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

Diego Ciccia

Kellogg School of Management, Northwestern University

Recap of last session

- Potential outcome framework can be extended to a panel data setting
  - No interference, no anticipation and no dynamics $\rightarrow$ matrix representation
  - ATT is well defined in the model-based perspective
- Realized outcomes are connected to the design via the latent factor model
  - Fundamental problem of causal inference: ATT is not identified
- DID and SC offer solutions under different assumptions
  - DID: parallel trends
  - SC: factor model + convex hull representation
- This session: **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
- (New book!) Damian Clarke. 2026. *Applied Microeconometrics*. MIT Press

Preliminaries

- Consider a non-staggered design, with (T_0, T_1) periods and (G_0, G_1) units
- **SDID** embeds properties of both DID and SC
  - vs. DID: weighted diff-in-diff estimator + synthetic parallel trends
  - vs. SC: both units and time weights are used
- The use of both unit and time weights ensures additional **robustness**

Key intuition

If **synthetic control** fails, could we put some **synthetic pre-period** to work?

Introducing SDID as a weighted fixed effects regression

- Recall from the previous session that

$$(\hat{\tau}^{DID}, .) = \underset{(\tau, \delta_g, \gamma_t)}{\arg\min} \sum_{g=1}^{G} \sum_{t=1}^{T} (Y_{g,t} - \delta_g - \gamma_t - \tau \cdot D_{g,t})^2 \tag{DID}$$

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

- The **Synthetic Differences-in-Differences** estimator $\hat{\tau}^{SDID}$ is defined as

$$(\hat{\tau}^{SDID}, .) = \underset{(\tau, \delta_g, \gamma_t)}{\arg\min} \sum_{g=1}^{G} \sum_{t=1}^{T} (Y_{g,t} - \delta_g - \gamma_t - \tau \cdot D_{g,t})^2 \hat{\omega}_g^{SDID} \hat{\lambda}_t^{SDID} \tag{SDID}$$

Synthetic weights, I

- Exactly as in SC, weights are **estimated** via a separate procedure (notice the hats!)
- These procedures are penalized, constrained linear regressions with constants

$$\hat{\omega} = \underset{\omega_0 \in \mathbb{R}, \omega \in [0,1]^{G_0}}{\arg\min} \sum_{t \in T_0} \left(\frac{1}{|G_1|} \sum_{g \in G_1} Y_{g,t} - \omega_0 - \sum_{g \in G_0} \omega_g Y_{g,t} \right)^2 + \xi_\omega \sum_{g \in G_0} \omega_g^2$$

$$\hat{\lambda} = \underset{\lambda_0 \in \mathbb{R}, \lambda \in [0,1]^{T_0}}{\arg\min} \sum_{g \in G_0} \left(\frac{1}{|T_1|} \sum_{t \in T_1} Y_{g,t} - \lambda_0 - \sum_{t \in T_0} \lambda_t Y_{g,t} \right)^2 + \xi_\lambda \sum_{t \in T_0} \lambda_t^2$$

s.t. $\sum_{g \in G_0} \omega_g = 1 \qquad \sum_{t \in T_0} \lambda_t = 1$

Synthetic weights, I

- Exactly as in SC, weights are **estimated** via a separate procedure (notice the hats!)
- These procedures are penalized, constrained linear regressions with constants

$$\hat{\omega} = \underset{\omega_0 \in \mathbb{R}, \omega \in [0,1]^{G_0}}{\arg\min} \sum_{t \in T_0} \left(\frac{1}{|G_1|} \sum_{g \in G_1} Y_{g,t} - \omega_0 - \sum_{g \in G_0} \omega_g Y_{g,t} \right)^2 + \xi_\omega \sum_{g \in G_0} \omega_g^2$$

$$\hat{\lambda} = \underset{\lambda_0 \in \mathbb{R}, \lambda \in [0,1]^{T_0}}{\arg\min} \sum_{g \in G_0} \left(\frac{1}{|T_1|} \sum_{t \in T_1} Y_{g,t} - \lambda_0 - \sum_{t \in T_0} \lambda_t Y_{g,t} \right)^2 + \xi_\lambda \sum_{t \in T_0} \lambda_t^2$$

