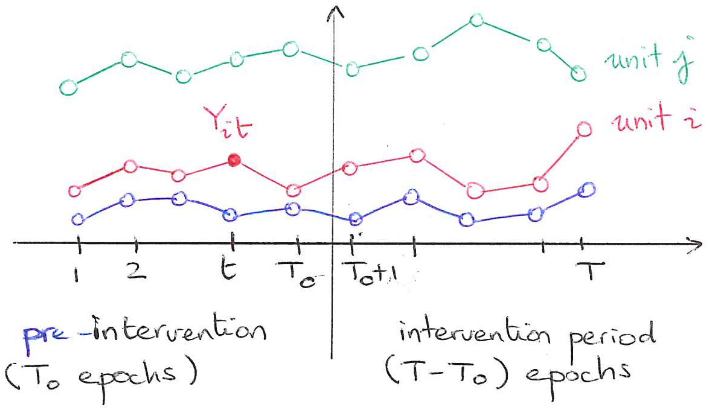
CI = SYNTHETIC CONTROLS

Consider a balanced panel data consisting of n units observed over T time periods. A subset of n_t units receive a treatment at time $T_0+1 \leqslant T$, while the remaining n_c units are never treated. Let W_{it} denote the treatment status of unit i at time t. Assuming a block treatment assignment, $W_{it} = \mathbb{1}(\{ i \geqslant n_c+1 ; t \geqslant T_0+1 \})$

Put $Y_{it}(0)$ = control potential outcome of unit i @ time t

$Y_{it}(1)$ = treatment P.O.

$Y_{it} = W_{it}Y_{it}(1) + (1-W_i)Y_{it}(0)$



We are interested in estimating the average treatment effect on the treated group over the intervention period

$$\begin{cases} ATT(T_0,T) = \dfrac{1}{T-T_0} \displaystyle\sum_{t=T_0+1}^{T} ATT_t \\ ATT_t = \mathbb{E}\left[Y_{it}(1) - Y_{it}(0) \mid W_{it}=1\right] \end{cases}$$

In CI = PANEL DATA METHODS, we derived a consistent estimator of $ATT(T_0,T)$ under the parallel trend (A) and no anticipation (B) assumptions (p 15/16):

$$\hat{\Delta} = \underset{\Delta,\alpha,\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{t=1}^{T} (Y_{it} - \alpha_i - \beta_t - \Delta W_{it})^2$$

$$= \left\{ \frac{1}{n_t(T-T_0)} \sum_{\substack{i \geqslant n_c+1 \\ t \geqslant T_0+1}} Y_{it} - \frac{1}{n_t T_0} \sum_{\substack{i \geqslant n_c+1 \\ t \leqslant T_0}} Y_{it} \right\}$$

$$- \left\{ \frac{1}{n_c(T-T_0)} \sum_{\substack{i \leqslant n_c \\ t \geqslant T_0+1}} Y_{it} - \frac{1}{n_c T_0} \sum_{\substack{i \leqslant n_c \\ t \leqslant T_0}} Y_{it} \right\}$$

= difference-in-differences estimator

$\longrightarrow ATT(T_0,T)$ as $n_t, n_c \to \infty$ under (A) & (B).

Guarantees (i.e. consistency, bias, asymptotic normality...) of estimators of $ATT(T_0,T)$ are commonly derived under the assumption of a data generating process. Many panel data methods assume that

$$Y_{it}(0) = \underset{\uparrow \\ \text{model component}}{l_{it}} + \varepsilon_{it} \leftarrow \text{additive noise}$$

See Chernozhukov et al (2019) for a discussion. The DID approach assumes that each unit i and each time period have a different offset : $l_{it} = \alpha_i + \beta_t$. When the underlying model departs from this simple additive structure, the parallel trend may break down and the DiD estimator of $ATT(T_0,T)$ may be biased.

