An Introduction to Modern Panel Data Methods: More on Synthetic Differences-in-Differences

Diego Ciccia

Kellogg School of Management, Northwestern University

Recap of last session

- Synthetic Differences-in-Differences
  - Unit and time weights
  - Enforcement of synthetic parallel trends
- Double robustnesss
  - SDID is well-behaved if *either* unit or time factors can be balanced out
- Cohort and event-time disaggregation
  - SDID aggregates estimators of cohort-specific and event-specific effects
- This session: **More on SDID**

Table of Contents

Sources

- Dmitry Arkhangelsky, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. 2021. *Synthetic Difference-in-Differences*. American Economic Review
- Damian Clarke, Daniel Pailañir, Susan Athey, and Guido Imbens. 2024. *On Synthetic Difference-in-Differences and Related Estimation Methods in Stata*. Stata Journal
- Clément de Chaisemartin and Xavier D'Haultfoeuille. 2026. *Causal Inference with Differences-in-Differences: Credible Answers to Hard Questions*. Book in progress

Asymptotic normality of SDID, I

- Arkhangelsky et al. (2021) prove the asymptotic distribution of $\hat{\tau}^{SDID}$
- They assume technical conditions, some of which are easy to interpret
  a. Either $|G_1|$ or $|T_1|$ grow large
  b. Both $|G_0|$ and $|T_0|$ grow large
  c. $|G_0|$ and $|T_0|$ grow at the same rate
  d. Both $\max \omega_g$ and $\max \lambda_t$ must go to zero
- Interpretation:
  a. Few treated units treated in few periods are ruled out
  b. Data must contain a large number of untreated units and periods
  c. Data cannot be rectangular $\rightarrow$ quasi-cross sections and long panels are ruled out
  d. Treated periods and units cannot be represented by a few untreated units or periods

Asymptotic normality of SDID, II

- Moreover, they assume that idiosyncratic errors are iid normal (homoskedasticity)
- Under these assumptions, they show that

$$\frac{\hat{\tau}^{SDID} - ATT}{\sqrt{V}} \xrightarrow{d} N(0,1)$$

  for some variance component V

- Some comments:
  - V is of the order $(|G_1||T_1|)^{-1}$, hence the relevance of Assumption (a)
  - V only depends on the variance of the idiosyncratic errors (and the design)
- If we are able to estimate V, then we can form a valid confidence interval fo ATT

$$\widehat{CI}_{0.95} = \hat{\tau}^{SDID} \pm 1.96\sqrt{\hat{V}}$$

Inference methods, I: Bootstrap

- We can approximate V via the **bootstrap** distribution of $\hat{\tau}^{SDID}$
- Bootstrap is perfomed by re-estimating SDID on (samples of) the current sample
  - Overall idea: we do not know the true distribution of $\hat{\tau}^{SDID}$
  - We assume that the data is our population, so $\hat{\tau}^{SDID}$ is the true parameter
  - We sample from the data and we re-estimate SDID
  - The set of SDID estimators forms some distribution (with some mean, variance, ...)
  - Under some conditions, this distribution is close to that of $\hat{\tau}^{SDID}$
- The simplest way of implementing bootstrap is to draw units **with replacement**
  - Draw G units from $\{1, .., G\}$, e.g. $(1, 1, 30, 5, 7, 17, 5, ...)$ (**Repeats are allowed**)
  - Estimate $\hat{\tau}_b^{SDID}$ on this sample and repeat for $b = 1, ..., B$
- The bootstrap estimator for V is

$$\hat{V}_{\text{bootstrap}} = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\tau}_b^{SDID} - \frac{1}{B}\sum_{b'=1}^{B}\hat{\tau}_{b'}^{SDID}\right)^2$$

Inference methods, II: Placebo inference

- With bootstrap, we may draw samples with no treated or no untreated units
  - If $|G_1| = 1$, the probability of drawing a fully untreated sample is ≈ 0.37 with $|G_0|$ large enough
- Solution: **placebo inference**
  - For each treated unit $g \in G_1$, randomly assign their $\mathbf{D}_g$ to an untreated unit
  - Keep only untreated units, and estimate $\hat{\tau}_p^{SDID}$ with this alternative treatment assignment, for $p = 1, ..., P$
- The placebo inference estimator for V is

$$\hat{V}_{\text{placebo}} = \frac{1}{B} \sum_{p=1}^{P} \left(\hat{\tau}_p^{SDID} - \frac{1}{P} \sum_{p'=1}^{P} \hat{\tau}_{p'}^{SDID} \right)^2$$