s.t. $\sum_{g \in G_0} \omega_g = 1 \qquad \sum_{t \in T_0} \lambda_t = 1$

Synthetic weights, I

- Exactly as in SC, weights are **estimated** via a separate procedure (notice the hats!)
- These procedures are penalized, constrained linear regressions with constants

$$\hat{\omega} = \underset{\omega_0 \in \mathbb{R}, \omega \in [0,1]^{G_0}}{\arg\min} \sum_{t \in T_0} \left(\frac{1}{|G_1|} \sum_{g \in G_1} Y_{g,t} - \omega_0 - \sum_{g \in G_0} \omega_g Y_{g,t} \right)^2 + \xi_\omega \sum_{g \in G_0} \omega_g^2$$

$$\hat{\lambda} = \underset{\lambda_0 \in \mathbb{R}, \lambda \in [0,1]^{T_0}}{\arg\min} \sum_{g \in G_0} \left(\frac{1}{|T_1|} \sum_{t \in T_1} Y_{g,t} - \lambda_0 - \sum_{t \in T_0} \lambda_t Y_{g,t} \right)^2 + \xi_\lambda \sum_{t \in T_0} \lambda_t^2$$

s.t. $\sum_{g \in G_0} \omega_g = 1 \qquad \sum_{t \in T_0} \lambda_t = 1$

Synthetic weights, I

- Exactly as in SC, weights are **estimated** via a separate procedure (notice the hats!)
- These procedures are penalized, constrained linear regressions with constants

$$\hat{\omega} = \underset{\omega_0 \in \mathbb{R}, \omega \in [0,1]^{G_0}}{\arg\min} \sum_{t \in T_0} \left(\frac{1}{|G_1|} \sum_{g \in G_1} Y_{g,t} - \omega_0 - \sum_{g \in G_0} \omega_g Y_{g,t} \right)^2 + \xi_\omega \sum_{g \in G_0} \omega_g^2$$

$$\hat{\lambda} = \underset{\lambda_0 \in \mathbb{R}, \lambda \in [0,1]^{T_0}}{\arg\min} \sum_{g \in G_0} \left(\frac{1}{|T_1|} \sum_{t \in T_1} Y_{g,t} - \lambda_0 - \sum_{t \in T_0} \lambda_t Y_{g,t} \right)^2 + \xi_\lambda \sum_{t \in T_0} \lambda_t^2$$

s.t. $\sum_{g \in G_0} \omega_g = 1 \qquad \sum_{t \in T_0} \lambda_t = 1$

Synthetic weights, I

- Exactly as in SC, weights are **estimated** via a separate procedure (notice the hats!)
- These procedures are penalized, constrained linear regressions with constants

$$\hat{\omega} = \underset{\omega_0 \in \mathbb{R}, \omega \in [0,1]^{G_0}}{\arg\min} \sum_{t \in T_0} \left(\frac{1}{|G_1|} \sum_{g \in G_1} Y_{g,t} - \omega_0 - \sum_{g \in G_0} \omega_g Y_{g,t} \right)^2 + \xi_\omega \sum_{g \in G_0} \omega_g^2$$

$$\hat{\lambda} = \underset{\lambda_0 \in \mathbb{R}, \lambda \in [0,1]^{T_0}}{\arg\min} \sum_{g \in G_0} \left(\frac{1}{|T_1|} \sum_{t \in T_1} Y_{g,t} - \lambda_0 - \sum_{t \in T_0} \lambda_t Y_{g,t} \right)^2 + \xi_\lambda \sum_{t \in T_0} \lambda_t^2$$

s.t. $\sum_{g \in G_0} \omega_g = 1 \qquad \sum_{t \in T_0} \lambda_t = 1$

- Role of constants: enforce synthetic parallel trends

Synthetic weights, II - Visualizing the construction of the weights

- Consider a simple setting with 1 treated unit and 1 treated period

$$Y = \begin{pmatrix} Y_{1,1} & Y_{1,2} & \cdots & Y_{1,T_0} & Y_{1,T_0+1} \\ Y_{2,1} & Y_{2,2} & \cdots & Y_{2,T_0} & Y_{2,T_0+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Y_{G_0,1} & Y_{G_0,2} & \cdots & Y_{G_0,T_0} & Y_{G_0,T_0+1} \\ Y_{G_0+1,1} & Y_{G_0+1,2} & \cdots & Y_{G_0+1,T_0} & Y_{G_0+1,T_0+1} \end{pmatrix}_{(G_0,1)\times(T_0,1)}$$