A natural generalization is $l_{it} = A_i^t B_t$ for $A_i, B_t \in \mathbb{R}^k$ [aka FACTOR MODELS]

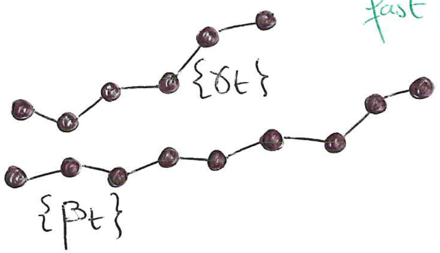
B_t → latent time varying factors

A_i → unknown unit factor loadings

Ex: $A_i = [\alpha_i \;\; 1]^t$

$B_t = [1 \;\; \beta_t]^t$ recovers $l_{it} = \alpha_i + \beta_t$

Ex: $B_t = [1 \;\; \underset{\text{fast rising trend}}{\beta_t} \;\; \underset{\text{slow rising trend}}{\gamma_t}]$



units with $A_i = [\alpha_i \;\; 1 \;\; 0]^t$ pick up the fast trend while units with $A_i = [\alpha_i \;\; 0 \;\; 1]^t$ the slow trend.

The (n x T) matrix of potential outcomes can be modeled as $\underset{(n\times T)}{\underline{\underline{Y}}(0)} = \underset{(n\times T)}{\underline{\underline{L}}} + \underset{(n\times T)}{\underline{\underline{\varepsilon}}}$ $\quad E(\underline{\underline{\varepsilon}} \mid \underline{\underline{L}}) = 0$

for a low rank matrix $\underline{\underline{L}}$. Some authors have proposed direct estimation of $\underline{\underline{L}}$ to recover the ATT, see for example Athey, Bayati, Doudchenko, Imbens, Khosravi (2017) via nuclear norm minimization. Synthetic Control methods address confounding bias without explicitely estimating $\underline{\underline{L}}$ ⇒ indirect approach. SC weight the control units to rebalance the treatment & control groups and create

---

parallel trends. The motivation being that balancing the treatment & control pre-intervention outcomes should also balance the latent factors A_i. The first section introduces the original SC approach and some of its variants. The next two sections detail two popular generalizations: Synthetic DiD and Augmented SC. We conclude with a short discussion on micro vs aggregated data.

I - CANONICAL SC

Recall that $ATT(T_0,T) = \frac{1}{(T-T_0)} \sum_{t=T_0+1}^{T} ATT_t$ with

$$ATT_t = E[Y_{it}(1) \mid \omega_{it}=1] - E[Y_{it}(0) \mid \omega_{it}=1]$$

↓ Estimated using $\overline{Y}_{nt} := \frac{1}{n_t} \sum_{i \in trt} Y_{it}$

↓ Estimated using SC. Notation: $\hat{Y}_{nt}(0)$

The SC estimator of $\widehat{ATT}_t$ is

$$\widehat{ATT}_t = \overline{Y}_{nt} - \hat{Y}_{nt}(0)$$

The problem reduces to a single treated unit [& $n_c \geq 1$ controls]

Without loss of generality, denote by Y_{nt} the time series observations of the treated unit [clear from context if it corresponds to a sample mean or not].

Observations are $\{Y_{it}\}$ for $i=1,\dots,n$, $t=1,\dots,T$

with control units denoted with index $i=1,\dots,n-1$.

Introducing matrix notation :

$$\begin{array}{cc} n_c \updownarrow & \multirow{2}{*}{$\left(\begin{array}{c|cc} \underline{\underline{Y}}_0 & \underline{Y}_{T_0+1} & \underline{Y}_T \\ \hline -\underline{Y}_n^t- & ? & ? \end{array}\right)$} \\ 1 \updownarrow & \\ & \quad\overset{\longleftrightarrow}{T_0} \qquad \overset{\longleftrightarrow}{T-T_0} \end{array}$$

SC performs a vertical regression, where each control unit receives a non-negative weight $\hat{\omega}_i \geqslant 0$ computed such that the weighted sum of the pre-intervention control observations matches the treatment :

$$Y_{nt} \approx \underbrace{\sum_{i=1}^{n_c} \hat{\omega}_i Y_{it}}_{\text{"the Synthetic Control"}} \qquad \forall t=1,\dots,T_0$$

Exact Matching may not be feasible

There are a multitude of criteria to fit the weights .

