Sequential Synthetic Difference in Differences: a review

Alexis Ladasic

Trinity College Dublin

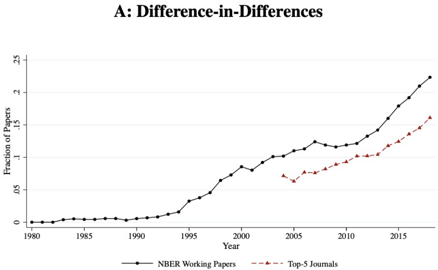
November 18, 2024

In summary

Growing literature

- Empirical research largely built on **event studies**: specific units are assigned to a treatment of interest and observed before, at the start and after treatment.
- The **difference-in-differences (DiD)** arguably one of the most used method to estimate treatment effect (e.g: policy change when some groups experience the shift while others don't)

Figure: Currie, et al. (2020)



Introduction and Literature Review (1/2)

- The **parallel trends (PT)** assumption is crucial for the validity of DiD
- Existing methods for **staggered (or sequential) adoption** designs often rely on the same identification assumption as traditional DiD
- **Synthethic DiD**: "most important innovation in the policy evaluation literature in the last 15 years" (Athey and Imbens, 2017).
    - Moves away from using a single control unit or simple average and instead uses weighted average
    - SDiD = "SC on steroids" (Fontana, 2024) as weights can change over time, potentially more treated units
- This paper proposes a new estimator, **Sequential Synthetic Difference in Differences (SSDD)**

Introduction and Literature Review (2/2)

1. **Sequential literature**:
   - See Bailey and Goodman-Bacon (2015): exploits variation in the timing of CHC establishment across counties between 1965-1974, comparing changes in mortality rates before and after CHCs began operating in treated counties to changes in mortality rates in untreated counties
   - Callaway and Sant'Anna (2021): define a group-time average treatment effects, $ATT(g,t)$, that is the average treatment effect in period t for the group of units first treated in period g.
2. **Synthethic literature**:
   - Alberto Abadie and Hainmueller (2010): minimum distance approach which requires a specific set of weights (nonnegative and sum to one)
   - Some interesting recent papers: Cunningham and Shah (2017), Magness and Makovi (2023)

Construction of the estimator (1/3)

Sequential Synthetic DiD

How to group?

$$Y_{a,t} = \frac{\sum_{i:A_i=a} Y_{i,t}}{n_a}$$

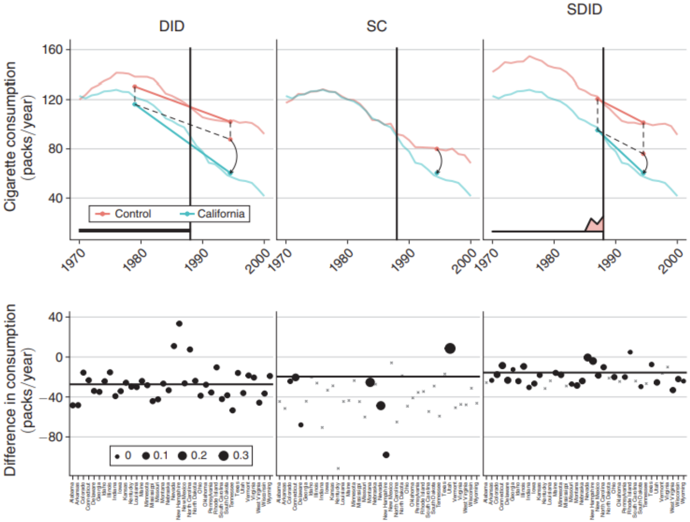
How to construct weights?

$$\lambda^{(a,k)}; \omega^{(a,k)}; \mu_a$$

Construction of the Estimator (2/3)

- The Sequential SDiD estimator adapts the Synthetic DiD (SDiD) estimator introduced in Arkhangelsky et al. (2021) to sequential settings
  - Check out the r code: synthdid
- The method proceeds in several steps (detailed maths in Annex)
  - Aggregate outcomes for units sharing the same adoption date.
  - For each adoption time and horizon, estimate the treatment effect using the SDiD estimator.
  - Use the resulting estimate to impute the missing counterfactual outcome for the treated units.
  - Repeat this exercise sequentially for all adoption times and horizons.
- Data-driven weights are used to enforce the parallel trends assumption in-sample