Synthetic weights, II - Visualizing the construction of the weights

- Consider a simple setting with 1 treated unit and 1 treated period

$$Y = \begin{pmatrix} Y_{1,1} & Y_{1,2} & \cdots & Y_{1,T_0} & Y_{1,T_0+1} \\ Y_{2,1} & Y_{2,2} & \cdots & Y_{2,T_0} & Y_{2,T_0+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Y_{G_0,1} & Y_{G_0,2} & \cdots & Y_{G_0,T_0} & Y_{G_0,T_0+1} \\ Y_{G_0+1,1} & Y_{G_0+1,2} & \cdots & Y_{G_0+1,T_0} & Y_{G_0+1,T_0+1} \end{pmatrix}_{(G_0,1)\times(T_0,1)}$$

- Unit weights: within time (pre-period), control units predict treated units

Synthetic weights, II - Visualizing the construction of the weights

- Consider a simple setting with 1 treated unit and 1 treated period

$$Y = \begin{pmatrix} Y_{1,1} & Y_{1,2} & \cdots & Y_{1,T_0} & Y_{1,T_0+1} \\ Y_{2,1} & Y_{2,2} & \cdots & Y_{2,T_0} & Y_{2,T_0+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Y_{G_0,1} & Y_{G_0,2} & \cdots & Y_{G_0,T_0} & Y_{G_0,T_0+1} \\ Y_{G_0+1,1} & Y_{G_0+1,2} & \cdots & Y_{G_0+1,T_0} & Y_{G_0+1,T_0+1} \end{pmatrix}_{(G_0,1)\times(T_0,1)}$$

- Unit weights: within time (pre-period), control units predict treated units
- Time weights: within unit (control), untreated periods predict treated periods

Synthetic weights, II - Visualizing the construction of the weights

- Consider a simple setting with 1 treated unit and 1 treated period

$$Y = \begin{pmatrix} Y_{1,1}(0) & Y_{1,2}(0) & \dots & Y_{1,T_0}(0) & Y_{1,T_0+1}(0) \\ Y_{2,1}(0) & Y_{2,2}(0) & \dots & Y_{2,T_0}(0) & Y_{2,T_0+1}(0) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Y_{G_0,1}(0) & Y_{G_0,2}(0) & \dots & Y_{G_0,T_0}(0) & Y_{G_0,T_0+1}(0) \\ Y_{G_0+1,1}(0) & Y_{G_0+1,2}(0) & \dots & Y_{G_0+1,T_0}(0) & Y_{G_0+1,T_0+1}(1) \end{pmatrix}_{(G_0,1)\times(T_0,1)}$$

- Unit weights: within time (pre-period), control units predict treated units
- Time weights: within unit (control), untreated periods predict treated periods
- Either case, only untreated outcomes are used to estimate weights
- **Goal**: partial out the unknown factor structure of untreated outcomes

Synthetic weights, III

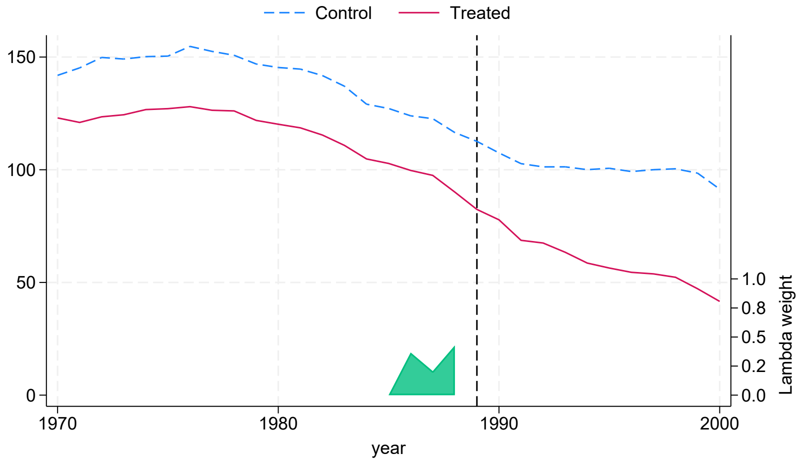
- Unit weights:
  - Regress avg. Y of treated units on control units' Y in the pre-period
  - **Goal**: Predict treated units' outcomes in pre-period via control units
  - $\rightarrow$ Unit-level component of untreated outcomes