Abadie (2003, 2010) [simplex regression]

$$\hat{w} = \underset{\substack{w \geqslant 0 \\ w^t \mathbb{1} = 1}}{\operatorname{argmin}} \| \underline{Y}_n - \underline{\underline{Y}}_0^t w \|_2^2 + \lambda \|w\|_2^2$$

$$\hat{Y}_{nt}(0) := \underline{Y}_t^t \hat{w} \qquad t \geqslant T_0+1$$

Remarks (i) Abadie (2003, 2010) do not consider the ridge penalty & a weighted OLS.

(ii) The constraints $w \geqslant 0$ & $w^t \mathbb{1} = 1$ ensures a sparse solution : many control units receive a 0 weight.

A popular alternative to simplex regression is the elastic net, see e.g Doudchenko & Imbens (2017)

Doudchenko & Imbens (2017) [elastic net]

$$\hat{w} = \underset{w}{\operatorname{argmin}} \| \underline{Y}_n - \underline{\underline{Y}}_0^t w \|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2$$

$$\hat{Y}_{nt} := \underline{Y}_t^t \hat{w} \qquad t \geqslant T_0+1$$

May add an intercept

× Guarantees under the latent factor model.

$\forall t$, $\forall i$ $\quad Y_{it}(0) = A_i^t B_t + \varepsilon_{it}$

$$\Rightarrow Y_{nt} - \hat{Y}_{nt}(0) = Y_{nt} - \sum_{i=1}^{n_c} \hat{w}_i Y_{it}, \quad t \geqslant T_0+1$$

$$= \left(A_n - \sum_{i=1}^{n_c} \hat{w}_i A_i\right)^t B_t$$

$$+ \left(\varepsilon_{nt} - \sum_{i=1}^{n_c} \hat{w}_i \varepsilon_{it}\right)$$

Under some conditions, balancing pre-intervention outcomes will balance the factor loadings

Put $X_i = (Y_{i1}, \dots, Y_{iT_0})^t$. Then:

$$\left(A_n - \sum_{i=1}^{n_c} \hat{w}_i A_i\right) = (B^t B)^{-1} B^t \left(X_n - \sum_{i=1}^{n_c} \hat{w}_i X_i\right)$$

$$- (B^t B)^{-1} B^t \left(\varepsilon_{n,1:T_0} - \sum_{i=1}^{n_c} \hat{w}_i \varepsilon_{i,1:T_0}\right)$$

$$B = \begin{pmatrix} - B_1^t - \\ \vdots \\ - B_{T_0}^t - \end{pmatrix} \in \mathbb{R}^{T_0 \times k} \qquad \varepsilon_{i,1:T_0} = (\varepsilon_{i1}, \dots, \varepsilon_{iT_0})^t$$

Under exact matching and some regularity conditions (e.g. B^tB should be invertible), Abadie, Diamond and Hainmuller (2010) provide a bound for $Y_{nt} - \hat{Y}_{nt}(0)$, and show that the bias can be made arbitrarily small as $T_0 \to \infty$. However, exact matching on pre-intervention outcomes is not always feasible & increasing T_0 does not help in reducing the bias.

The SC estimator of ATT(T_0, T) is

$$\widehat{ATT}(T_0,T) = \frac{1}{T-T_0} \sum_{t=T_0+1}^{T} \left(Y_{nt} - \hat{Y}_t(0)\right)$$

$$= \left\{\frac{1}{T-T_0} \sum_{t=T_0+1}^{T} Y_{nt}\right\} \leftarrow \text{trt}$$

$$- \left\{\frac{1}{T-T_0} \sum_{t=T_0+1}^{T} \sum_{i=1}^{n_c} \hat{w}_i Y_{it}\right\} \swarrow \text{ctl}$$

= "weighted" difference estimator

The SC estimator is "just" a difference estimator, with proper weighting of the control units. Assuming $w_i = \frac{1}{n_c}$ $\forall i \in \mathcal{C}$

---

recover the difference estimator [but randomization is required to recover / identify ATT(T_0, T) = ATE].