A comparison between DiD, SC & SDiD



Construction of the Estimator (3/3)

The last two steps are as follows:

$$\hat{\tau}_{a,k}^{\text{SSDiD}} = Y_{a,a+k} - \left(\sum_{j>a} \hat{\omega}_j^{(a,k)} Y_{j,a+k} \right) - \left[\sum_{l<a+k} \hat{\lambda}_l^{(a,k)} \left(Y_{a,l} - \sum_{j>a} \hat{\omega}_j^{(a,k)} Y_{j,l} \right) \right]$$

Then we can compute the ATT across adoption times:

$$\hat{\tau}_k^{\text{SSDiD}}(\mu) = \sum_{a \in \{a_{min},\dots,a_{max}\}} \mu_a \hat{\tau}_{a,k}^{\text{SSDiD}}$$

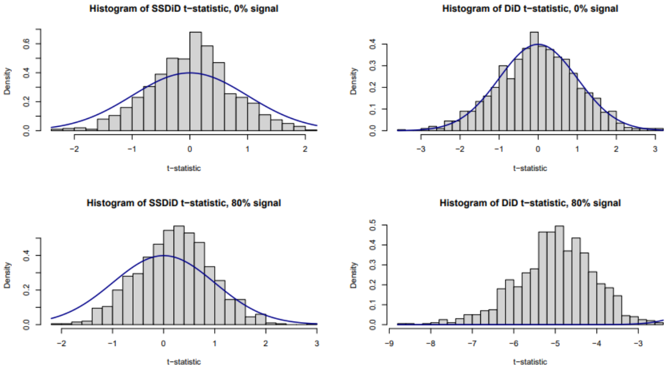
Key question here is the choice of the weights of each sequential estimator: authors choose weights that are proportional to the shares π_a.

Inference & Concrete example (1/3)

- Inference is conducted using Bayesian bootstrap
  - Introduced by Rubin (1981)
  - Key difference with frequentist version: parameters to be estimated are not considered fixed unknown values, but random variables that can be estimated from the sampling
- Authors construct weighted analogs of aggregated outcomes using independent exponential random variables $\boldsymbol{\xi} := \xi_i | \xi_i \sim Exp(1)$
  - Apply Algorithm 1 to the weighted outcomes (for the bayesian bootstrap)
  - Compute variance over the constructed estimator $\hat{\tau}_k^{SSDD}(\mu, \xi)$
  - Then use the variance to conduct more classical inference
- Application is done on Bailey and Goodman-Bacon (2015):
  1. Results show that the Sequential SDiD estimator produces estimates comparable to the standard DiD estimator
  2. Outperforms in the presence of a noisy signal (interactive fixed effects)

Inference & Concrete example (2/3)

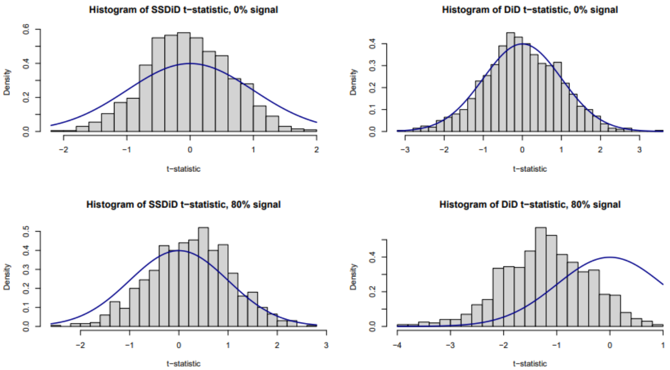
Figure 2: Distribution of t-statistics for τ_0



Notes: Each point corresponds to $\hat{\tau}_5$, with corresponding estimates constructed using standard DiD and Sequential SDiD as described in Algorithm 1. The grey dotted lines correspond to 95% confidence intervals constructed using Bayesian bootstrap described in Section 2 with 1000 simulations.

Inference & Concrete example (2/3)

Figure 3: Distribution of t-statistics for τ_4



Notes: Each point corresponds to $\hat{\tau}_4$, with corresponding estimates constructed using standard DiD and Sequential SDiD as described in Algorithm 1. The grey dotted lines correspond to 95% confidence intervals constructed using Bayesian bootstrap described in Section 2 with 1000 simulations.