- Time weights:
  - Regress avg. Y of control units in the post-period on control units' Y in the pre
  - **Goal**: Predict control units' outcomes in post-period via control units' pre-period
  - $\rightarrow$ Time-level component of untreated outcomes

- Final weights:

$$\hat{\omega}_g^{SDID} = \begin{cases} \hat{\omega}_g & \text{if } g \in G_0 \\ \frac{1}{|G_1|} & \text{if } g \in G_1 \end{cases} \qquad \hat{\lambda}_t^{SDID} = \begin{cases} \hat{\lambda}_t & \text{if } t \in T_0 \\ \frac{1}{|T_1|} & \text{if } t \in T_1 \end{cases}$$

SDID on Abadie et al. (2010)



Synthetic Differences-in-Differences estimator, I

- The solution to (SDID) can be written down as

$$\hat{\tau}^{SDID} = \frac{1}{|T_1|} \sum_{t \in T_1} \left(\frac{1}{|G_1|} \sum_{g \in G_1} Y_{g,t} - \sum_{g \in G_0} \hat{\omega}_g Y_{g,t} \right) - \sum_{t \in T_0} \hat{\lambda}_t \left(\frac{1}{|G_1|} \sum_{g \in G_1} Y_{g,t} - \sum_{g \in G_0} \hat{\omega}_g Y_{g,t} \right)$$

- The SDID estimator ...
  - ... compares *synthetic* differential outcomes, i.e., treated units vs synthetic control ...
  - ... between post and *synthetic* pre-period, i.e., pre-periods weighted by $\hat{\lambda}_t$

Synthetic Differences-in-Differences estimator, II

- The solution to (SDID) can *also* be written down as

$$\hat{\tau}^{SDID} = \frac{1}{|G_1|} \sum_{g \in G_1} \left(\frac{1}{|T_1|} \sum_{t \in T_1} Y_{g,t} - \sum_{t \in T_0} \hat{\lambda}_t Y_{g,t} \right) - \sum_{g \in G_0} \hat{\omega}_g \left(\frac{1}{|T_1|} \sum_{t \in T_1} Y_{g,t} - \sum_{t \in T_0} \hat{\lambda}_t Y_{g,t} \right)$$

- The SDID estimator ...
  - ... compares *synthetic* trends, i.e., treated post vs synthetic pre-period ...
  - ... between treated and *synthetic* control units, i.e., untreated units weighted by $\hat{\omega}_g$
- This equivalence is strongly related to the *double robustness* property of SDID
  - When unit weights fail, time weights can still achieve unbiasedness

Double robustness, I

- In order to explain double robustness, I make two simplifying assumptions
  1. There is only one treated unit (g_1) and one treated period (t_1)
  2. Weights are non-stochastic, i.e., $\hat{\omega}$ and $\hat{\lambda}$ are just some fixed numbers
- Assumption 1 is immaterial, it just simplifies notation
  - Every step goes through by replacing Y_{g_1,t_1} with some time/group average
- Assumption 2 is stronger, but holds for so-called *oracle* weights
  - $(\hat{\omega}, \hat{\lambda})$ are random, since they are regression coefficients from a random sample
  - Oracle weights solve a *population* version of the weights' regressions
  - Arkhangelsky et al. (2021) discuss conditions under which $(\hat{\omega}, \hat{\lambda})$ are *close* to oracle
  - To enforce non-randomness, I replace $(\hat{\omega}, \hat{\lambda})$ with (ω, λ)

Double robustness, II

- Under these assumptions, the SDID estimator reduces to

$$\tilde{\tau}^{SDID} = \left(Y_{g_1,t_1} - \sum_{g \in G_0} \omega_g Y_{g,t_1} \right) - \sum_{t \in T_0} \lambda_t \left(Y_{g_1,t} - \sum_{g \in G_0} \omega_g Y_{g,t} \right)$$