It turns out that the SC point estimate can be computed as the weighted least square estimate in a one-way fixed effect model:

Theorem.

$$\hat{\Delta} = \underset{\beta, \Delta}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{t=T_0+1}^{T} \hat{w}_i \left(Y_{it} - \beta_t - \Delta W_{it}\right)^2$$

Non-negative, sum to 1 time FE.

$$= \frac{1}{T-T_0} \sum_{t=T_0+1}^{T} Y_{nt} - \frac{1}{T-T_0} \sum_{t=T_0+1}^{T} \sum_{i=1}^{n_c} \hat{w}_i Y_{it}$$

= SC estimator of ATT(T_0,T).

[See Appendix A for a proof]

[$w_i = \frac{1}{n_c}$ represents the diff estimator as the OLS solution of a one-way FE model]

The result above shows that SC estimators omit unit FEs. Generalizations with both unit & time FEs yields Synthetic DID (SDID) type of estimators:

Generalization

$$\hat{\Delta} = \underset{\alpha, \beta, \Delta}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{w}_i \left(Y_{it} - \alpha_i - \beta_t - \Delta W_{it}\right)^2$$

$$= \left\{ \frac{1}{T-T_0} \sum_{t=T_0+1}^{T} Y_{nt} - \frac{1}{T_0} \sum_{t=1}^{T_0} Y_{nt} \right\} \leftarrow \text{trt}$$

$$- \left\{ \frac{1}{T-T_0} \sum_{t=T_0+1}^{T} \sum_{i=1}^{n_c} \hat{\omega}_i Y_{it} - \frac{1}{T_0} \sum_{t=1}^{T_0} \sum_{i=1}^{n_c} \hat{\omega}_i Y_{it} \right\} \leftarrow \text{ctl}$$

= difference-in-differences estimator of ATT(T_0,T) where each control unit receives a weight w_i.

II_ SYNTHETIC DID

Arkhangelsky et al (2021) go one step further and introduce weights λ_t to balance pre-intervention time periods with intervention ones:

$$(\hat{\lambda}_0, \hat{\lambda}) = \underset{\substack{\lambda_0 \in \mathbb{R},\ \lambda \in \Lambda}}{\operatorname{argmin}} \sum_{i=1}^{n_c} \left(\lambda_0 + \sum_{t=1}^{T_0} \lambda_t Y_{it} - \frac{1}{T-T_0} \sum_{t=T_0+1}^{T} Y_{it} \right)^2$$

Horizontal Regression over the control units

$$\text{where } \Lambda = \left\{ \lambda \in \mathbb{R}_+^T \ \middle|\ \sum_{t=1}^{T_0} \lambda_t = 1 \ ,\ \lambda_t = \frac{1}{T-T_0}, t \geqslant T_0+1 \right\}$$

Ensures a sparse solution.

⇒ Weights λ_t complement weights w_i.

Horizontal (HZ) : select in pre-intervention times t similar to the intervention period.

Vertical (VT) : select in the control pool units that are similar to the treated unit.

Both unit & time weights are used in a TWFE regression model to estimate the ATT(T_0,T):

Synthetic DID

$$\hat{\Delta} = \underset{\alpha, \beta, \Delta}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{\omega}_i \hat{\lambda}_t \left(Y_{it} - \alpha_i - \beta_t - \Delta W_{it} \right)^4$$

$$= \left\{ \frac{1}{T-T_0} \sum_{t=T_0+1}^{T} Y_{nt} - \sum_{t=1}^{T_0} \hat{\lambda}_t Y_{nt} \right\} \leftarrow \text{trt}$$

$$- \left\{ \frac{1}{T-T_0} \sum_{t=T_0+1}^{T} \sum_{i=1}^{n_c} \hat{\omega}_i Y_{it} - \sum_{t=1}^{T_0} \sum_{i=1}^{n_c} \hat{\omega}_i \hat{\lambda}_t Y_{it} \right\} \leftarrow \text{ctl}$$

Akhangelsky et al (2021) fit unit weights using a simplex ridge regression + intercept and time weights using a simplex regression with no penalty + intercept.