Inference methods, III: Jackknife

- Both bootstrap and placebo inference are computationally intensive
  - In case someone wants a coding project to work on: parallelization of SDID
- **Jackknife** methods work by dropping one observation at a time
  - The idea is that dispersion between observations drives the variance
  - Let $\hat{\tau}^{SDID}_{(-g)}$ be the SDID estimator computed by omitting unit g
- The jackknife estimator for V is

$$\hat{V}_{\text{jackknife}} = \frac{G-1}{G} \sum_{g \in G} \left(\hat{\tau}^{SDID}_{(-g)} - \hat{\tau}^{SDID} \right)^2$$

! The jackknife estimator is not defined with only 1 treated unit

Example with Stata: Bhalotra et al. (2023), redux

```
clear
webuse set www.damianclarke.net/stata/
webuse quota_example.dta, clear
keep if inlist(quotaYear, 2003, .)
```

- Bootstrap (default)

```
sdid womparl country year quota
```

- Placebo inference

```
sdid womparl country year quota, vce(placebo)
```

- Jackknife

```
sdid womparl country year quota, vce(jackknife)
```

Clustering, I

- Every inference procedure described so far assumes that units are independent
- This assumption fails in many settings (e.g., students within class)
- The solution is to redefine inference at a more aggregated level: **clusters**
  - We are agnostic about the dependence of units within clusters
  - Units in different clusters are assumed to be independent
- Ignoring between-group dependence may badly affect inference
  - See some coverage tests on SDID with clustered data

Clustering, II

- Fortunately, SDID allows for clustered inference
  - Clustered placebo inference requires the same number of units per cluster
- Load up version of Bhalotra et al. (2023) with geographical clusters

```
use "https://raw.githubusercontent.com/DiegoCiccia/sdid/main/
    simulations/cluster_data/quota_example_cluster.dta", clear
keep if inlist(quotaYear, 2003, .)
```

- Use clustered bootstrap (draw groups of countries with replacement)

```
sdid womparl country year quota, vce(bootstrap) cluster(area)
```

Controlling for covariates, I: Projected

- Outcomes are often affected by other variables than treatment alone
  - $\hat{\tau}^{SDID}$ may pick up differential trends of covariates that are not controlled for

- In general, SDID performs covariate adjustment linearly, i.e., replacing $Y_{g,t}$ with

$$Y_{g,t}^{\text{adj}} = Y_{g,t} - X_{g,t}'\hat{\pi}$$

  for some covariate $X_{g,t} \in \mathbb{R}^p$ and some coefficient $\hat{\pi}$

- This adjustment can be perfomed using two routines: **optimized** and **projected**

- The projected routine feeds in SDID the residual of a TWFE regression of $Y_{g,t}$ on $X_{g,t}$ in the sample of not-yet-treated, i.e., $D_{g,t} = 0$, that is,

$$(\hat{\pi}, .) = \arg\min_{\pi,\delta,\gamma} \sum_{g,t:D_{g,t}=0} (Y_{g,t} - \delta_g - \gamma_t - X_{g,t}'\pi)^2$$

Controlling for covariates, II: Optimized

- The optimized routine adds covariate adjustment to estimation of the weights

$$\ell_{\text{unit}}(\omega_0, \omega, \pi) = \sum_{t \in T_0} \left(\frac{1}{|G_1|} \sum_{g \in G_1} (Y_{g,t} - X'_{g,t}\pi) - \omega_0 - \sum_{g \in G_0} \omega_g (Y_{g,t} - X'_{g,t}\pi) \right)^2 + \xi_\omega \sum_{g \in G_0} \omega_g^2$$

$$\ell_{\text{time}}(\lambda_0, \lambda, \pi) = \sum_{g \in G_0} \left(\frac{1}{|T_1|} \sum_{t \in T_1} (Y_{g,t} - X'_{g,t}\pi) - \lambda_0 - \sum_{t \in T_0} \lambda_t (Y_{g,t} - X'_{g,t}\pi) \right)^2 + \xi_\lambda \sum_{t \in T_0} \lambda_t^2$$

$$(\hat{\omega}_0, \hat{\omega}, \hat{\lambda}_0, \hat{\lambda}, \hat{\pi}) = \underset{\substack{\omega_0, \lambda_0 \in \mathbb{R} \\ \pi \in \mathbb{R}^p \\ \omega \in [0,1]^{|G_0|} \\ \lambda \in [0,1]^{|T_0|}}}{\arg\min} \ \ell_{\text{unit}}(\omega_0, \omega, \pi) + \ell_{\text{time}}(\lambda_0, \lambda, \pi) \quad \text{s.t.} \quad \sum_{g \in G_0} \omega_g = 1, \sum_{t \in T_0} \lambda_t = 1$$