- Consider taking the expectation of $\tilde{\tau}^{SDID}$ and add-and-subtract $\mathbb{E}[Y_{g_1,t_1}(0)]$

$$\mathbb{E}[\tilde{\tau}^{SDID}] = \mathbb{E}[Y_{g_1,t_1}(1) - Y_{g_1,t_1}(0)]$$
$$+ \left(\mathbb{E}[Y_{g_1,t_1}(0)] - \sum_{g \in G_0} \omega_g \mathbb{E}[Y_{g,t_1}(0)] \right) - \sum_{t \in T_0} \lambda_t \left(\mathbb{E}[Y_{g_1,t}(0)] - \sum_{g \in G_0} \omega_g \mathbb{E}[Y_{g,t}(0)] \right)$$

Double robustness, II

- Under these assumptions, the SDID estimator reduces to

$$\tilde{\tau}^{SDID} = \left(Y_{g_1,t_1} - \sum_{g \in G_0} \omega_g Y_{g,t_1}\right) - \sum_{t \in T_0} \lambda_t \left(Y_{g_1,t} - \sum_{g \in G_0} \omega_g Y_{g,t}\right)$$

- Consider taking the expectation of $\tilde{\tau}^{SDID}$ and add-and-subtract $\mathbb{E}[Y_{g_1,t_1}(0)]$

$$\mathbb{E}[\tilde{\tau}^{SDID}] = \overbrace{\mathbb{E}[Y_{g_1,t_1}(1) - Y_{g_1,t_1}(0)]}^{\text{ATT}}$$
$$+ \underbrace{\left(\mathbb{E}[Y_{g_1,t_1}(0)] - \sum_{g \in G_0} \omega_g \mathbb{E}[Y_{g,t_1}(0)]\right) - \sum_{t \in T_0} \lambda_t \left(\mathbb{E}[Y_{g_1,t}(0)] - \sum_{g \in G_0} \omega_g \mathbb{E}[Y_{g,t}(0)]\right)}_{\text{Bias}}$$

Double robustness, III - Factor model assumption

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

- Let's use the factor model assumption to decompose the bias:

$$\text{Bias} = \left(\mathbb{E}[Y_{g_1,t_1}(0)] - \sum_{g \in G_0} \omega_g \mathbb{E}[Y_{g,t}(0)] \right) - \sum_{t \in T_0} \lambda_t \left(\mathbb{E}[Y_{g_1,t}(0)] - \sum_{g \in G_0} \omega_g \mathbb{E}[Y_{g,t}(0)] \right)$$

$$= \left(\Delta_{g_1} - \sum_{g \in G_0} \omega_g \Delta_g \right)' \left(\Gamma_{t_1} - \sum_{t \in T_0} \lambda_t \Gamma_t \right)$$

Double robustness, IV

$$\text{Bias} = \left(\Delta_{g_1} - \sum_{g \in G_0} \omega_g \Delta_g\right)' \left(\Gamma_{t_1} - \sum_{t \in T_0} \lambda_t \Gamma_t\right)$$

- If either
  1. treated unit factor is in convex hull of untreated units' factors:
     $\Delta_{g_1} = \sum_{g \in G_0} \omega_g \Delta_g$
  2. untreated period factor is in convex hull of untreated periods' factors
     $\Gamma_{t_1} = \sum_{t \in T_0} \lambda_t \Gamma_t$

  then, bias $= 0$.

- Contrary to SC, SDID is unbiased when either unit or time factors are balanced out

Double robustness, V - Why the definition of the weights works

$$\text{Bias} = \left(\Delta_{g_1} - \sum_{g \in G_0} \omega_g \Delta_g\right)' \left(\Gamma_{t_1} - \sum_{t \in T_0} \lambda_t \Gamma_t\right)$$

- Bias $= 0$ also if weights balance out the systematic components up to a constant
- Consider the unit weights definition, and assume that an exact solution is found

$$Y_{g_1,t} = \omega_0 + \sum_{g \in G_0} \omega_g Y_{g,t} \quad \Longrightarrow \quad \left(\Delta_{g_1} - \sum_{g \in G_0} \omega_g \Delta_g\right)' \Gamma_t = \omega_0$$

$$\Longrightarrow \quad \left(\Delta_{g_1} - \sum_{g \in G_0} \omega_g \Delta_g\right)' \left(\Gamma_t - \Gamma_{t-1}\right) = 0$$

- Constants in synthetic weights are there to be differenced out, yielding bias $= 0$

What is SDID unbiased for?