The SDID estimator of ATT(T_0,T) takes the form

$$\widehat{ATT}(T_0, T) = \frac{1}{T-T_0} \sum_{t=T_0+1}^{T} \widehat{ATT}_t$$

$$\widehat{ATT}_t = Y_{nt} - \hat{Y}_{nt}(0) \quad , \quad t \geqslant T_0+1$$

$$\hat{Y}_{nt}(0) = \langle \hat{\lambda}, \underline{Y}_n \rangle + \langle \hat{\omega}, \underline{Y}_t \rangle - \langle \hat{\omega}, \underline{\underline{Y}}_0 \hat{\lambda} \rangle$$

The authors provide asymptotic guarantees that $\widehat{ATT}(T_0,T)$ identify ATT(T_0,T) under general regularity conditions & a latent factor model. Assuming a large panel with $n_c \to \infty$, $T_0 \to \infty$, $n_t(T-T_0) \to \infty$, they prove that $\widehat{ATT}(T_0, T)$ is asymptotically normally distributed, centered around $ATT(T_0,T)$, with variance $O\left(\frac{1}{n_t(T-T_0)}\right)$

- **Remark** : HZ vs VT regression

SDID makes use of information contained in HZ and VT regressions to estimate the ATT. There are cases, however, where both HZ & VT yield the same point estimates, as noted in Shen, Ding, Sekhon & Yu (2023).

Recall the notation
$$\begin{array}{c|c|c} & T_0 & 1 \\ \hline n_c & \underline{\underline{Y}}_0 & \underline{Y}_T \\ 1 & -\ \underline{Y}_n^t\ - & ? \end{array}$$

$$HZ : \hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \| \underline{Y}_T - \underline{\underline{Y}}_0 \alpha \|_2^2 \qquad \hat{Y}^{HZ}_{nT}(0) = \underline{Y}_n^t \hat{\alpha}$$

$$VT : \hat{\beta} = \underset{\beta}{\operatorname{argmin}} \| \underline{Y}_n - \underline{\underline{Y}}_0^t \beta \|_2^2 \qquad \hat{Y}^{VT}_{nT}(0) = \underline{Y}_T^t \hat{\beta}$$

If columns are linearly independent (rank $\underline{\underline{Y}}_0 = T_0 < n_c$)
then $\hat{\alpha} = (Y_0^t Y_0)^{-1} Y_0^t Y_T$

If raws are linearly independent (rank $Y_0 = n_c < T_0$)
then $\hat{\beta} = (Y_0 Y_0^t)^{-1} Y_0 Y_n$

Assuming more generally that rank $Y_0 = R \leq \min(n_c, T_0)$, and denoting $Y_0^\dagger$ the pseudo inverse of Y_0 [some properties : $(Y_0^\dagger)^t$ is the pseudo inverse of Y_0^t since $Y_0 = U \Sigma V^t = \sum_{\ell=1}^{R} \sigma_\ell u_\ell v_\ell^t$ (SVD decomposition) and

$Y_0^\dagger = \sum_{\ell=1}^{R} \frac{1}{\sigma_\ell} v_\ell u_\ell^t = V \Sigma^{-1} U^t$, with
$U \in \mathbb{R}^{n_c \times R}$, $V \in \mathbb{R}^{T_0 \times R}$, $\Sigma \in \mathbb{R}^{R \times R} = \operatorname{diag}(\sigma_1, \ldots, \sigma_R)$.
Also, $Y_0^\dagger Y_0 Y_0^\dagger = Y_0^\dagger$]

Then $\hat{\alpha} = Y_0^\dagger Y_T \Rightarrow \hat{Y}^{HZ}_{nT}(0) = Y_N Y_0^\dagger Y_T$
$\hat{\beta} = (Y_0^\dagger)^t Y_n \Rightarrow \hat{Y}^{VT}_{nT}(0) = Y_T^t (Y_0^\dagger)^t Y_N$
and we see that $\hat{Y}^{HZ}_{nT}(0) = \hat{Y}^{VT}_{nT}(0)$.

