_{3,1}(0) & Y_{3,2}(0) & Y_{3,3}(0) \end{pmatrix} = \mathbb{E}[\mathbf{Y}(\mathbf{0})] + \boldsymbol{\Theta} \cdot \mathbf{D} + \mathbf{U}$$

where $\mathbb{E}[\mathbf{U}|\mathbf{D}] = 0$ and, letting $\theta_{g,t} = \mathbb{E}[Y_{g,t}(1) - Y_{g,t}(0)]$,

$$\mathbb{E}[\mathbf{Y}(\mathbf{0})] = \begin{pmatrix} \mathbb{E}[Y_{1,1}(0)] & \mathbb{E}[Y_{1,2}(0)] & \mathbb{E}[Y_{1,3}(0)] \\ \mathbb{E}[Y_{2,1}(0)] & \mathbb{E}[Y_{2,2}(0)] & \mathbb{E}[Y_{2,3}(0)] \\ \mathbb{E}[Y_{3,1}(0)] & \mathbb{E}[Y_{3,2}(0)] & \mathbb{E}[Y_{3,3}(0)] \end{pmatrix} \qquad \boldsymbol{\Theta} = \begin{pmatrix} \theta_{1,1} & \theta_{1,2} & \theta_{1,3} \\ \theta_{2,1} & \theta_{2,2} & \theta_{2,3} \\ \theta_{3,1} & \theta_{3,2} & \theta_{3,3} \end{pmatrix}$$

Potential outcomes in panel data, V

$$\mathbf{Y} = \mathbb{E}[\mathbf{Y}(\mathbf{0})] + \boldsymbol{\Theta} \cdot \mathbf{D} + \mathbf{U} \quad \overset{(g,t) \text{ level}}{\rightarrow} \quad Y_{g,t} = E[Y_{g,t}(0)] + \theta_{g,t} D_{g,t} + U_{g,t}$$

- Linear representation of $\mathbf{Y}$
  - $\mathbb{E}[Y(0)]$: systematic component
  - Θ: treatment effects
  - $\mathbf{U}$: idiosyncratic component
- This **latent factor** model is the key starting point of DID/SC/SDID
  - It looks like the usual linear model, so why don't we just run OLS?
- **Fundamental problem of causal inference**
  - $|G_1| \times |T_1|$ entries of $\mathbb{E}[\mathbf{Y}(\mathbf{0})]$ are unknown (more formally, *not identified*)
  - Even if we are only interested in Θ, we cannot "control" for $\mathbb{E}[\mathbf{Y}(\mathbf{0})]$
  - We cannot retrieve Θ, e.g., by linear regression, under the stated assumptions

What do DID, SC and SDID have in common?

- Let now recast $\mathbf{\Theta}$ to a known object

$$ATT = \frac{1}{|G_1|}\frac{1}{|T_1|}\sum_{g\in G_1}\sum_{t\in T_1}\mathbb{E}[Y_{g,t}(1) - Y_{g,t}(0)]$$
$$= \frac{1}{|G_1|}\frac{1}{|T_1|}\sum_{g\in G_1}\sum_{t\in T_1}\mathbf{\Theta}_{g,t}$$

  that is, ATT is just the average of the entries of Θ at which $\mathbf{D}_{g,t} = 1$

- Scope of DID/SC/SDID:

Under what assumptions on $\mathbb{E}[\mathbf{Y}(\mathbf{0})]$ can we recover ATT from data $(\mathbf{Y}, \mathbf{D})$?

Parallel trends, I

- DID answers the last question by imposing **parallel trends**

$$\mathbb{E}[Y_{g,t}(0) - Y_{g,t-1}(0)] = \mathbb{E}[Y_{g',t}(0) - Y_{g',t-1}(0)] \qquad \text{(Parallel Trends)}$$

  for all $g, g' \in \{1, ..., G\}$ and $t \geq 2$

- Units' untreated outcomes are required to change on average by the same *amount*
  - This assumption has nothing to do with *levels*, i.e., value of $\mathbb{E}[Y(0)_{g,t}]$
- Parallel trends has also an intuitive appeal to it
  - Without the treatment, units would have *evolved* in parallel
- Parallel trends solves the Fundamental problem in a direct way
  - Say unit 2 is treated in period 2, then the unknown counterfactual is identified

$$\mathbb{E}[Y_{2,2}(0)] = \mathbb{E}[Y_{2,1}(0)] + \mathbb{E}[Y_{1,2}(0) - Y_{1,1}(0)] = \mathbb{E}[Y_{2,1}] + \mathbb{E}[Y_{1,2} - Y_{1,1}]$$