Further Thoughts, Limits and Potential Applications

- SSDD offers several advantages: (i) sequential treatment rollout; (ii) robustness to violations of the PT; (iii) can handle interactive fixed effects and heterogenous effects in the treatment
- Limited by idiosyncratic errors, large adoption cohorts, computationally intensive
- Another good test to see how well SSDD is doing would probably be on anti-takeover laws
    - Literature is large on the question (Karpoff and Wittry, 2018) but heavily criticized (Baker et al., 2022)

Annex

1. Code
2. Mathematical construction of the aggregator
3. What's trending in DiD?

Code

Algorithm 1: Sequential SDiD

Data: Aggregated data, $a_{\min}$, $a_{\max}$, K, η

Result: $\{\hat{\tau}_{a,k}^{SSDiD}\}_{a\in\{a_{\max},\ldots,a_{\min}\}}^{k\in\{0,\ldots,K\}}$

for $k \in \{0,\ldots,K\}$ **do**

  for $a \in \{a_{\min},\ldots,a_{\max}\}$ **do**

    Construct the weights:

$$\hat{\omega}^{(a,k)} := \arg\min_{\sum_{j>a}\omega_j=1}\left\{\sum_{l<a+k}\left(\sum_{j>a}\omega_j Y_{j,l} - \omega_0 - Y_{a,l}\right)^2 + \eta^2\sum_{j>a}\omega_j^2\pi_j\right\},$$

$$\hat{\lambda}^{(a,k)} := \arg\min_{\sum_{l<a+k}\lambda_l=1}\left\{\sum_{j>a}\left(\sum_{l<a+k}\lambda_l Y_{j,l} - \lambda_0 - Y_{j,a}\right)^2 + \eta^2\sum_{l<a+k}\lambda_l^2\right\},$$

    Construct the estimator:

$$\hat{\tau}_{a,k}^{SSDiD} := \left(Y_{a,a+k} - \sum_{j>a}\hat{\omega}_j^{(a,k)}Y_{j,a+k}\right) - \sum_{l<a+k}\hat{\lambda}_l^{(a,k)}\left(Y_{a,l} - \sum_{j>a}\hat{\omega}_j^{(a,k)}Y_{j,l}\right)$$

    Define $Y_{a,a+k} := Y_{a,a+k} - \hat{\tau}_{a,k}^{SSDiD}$

  end

end

Step 1: Data Aggregation

We start by grouping units that adopt the treatment at the same time:

$$Y_{a,t} = \frac{\sum_{i:A_i=a} Y_{i,t}}{n_a}$$

Where:

- $Y_{a,t}$ is the average outcome for adoption cohort a at time t
- A_i is the adoption time for unit i
- n_a is the number of units in cohort a
- $Y_{i,t}$ is the outcome for unit i at time t

Step 2: Weight Construction (Unit Weights)

For each (a, k), we construct unit weights to find good comparisons:

$$\hat{\omega}^{(a,k)} =_{\sum_{j>a}\omega_j=1} \left\{ \sum_{l<a+k} \sum_{j>a} (\omega_j Y_{j,l} - \omega_0 - Y_{a,l})^2 + \eta^2 \sum_{j>a} \frac{\omega_j^2}{\pi_j} \right\}$$

Where:

- ω_j are weights for control units (treated later)
- η^2 is a regularization term
- π_j are cohort shares (n_j/n, where n is total sample size)
- l indexes time periods before treatment

This creates a "synthetic control" that matches pre-treatment outcomes of the treated unit.

Step 2 (continued): Weight Construction (Time Weights)

We also construct time weights to balance across time periods:

$$\hat{\lambda}^{(a,k)} =_{\sum_{l<a+k} \lambda_l = 1} \left\{ \sum_{j>a} \sum_{l<a+k} (\lambda_l Y_{j,l} - \lambda_0 - Y_{j,a+k})^2 + \eta^2 \sum_{l<a+k} \lambda_l^2 \right\}$$

Where:

- λ_l are weights for pre-treatment time periods
- η^2 is the same regularization term as before

Step 3: Treatment Effect Estimation

We estimate the treatment effect for each (a, k):

$$\hat{\tau}_{a,k}^{\text{SSDiD}} = Y_{a,a+k} - \left(\sum_{j>a} \hat{\omega}_j^{(a,k)} Y_{j,a+k} \right) - \left[\sum_{l<a+k} \hat{\lambda}_l^{(a,k)} \left(Y_{a,l} - \sum_{j>a} \hat{\omega}_j^{(a,k)} Y_{j,l} \right) \right]$$

Where:

- $Y_{a,a+k}$ is the actual outcome k periods after treatment
- $\sum_{j>a} \hat{\omega}_j^{(a,k)} Y_{j,a+k}$ is the weighted control outcome
- The last term adjusts for pre-existing differences

This compares the treated unit's outcome to a weighted average of control units, adjusting for pre-treatment differences.