Symmetry in HZ and VT regression under ℓ_2-norm minimalization.

In addition,
$$\begin{aligned} \hat{\beta}^t Y_0 \hat{\alpha} &= Y_N^t Y_0^\dagger Y_0 Y_0^\dagger Y_T \\ &= Y_N^t Y_0^\dagger Y_T \\ &= \hat{Y}^{HZ}_{nT}(0) = \hat{Y}^{VT}_{nT}(0). \end{aligned}$$

So that the LS estimate $< \hat{\alpha}, Y_0^t \hat{\beta} >$ is expressed in terms of both $\hat{\alpha}$ and $\hat{\beta}$.

III_ AUGMENTED SC (ASC)

ASC starts with the canonical SC estimator of the weights & makes we of an outcome model to estimate & correct for the bias due to imperfect pre-intervention fit:

$$\underset{t \geq T_0+1}{\hat{Y}_{nt}(0)} = \underset{\text{outcome model estimate}}{\hat{M}_{nt}(0)} + \sum_{i=1}^{n_c} \underset{\text{SC weight}}{\hat{w}_i} \underbrace{(Y_{it} - \hat{M}_{it}(0))}_{\text{bias from the outcome model}}$$

- ↘ Similar to an AIPW estimator
- ↘ ASC belongs to the class of Doubly Robust estimators.

Ben-Michael et al (2021) discuss in depth the properties of this estimator for a class of linear outcome models fitted using Ridge Regression:

$$\hat{M}_{it}(0) = \sum_{t=1}^{T_0} \hat{\mu}_t Y_{it}$$

$i = 1, \dots, n$

pre-intervention outcomes only

where

$$(\hat{\mu}_0, \hat{\mu}) = \underset{\substack{\mu_0, \mu \\ \in \mathbb{R} \;\; \mathbb{R}^{T_0}}}{\operatorname{argmin}} \sum_{i=1}^{n_c} \Big(\sum_{t=1}^{T_0} \mu_t Y_{it} \underset{(+\mu_0)}{} - \frac{1}{T-T_0} \sum_{t=T_0+1}^{T} Y_{it} \Big)^2 + \lambda \|\mu\|_2^2$$

With or without intercept

Average outcome of ctl unit i in the intervention period

Under a linear outcome model, the ASC estimator of $Y_{nt}(0)$, $t \geqslant T_0+1$, reduces to:

$$\hat{Y}_{nt}(0) = \langle \underline{\hat{\mu}}, \underline{Y}_n \rangle + \langle \underline{\hat{\omega}}, \underline{Y}_t \rangle - \langle \underline{\hat{\omega}}, \underline{\underline{Y}}_0 \underline{\hat{\mu}} \rangle$$

↑ Same expression as the SDID estimator.
The weights ($\hat{\mu}$ or $\hat{\lambda}$) are not fitted the same way (RR vs simplex regression).

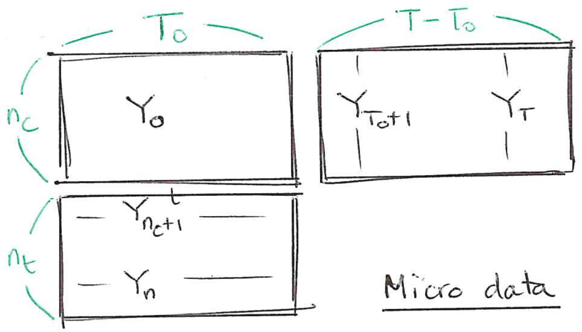
× **Remark**: Setting $\hat{\mu} \leftarrow \frac{1}{T_0} \mathbb{1}$ (resp. $\hat{\lambda}$ for SDID) and $\hat{\omega} \leftarrow \frac{1}{n_c} \mathbb{1}$ yields the DiD estimator.
(see also page 10)

Take Away:

with randomization.