Parallel trends, II

- Parallel trends is equivalent to assuming a specific functional form for $\mathbb{E}[Y_{g,t}(0)]$

$$\mathbb{E}[Y_{g,t}(0)] = \delta_g + \gamma_t$$

  i.e., average untreated outcomes can be separated in a time and unit component

- This means that the latent factor model reduces to

$$Y_{g,t} = \delta_g + \gamma_t + \theta_{g,t} D_{g,t} + U_{g,t}$$

- This specification splits the unknown $\mathbb{E}[Y_{g,t}(0)]$ into **time and unit fixed effects**

Two-Way Fixed Effects (TWFE) regression

- Suppose we run the following least-squares regression

$$(\beta, ., .) = \underset{(b,\delta,\gamma)}{\arg\min} \sum_{g=1}^{G} \sum_{t=1}^{T} \mathbb{E}\left[(Y_{g,t} - \delta_g - \gamma_t - b \cdot D_{g,t})^2\right]$$

- This specification is known as the **Two-Way Fixed Effects** (TWFE) regression
- It turns out that, under all the assumptions stated so far,

$$\beta = \frac{1}{|G_1||T_1|} \sum_{g \in G_1} \sum_{t \in T_1} \mathbb{E}[Y_{g,t}(1) - Y_{g,t}(0)] = ATT$$

- As a result, one can consistently estimate ATT in this setting by running

```
reghdfe Y D, absorb(G T)
```

Where are the differences in differences that we all love?

- The sample counterpart of TWFE is

$$(\hat{\beta}, ., .) = \underset{(b,\delta,\gamma)}{\arg\min} \sum_{g=1}^{G} \sum_{t=1}^{T} (Y_{g,t} - \delta_g - \gamma_t - b \cdot D_{g,t})^2$$

- It turns out that the TWFE estimator is exactly equal to the **diff-in-diff** estimator
  - The DID estimator compares average outcome differentials of treated and untreated units, before and after the start of the treatment

$$\hat{\beta} = \left(\frac{1}{|G_1||T_1|} \sum_{\substack{g \in G_1 \\ t \in T_1}} Y_{g,t} - \frac{1}{|G_1||T_0|} \sum_{\substack{g \in G_1 \\ t \in T_0}} Y_{g,t} \right) - \left(\frac{1}{|G_0||T_1|} \sum_{\substack{g \in G_0 \\ t \in T_1}} Y_{g,t} - \frac{1}{|G_0||T_0|} \sum_{\substack{g \in G_0 \\ t \in T_0}} Y_{g,t} \right)$$

Caveats about TWFE

- In the basic design, TWFE is equal to the DID estimator
  - Moreover, TWFE is unbiased for ATT, i.e., $\mathbb{E}[\hat{\beta}] = ATT$
- This equivalence only holds in the basic design
  - With multiple cohorts, the DID estimator is not clearly defined
  - Also, $\mathbb{E}[\hat{\beta}]$ is not guarenteed to be equal to ATT
- Beyond the basic design, $\mathbb{E}[\hat{\beta}]$ is a weighted sum of $\theta_{g,t}$s
  - Weights may be **negative**
  - As such, one could get that $\mathbb{E}[\hat{\beta}] > 0$ even if $\theta_{g,t} < 0$
  - Policy-wise, we could pass a policy that makes everyone worse off
  - More on these issues in Chapter 5 of *Credible Answers to Hard Questions*