Step 4: Outcome Adjustment

After computing $\hat{\tau}_{a,k}^{\text{SSDiD}}$, we update the outcome:

$$Y_{a,a+k} := Y_{a,a+k} - \hat{\tau}_{a,k}^{\text{SSDiD}}$$

This step:

- "Removes" the estimated treatment effect
- Allows the method to "learn" and improve estimates over time
- Key to handling dynamic treatment effects

Step 5: Aggregation of Estimates

Finally, we compute the effect across adoption times:

$$\hat{\tau}_k^{\text{SSDiD}}(\mu) = \sum_{a \in \{a_{min}, \ldots, a_{max}\}} \mu_a \hat{\tau}_{a,k}^{\text{SSDiD}}$$

Where:

- μ_a are user-specified weights
- Often set proportional to cohort sizes: $\mu_a = \frac{\pi_a}{\sum_{a \in \{a_{min}, \ldots, a_{max}\}} \pi_a}$
- a_{min} and a_{max} are the earliest and latest adoption times considered

This gives us our final estimate of the treatment effect k periods after adoption, allowing flexible weighting of different adoption cohorts.

Trends in modern DiD (Sant'Anna, 2024)

Recent boom of DiD methods

In the last years, we have seen many methodological advances in DiD: (by no means an exhaustive list)

- Athey and Imbens (2022)
- Borusyak, Jaravel and Spiess (2024)
- de Chaisemartin and D'Haultfœuille (2020)
- Goodman-Bacon (2021)
- Sun and Abraham (2021)

"Backward Engineering" causal interpretations for TWFE regressions and propose some alternative DiD estimators

- Callaway and Sant'Anna (2021)
- Sant'Anna and Zhao (2020)
- Lee and Wooldridge (2023)
- Wooldridge (2021)

"Forward Engineering" DiD estimators conditional on covariates

- Rambachan and Roth (2023)
- Roth (2022)

Issues with pre-tests and how to handle PT as approximation

- Roth and Sant'Anna (2023a,b)

Sensitivity to functional form and random treatment timing

- Callaway, Goodman-Bacon and Sant'Anna (2024a)
- de Chaisemartin, D'Haultfoeuille, Pasquier, Sow and Vazquez-Bare (2024)

DiD with continuous and multi-valued treatments

- Ghanem, Sant'Anna and Wüthrich (2022)
- Marx, Tamer and Tang (2023)

Better understanding PT and selection

- Callaway (2021)
- Callaway and Li (2019)
- Tchetgen Tchetgen, Park and Richardson (2024)
- Wooldridge (2023)

Nonlinear DiD Models

EMORY UNIVERSITY 2

Alberto Abadie, A. D. and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118.

Athey, S. and Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32.

Bailey, M. J. and Goodman-Bacon, A. (2015). The war on poverty's experiment in public medicine: Community health centers and the mortality of older americans. *American Economic Review*, 105(3):1067–1104.

Baker, A. C., Larcker, D. F., and Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2):370–395.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230. Themed Issue: Treatment Effect 1.

Cunningham, S. and Shah, M. (2017). Decriminalizing Indoor Prostitution: Implications for Sexual Violence and Public Health. *The Review of Economic Studies*, 85(3):1683–1715.

Karpoff, J. and Wittry, M. D. (2018). Institutional and legal context in natural experiments: The case of state antitakeover laws. *The Journal of Finance*, 73(2):657–714.

Magness, P. W. and Makovi, M. (2023). The mainstreaming of marx: Measuring the effect of the russian revolution on karl marx's influence. *Journal of Political Economy*, 131(6):1507–1545.

Rubin, D. B. (1981). The Bayesian Bootstrap. *The Annals of Statistics*, 9(1):130 – 134.