SC is to diff
what ASC/SDID is to diff-in-diff.

IV_ MICRO VS AGGREGATED DATA.



($n_c + n_t = n$)

Suppose that the data is clustered
- ↘ individuals grouped by city / country
- ↘ articles grouped by type (jeans / shoes / ...)

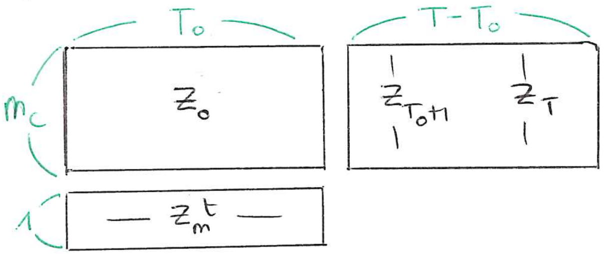
We may consider aggregating units for each cluster before fitting a SC or DID.

Consider a single treatment cluster & m_c control clusters.

Put $z_{jt} = \frac{1}{n_j} \sum_{i \in [n_j]} Y_{it}$ $t = 1, \dots, T$
$j = 1, \dots, m_c + 1$

where $[n_j] = \{ \text{units } i = 1, \dots, n \text{ belonging to cluster } j \}$
$n_j = \#$ units in $[n_j]$.

- Aggregated data.



$(m = m_c + 1)$

We are still interested in estimating a unit-level ATT:

$$ATT_t = \mathbb{E}[Y_{it}(1) - Y_{it}(0) \mid W_{it} = 1]$$

$$\widehat{ATT}_t = z_{mt} - \hat{z}_{mt}(0)$$

where

[ASC/SDID] $\hat{z}_{mt}(0) = \langle \hat{\lambda}, z_m \rangle + \langle \hat{\omega}, z_t \rangle - \langle \hat{\omega}, z_0 \hat{\lambda} \rangle$

$$\widehat{ATT}_t = \left\{ z_{mt} - \sum_{t'=1}^{T_0} \hat{\lambda}_{t'} z_{mt'} \right\} \leftarrow \text{trt group}$$
$$- \left\{ \sum_{j=1}^{m_c} \hat{\omega}_j z_{jt} - \sum_{j=1}^{m_c} \sum_{t'=1}^{T_0} \hat{\omega}_j \hat{\lambda}_{t'} z_{jt'} \right\}$$

↖ control group.

We recover DID with $\hat{\lambda} \leftarrow \frac{1}{T_0} \mathbb{1}$

$\hat{\omega} \leftarrow \left(\frac{n_1}{n_c}, \dots, \frac{n_{m_0}}{n_c} \right)^t$

since

$$\widehat{ATT}_t = \left\{ \overset{z_{mt}\,=}{\boxed{\frac{1}{n_t} \sum_{i=n_c+1}^{n} Y_{it}}} - \frac{1}{T_0} \sum_{t'=1}^{T_0} \overset{z_{mt'}\,=}{\boxed{\frac{1}{n_t} \sum_{i=n_c+1}^{n} Y_{it'}}} \right\}$$
$$- \left\{ \sum_{j=1}^{m_c} \frac{\cancel{n_j}}{n_c} \underset{=\,z_{jt}}{\boxed{\frac{1}{\cancel{n_j}} \sum_{i \in [n_j]} Y_{it}}} - \frac{1}{T_0} \sum_{j=1}^{m_c} \frac{n_j}{n_c} \sum_{t'=1}^{T_0} \underset{=\,z_{jt'}}{\boxed{\frac{1}{n_j} \sum_{i \in [n_j]} Y_{it'}}} \right.$$

$$= \left\{ \frac{1}{n_t} \sum_{i \in trt} Y_{it} - \frac{1}{n_t T_0} \sum_{t'=1}^{T_0} \sum_{i \in trt} Y_{it'} \right\}$$
$$- \left\{ \frac{1}{n_c} \sum_{i \in ctl} Y_{it} - \frac{1}{n_c T_0} \sum_{t'=1}^{T_0} \sum_{i \in ctl} Y_{it} \right\}.$$