A factor model approach, I

- Parallel trends assumption is often not verified
  - Treated and untreated groups are generally on different outcome trajectories
  - People join a training program *because* they experience a decline in their income
- The additive structure implied by parallel trends is not necessary to get ATT
- The baseline motivation for a **synthetic control** approach is more flexible form for the systematic component of the latent factor model

$$\mathbb{E}[Y_{g,t}(0)] = \Delta_g' \Gamma_t = \sum_{j=1}^{k} \Delta_{g,j} \Gamma_{t,j}$$

where Δ_g (**unit factor**) and Γ_t (**time factor**) are vectors in $\mathbb{R}^k$
  - E.g., $\Delta_g = (\delta_{g,1}, ..., \delta_{g,k})$ and $\Gamma_t = (\gamma_{t,1}, ..., \gamma_{t,k})$

A factor model approach, II

- This specification actually embeds parallel trends
  - Say $\Delta_g = (\delta_g, 1)$ and $\Gamma_t = (1, \gamma_t)$, then $\mathbb{E}[Y_{g,t}(0)] = \delta_g + \gamma_t$
- It also allows for non-parallel trends

$$\mathbb{E}[Y_{g,t}(0) - Y_{g,t-1}(0)] - \mathbb{E}[Y_{\tilde{g},t}(0) - Y_{\tilde{g},t-1}(0)] = (\Delta_g - \Delta_{\tilde{g}})'(\Gamma_t - \Gamma_{t-1})$$

  where the right-hand side need not to be 0
- SC was originally proposed for settings where only 1 or few units were treated
  - To enforce this, let $g_1 = |G_0| + 1$ and $t_1 = |T_0| + 1$

A factor model approach, III

- The ATT reduces to
$$ATT = \mathbb{E}[Y_{g_1,t_1}(1) - Y_{g_1,t_1}(0)]$$
- The Fundamental problem strikes back: $\mathbb{E}[Y_{g_1,t_1}(0)]$ is not observed
- Say, however, that there are weights ω in $[0, 1]$ such that
$$\sum_{g \in G_0} \omega_g \Delta_g = \Delta_{g_1} \qquad \sum_{g \in G_0} \omega_g = 1$$
- The unit factor of the treated unit lies in the **convex hull** of control units' factors
- A **synthetic** control unit can be formed by convex combination of untreated units
$$Y_t^* = \sum_{g \in G_0} \omega_g Y_{g,t}$$

A factor model approach, IV

$$(1)\ \ \mathbb{E}[Y_{g,t}(0)] = \Delta_g' \Gamma_t \qquad (2)\ \ \sum_{g \in G_0} \omega_g \Delta_g = \Delta_{g_1} \qquad (3)\ \ Y_t^* = \sum_{g \in G_0} \omega_g Y_{g,t}$$

- Consider the cross-sectional difference between g_1 and the SC in the post period

$$\begin{aligned}
\mathbb{E}[Y_{g_1,t_1}] - \mathbb{E}[Y_t^*] &= ATT + \mathbb{E}[Y_{g_1,t_1}(0)] - \sum_{g \in G_0} \omega_g \mathbb{E}[Y_{g,t_1}(0)] && (3)\\
&= ATT + \Delta_{g_1}' \Gamma_{t_1} - \sum_{g \in G_0} \omega_g \Delta_g' \Gamma_{t_1} && (1)\\
&= ATT + \left(\Delta_{g_1} - \sum_{g \in G_0} \omega_g \Delta_g \right)' \Gamma_{t_1}\\
&= ATT && (2)
\end{aligned}$$

- SC identifies the ATT under the factor model (1) and convex hull (2) assumptions

Implementation of SC, I

- The simplest way to compute the weights is to project outcomes in the pre period

$$\hat{\omega} = \underset{\omega \in [0,1]^{G_0}}{\arg\min} \sum_{t \in T_0} \left(Y_{g_1,t} - \sum_{g \in G_0} \omega_g Y_{g,t} \right)^2 \qquad \text{s.t.} \quad \sum_{g \in G_0} \omega_g = 1$$