- A great selling point of SDID is that G_1 and T_1 are not required to be small
- With multiple treated periods and units, the parameter that SDID is unbiased for is

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

- This is the same ATT we wished to recover with DID under parallel trends

Staggered adoption algorithm, I

- TWFE estimators get into trouble when units' treatment starts in different periods
- SDID avoids this type of concerns by running *separate* estimation procedures
- Data is split in subdatasets, each corresponding to a single treatment timing

$$\mathbf{D} = \begin{pmatrix} (g,t) & 1 & 2 & 3 & 4 \\ 1 & 0 & 0 & 1 & 1 \\ 2 & 0 & 0 & 0 & 1 \\ 3 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 \end{pmatrix} \rightarrow \begin{matrix} \mathbf{D}_1 = \begin{pmatrix} (g,t) & 1 & 2 & 3 & 4 \\ 1 & 0 & 0 & 1 & 1 \\ 3 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 \end{pmatrix} \\ \\ \mathbf{D}_2 = \begin{pmatrix} (g,t) & 1 & 2 & 3 & 4 \\ 2 & 0 & 0 & 0 & 1 \\ 3 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 \end{pmatrix} \end{matrix}$$

Staggered adoption algorithm, II: some notation

- Let A be the set of periods at which some cohort's treatment started, e.g., $\{3, 4\}$
- Each treated cohort is denoted by its first treated period
  - Cohort $a = 3$ is the group of treated units whose treatment started at $t = 3$
- Let $G_{1,a}$ be the set of treated units in cohort a
- Let $T_{0,a}$ and $T_{1,a}$ be the set of pre and post periods for cohort a
- The total number of post-periods for cohort a is $|T_{1,a}| = T - a + 1$
- The total number of post periods across all treated units is

$$T^{tot} = \sum_{a \in A} |G_{1,a}| \cdot |T_{1,a}|$$

Staggered adoption algorithm, III

- SDID is run on *each* of these datasets
  - Let $\hat{\tau}_a^{SDID}$ be the SDID estimator using only never-treated units and cohort a.
  - ! Important detail: weights $(\hat{\omega}, \hat{\lambda})$ are recomputed for each cohort.

- The aggregated SDID estimator is

$$\hat{\tau}^{SDID} = \sum_{a \in A} \frac{|G_{1,a}| \cdot |T_{1,a}|}{T^{tot}} \cdot \hat{\tau}_a^{SDID}$$

- $\hat{\tau}^{SDID}$ is an estimator for

$$ATT = \frac{1}{T^{tot}} \sum_{a \in A} \sum_{g \in G_{1,a}} \sum_{t \in T_{1,a}} \mathbb{E}[Y_{g,t}(1) - Y_{g,t}(0)]$$

Disaggregating the ATT, I

- ATT is actually some average of period-specific treatment effects
- With one cohort, the ATT can be rearranged as

$$ATT = \frac{1}{|T_1|} \sum_{\ell=1}^{|T_1|} \frac{1}{|G_1|} \sum_{g \in G_1} \mathbb{E}\left[Y_{g,|T_0|+\ell}(1) - Y_{g,|T_0|+\ell}(0)\right]$$

$$= \frac{1}{|T_1|} \sum_{\ell=1}^{|T_1|} ATT_\ell$$

- ATT_ℓ is the ATT, ℓ periods after the start of the treatment
- ATT_ℓ is defined for $|T_1| = T - |T_0|$ periods, i.e., the total number of post-periods

Disaggregating the ATT, II

- With more than one cohort, ATT_ℓ may not be defined for some cohorts
  - e.g., $T = 4$ and $A = \{3, 4\}$, ATT_2 only exists for $a = 3$
- As such, we need a bit more of notation
- Let $\underline{a}$ be the the first treated cohort, i.e., $\underline{a} = \min(A)$
- The maximum number of post treatment periods in a staggered design is

$$L = T - \underline{a} + 1$$

! This is also a *quick maths* trick for other uses (e.g., running `did_multiplegt`)

- Let A_ℓ be the set of cohorts for which the ℓ-th post period exists
  - e.g., $T = 4$ and $A = \{3, 4\}$, $A_1 = \{3, 4\}$ and $A_2 = \{3\}$