Controlling for covariates, III

- Both methods can be used in sdid
- Load a reduced version of Bhalotra et al. (2023)

```
webuse set www.damianclarke.net/stata/
webuse quota_example.dta, clear
drop if lngdp == .
keep if inlist(quotaYear, ., 2003)
```

- Optimized (default)

```
sdid womparl country year quota, vce(bootstrap) covariates(lngdp)
```

- Projected

```
sdid womparl country year quota, vce(bootstrap) covariates(lngdp,projected)
```

Diagnostics on SC and SDID

- The validity of DID is generally assessed via pre-trends tests
  - Under the null of parallel trends and no anticipation, differences in outcomes across the pre-periods should be zero in expectation
- It is natural to think of similar approaches to test the assumptions of SC and SDID
- Placebo tests of SC and SDID is still a venue for future research
- The diagnostic check that is often implemented in SC is **backtesting**
- The concern that backtesting addresses is:

Is the estimate due to actual treatment effects, or to the fact that treated periods are not used to form the weights?

Backtesting, I

- To purge this concern, one can think of setting up an **hold out** set of pre-periods
  - The intuition is similar to train and test subsets in machine learning
- Implementation:
  - Replace $\tilde{D}_{g,t} = 1$ for $g \in G_1$ and $t = |T_0| - p, ..., |T_0|$, for $p = 0, ..., |T_0| - 1$
  - This amounts to anticipating treatment rollout for treated units by $p + 1$ periods
  - Redefine $\tilde{T}_0$ and $\tilde{T}_1$, where $\tilde{T}_1$ now includes placebo and actual treated periods
  - Weights will only be computed on $\tilde{T}_0$, hence the hold out rationale
- Doubt is cast on the SC results if
  - Post-estimates are very different from using all T_0
  - Estimates in placebo post periods are large compared to actual post periods
- We can use `sdid_event` to make inference on event-time SC

Backtesting, II

- We can run a backtesting exercise on Abadie et al. (2010) with $p = 5$

```
webuse set www.damianclarke.net/stata/
webuse prop99_example.dta, clear
sdid packspercapita state year treated, method(sc) vce(noinference) ///
graph g2_opt(title("Start of the treatment = 1989"))

sort state year
bys state: gen treated_p_5 = treated[_n+5]
replace treated_p_5 = (state == "California") if missing(treated_p_5)

sdid packspercapita state year treated_p_5, method(sc) ///
vce(noinference) graph g2_opt(xline(1989, lc(red) lp(solid)) ///
title("Start of the treatment = 1984"))

sdid_event packspercapita state year treated_p_5, method(sc) vce(placebo)
```

Backtesting with SDID, I

- *Credible answers to hard questions*: backtesting cannot be directly used in SDID
  - If we backdate the treatment, $p+1$ hold out periods are not used for unit weights
  - But they are used to form time weights (as placebo post periods to be fitted)
  - As such, they are not really *held out*
- A workaround is to perform a **full placebo** exercise
  - Backdate the treatment by $p+1$ periods
  - Keep only observations in T_0
  - Run SDID using the placebo post periods as the *only* post periods
- Notice that the logic of the full placebo test is not (just) to check for overfitting
  - The overall idea is to have a version of SDID with only untreated units and periods

Backtesting with SDID, II

- We can run a full placebo exercise on Abadie et al. (2010) with $p = 5$

```
webuse set www.damianclarke.net/stata/
webuse prop99_example.dta, clear
sdid packspercapita state year treated, method(sdid) vce(placebo)

preserve
sort state year
bys state: gen treated_p_5 = treated[_n+5]
replace treated_p_5 = (state == "California") if missing(treated_p_5)
keep if year <= 1989
sdid_event packspercapita state year treated_p_5, method(sdid) vce(placebo)
restore
```

Conclusion

- A lot of questions related to the implementation of SDID are still open
  - DID and SC have been extensively studied
  - SDID is quite new, and not all results on DID and SC readily apply
- Nonetheless, SDID has received increased attention in applied work
  - This motivates future work in applied econometrics
- I conclude by pointing out some published papers using SDID as main specification
  - They may serve as a blueprint to implement SDID in other settings

	Title	Journal
Dench et al. (2024)	*The effects of post-Dobbs abortion bans on fertility*	JPubE
Beland et al. (2026)	*Curbing Pretextual Stops: Police Efficiency and Racial Disparities*	JPubE
Li and Zhao (2026)	*The Effectiveness of Carbon Emission Trading System: Evidence from China's Regional Markets.*	JDE
Flynn (2025)	*Contraceptive Access and Infant Health Outcomes*	JHE

A collection of papers published in economics (top field) journals with SDID as main specification

Thanks!