- The SC estimator is

$$\hat{\tau}^{SC} = Y_{g_1,t_1} - \sum_{g \in G_0} \hat{\omega}_g Y_{g,t_1}$$

- In this way, the validity of SC hinges on the condition that

$$\sum_{g \in G_0} \hat{\omega}_g Y_{g,t} \approx Y_{g_1,t} \text{ for } t \in T_0 \quad \text{then} \quad \sum_{g \in G_0} \hat{\omega}_g \Delta_g \approx \Delta_{g_1}$$

that is, estimated weights well approximate those that would give us the ATT

Implementation of SC, II

- In a more general setting, suppose that G_1 and T_1 are small, with $|G_1|, |T_1| > 1$
- Define the weights by balancing the average treated units' outcomes

$$\hat{\omega} = \underset{\omega \in [0,1]^{G_0}}{\arg\min} \sum_{t \in T_0} \left(\frac{1}{|G_1|} \sum_{g \in G_1} Y_{g,t} - \sum_{g \in G_0} \omega_g Y_{g,t} \right)^2 \qquad \text{s.t.} \quad \sum_{g \in G_0} \omega_g = 1$$

- The SC estimator is a cross-sectional difference, so we can define for $t \in T_1$

$$\hat{\tau}_t^{SC} = \frac{1}{|G_1|} \sum_{g \in G_1} Y_{g,t} - \sum_{g \in G_0} \hat{\omega}_g Y_{g,t}$$

- Under regularity conditions, this is an (asymptotically) unbiased estimator for

$$ATT_t = \frac{1}{|G_1|} \sum_{g \in G_1} \mathbb{E}[Y_{g,t}(1) - Y_{g,t}(0)]$$

Implementation of SC, III - Aggregated estimator

- Consider the following weighted fixed regression

$$(\hat{\tau}^{SC}, .) = \arg\min_{b,\gamma} \sum_{g=1}^{G} \sum_{t=1}^{T} (Y_{g,t} - \gamma_t - b \cdot D_{g,t})^2 \hat{\omega}_g^{SC}$$

where

$$\hat{\omega}_g^{SC} = \begin{cases} \hat{\omega}_g & \text{if } g \in G_0 \\ \frac{1}{|G_1|} & \text{if } g \in G_1 \end{cases}$$

- The weighted fixed effects estimator is equal to the average of all SC estimators

$$\hat{\tau}^{SC} = \frac{1}{|T_1|} \sum_{t \in T_1} \left(\frac{1}{|G_1|} \sum_{g \in G_1} Y_{g,t} - \sum_{g \in G_0} \hat{\omega}_g Y_{g,t} \right) = \frac{1}{|T_1|} \sum_{t \in T_1} \hat{\tau}_t^{SC}$$

Abadie et al. (2010)

- Context: California's "Proposition 99" tobacco control policy (1989)
- California increased sales tax for cigarettes by 25 cents
- Find significant reduction in per-capita cigarette sales

! This paper is the standard example of SC

- Data: state-by-year panel
  - Years: 1970-2000
  - 38 control states, only 1 treated
- We use per-capita cigarette sales as the relevant outcome

Implementation using sdid

- The sdid package (Stata, R) can also be used to estimate DID and SC
- Install the package

```
ssc install sdid, replace
```

- Load the data

```
clear
webuse set www.damianclarke.net/stata/
webuse prop99_example.dta, clear
```

- SDID syntax

```
sdid Y G T D [if] [in] [weight] [, options]
```

DID and SC - Mind the line breaks when copying-and-pasting!

- DID - Graph comparing $Y_{g_1,t}$ vs $\frac{1}{|G_0|}\sum_{g\in G_0} Y_{g,t}$

```
sdid packspercapita state year treated, method(did) vce(noinference)
    graph
```

- SC - Graph comparing $Y_{g_1,t}$ vs $\sum_{g\in G_0} \hat{\omega}_g Y_{g,t}$

```
sdid packspercapita state year treated, method(sc) vce(noinference)
    graph
```

- SC - Graph comparing $\frac{1}{|T_1|}\sum_{t\in T_1} Y_{g_1,t}$ vs $\frac{1}{|T_1|}\sum_{t\in T_1} Y_{g,t}$, size $= \hat{\omega}_g$

```
sdid packspercapita state year treated, method(sc) vce(noinference)
    graph g1on
```

Next session

- Synthetic Differences-in-Differences
  - Estimation & TWFE representation
  - Unit and time synthetic weights
  - Double-robustness
- Staggered adoption design & SDID
  - Cohort-level SDID estimators
- Event study disaggregation
  - Event-time level SDID estimators

Thanks!