Disaggregating the ATT, III - A matrioska of ATTs

- Let $N_{1,\ell} = \sum_{a \in A_\ell} |G_{1,a}|$ be the total number of units that have an ℓ-post period
- With more than one cohort, the ATT can be disaggregated as

$$ATT = \sum_{\ell=1}^{L} \frac{N_{1,\ell}}{T^{tot}} ATT_\ell$$

where

$$ATT_\ell = \sum_{a \in A_\ell} \frac{|G_{1,a}|}{N_{1,\ell}} ATT_{a,\ell} \qquad ATT_{a,\ell} = \frac{1}{|G_{1,a}|} \sum_{g \in G_{1,a}} \mathbb{E}[Y_{g,a-1+\ell}(1) - Y_{g,a-1+\ell}(0)]$$

- Finally, ATT_ℓ is the ATT across cohorts, ℓ periods after the start of each treatment

Disaggregating SDID - Another matrioska

- The building block to disaggregate SDID in a cohort-specific ATT estimator is

$$\hat{\tau}_{a,\ell}^{SDID} = \frac{1}{|G_{1,a}|} \sum_{g \in G_{1,a}} Y_{g,a-1+\ell} - \sum_{g \in G_0} \hat{\omega}_{a,g} Y_{g,a-1+\ell} - \sum_{t \in T_{0,a}} \hat{\lambda}_{a,t} \left(\frac{1}{|G_{1,a}|} \sum_{g \in G_{1,a}} Y_{g,t} - \sum_{g \in G_0} \hat{\omega}_{a,g} Y_{g,t} \right)$$

- $\hat{\tau}_{a,\ell}^{SDID}$ is the cohort-specific ℓ-th synthetic event study

! The benchmark is the whole (synthetic) pre-period, instead of in $a-1$

- Using the same steps as for the ATT, it holds that

$$\hat{\tau}_{\ell}^{SDID} = \sum_{a \in A_\ell} \frac{|G_{1,a}|}{N_{1,\ell}} \hat{\tau}_{a,\ell}^{SDID} \qquad \hat{\tau}^{SDID} = \sum_{\ell=1}^{L} \frac{N_{1,\ell}}{T^{tot}} \hat{\tau}_{\ell}^{SDID}$$

Stata packages for implementing SDID

- Estimators discussed in these slides can be implemented using the `sdid` suite
- Packages are available in Stata and R, although Stata is much better maintained
- Install `sdid` (if you have not done that last session)

```
ssc install sdid, replace
```

- Install `sdid_event` and dependencies

```
ssc install unique, replace
ssc install sdid_event, replace
```

Bhalotra et al. (2023)

- Impact of parliamentary gender quotas on maternal mortality
- Find sustained 7%-12% reductions in female mortality due to child birth
- Mechanism: more women in parliament

! This paper uses almost all the modern DiD tools available in 2021

- Data: country-by-year panel
  - Years: 1990-2015
  - Countries: 115, 9 of which implement a parliamentary gender quota reform
  - Countries introduce the reform in different years → staggered design
- We use female parliamentary participation as the relevant outcome

Data loading

- Load the data

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

- Get a sense of treatment timing: detect cohorts

```
preserve
collapse (first) quotaYear, by(country)
tostring quotaYear, gen(cohort)
replace cohort = "Never-treated" if cohort == "."
label var cohort "Cohort"
tab cohort
restore
```

Running SDID, I

- SDID syntax

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

- Estimation only (no inference):

```
sdid womparl country year quota, vce(noinference) graph
```

- Retrieve weights:

```
mat li e(lambda)
mat li e(omega)
```

Running SDID, II

- SDID event-study:

```
sdid_event womparl country year quota, vce(noinference)
```

- Cohort-level estimators from sdid

```
sdid womparl country year quota, vce(noinference)
mat li e(tau)
```

- Full disaggregation via sdid_event

```
sdid_event womparl country year quota, vce(noinference) disag
```

Next session

- Inference:
  - Asymptotic normality of SDID estimator
  - Bootstrap, placebo and jackknife inference
  - Clustering
- Covariates:
  - Projected and optimized methods for covariate adjustment
- Placebo tests:
  - Backtesting with SDID

Thanks!