& $\frac{1}{T - T_0} \sum_{t=T_0+1}^{T} \widehat{ATT}_t = DID.$

⇒ No loss of generality in aggregating data per cluster.

Appendix A : Proof of the Theorem page 8 .

Model $Y_{it} = \beta_t + \Delta w_{it} + \varepsilon_{it}$ in matrix notation :

$$\underline{\underline{X}} = \underbrace{\left[\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & & 0 & 1 \\ \hline 0 & 1 & & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 1 & & 0 & 1 \\ \hline & & \cdots/\cdots & & \cdots \\ \hline 0 & 0 & & 1 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & & 1 & 1 \end{array}\right]}_{T-T_0} \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} n \\ \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} n \\ \\ \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} n \end{array} \equiv \left[\begin{array}{c|c} X_1 & x \\ \hline X_2 & x \\ \hline \cdots/\cdots & \\ \hline X_{T-T_0} & x \end{array}\right]$$

$\underline{\underline{X}} \in \mathbb{R}^{n(T-T_0) \times (T-T_0+1)}$

where

- $X_t \equiv \left[\begin{array}{ccc} & \overset{t}{\downarrow} & \\ & 1 & \\ 0 & \vdots & 0 \\ & 1 & \end{array}\right] \in \mathbb{R}^{n \times (T-T_0)}$

- $x = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^n$

$\underline{\beta} = (\beta_{T_0+1}, \ldots, \beta_T, \Delta)^t \in \mathbb{R}^{T-T_0+1}$

$\underline{Y} = (\underbrace{Y_{1,T_0+1}, \ldots, Y_{n,T_0+1}}_{n} \mid \ldots \mid \underbrace{Y_{1T}, \ldots, Y_{nT}}_{n})^t \in \mathbb{R}^{n(T-T_0)}$

$$\boxed{\underline{Y} = \underline{\underline{X}}\,\underline{\beta} + \underline{\varepsilon}}$$

$\uparrow$ defined similarly

• Weight Matrix

$$\underline{\underline{W}} = \left[\begin{array}{cc|cc} \underline{\underline{Z}} & 0 & & \\ \hline 0 & \underline{\underline{Z}} & & \\ & & \ddots & \\ \hline & & & 0 \\ & & 0 & \underline{\underline{Z}} \end{array}\right] \in \mathbb{R}^{n(T-T_0) \times n(T-T_0)}$$

$(T-T_0)$ matrices

$$\underline{\underline{Z}} = \begin{pmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{pmatrix} \in \mathbb{R}^{n \times n}$$

with $\sum_{i=1}^{n-1} w_i = 1$; $w_n = 1$.

Then :

- $$\underline{\underline{X}}^t \underline{\underline{W}}\,\underline{\underline{X}} = \left(\begin{array}{c|c} 2\, I_{(T-T_0)\times(T-T_0)} & \mathbb{1}_{T-T_0} \\ \hline \mathbb{1}^t_{T-T_0} & T-T_0 \end{array}\right)$$

- $$(\underline{\underline{X}}^t \underline{\underline{W}}\,\underline{\underline{X}})^{-1} = \frac{1}{T-T_0}\left(\begin{array}{c|c} \frac{1}{2}[(T-T_0) I_{T-T_0} + J_{T-T_0} & -\mathbb{1}_{T-T_0} \\ \hline -\mathbb{1}^t_{T-T_0} & 2 \end{array}\right)$$

- $$\underline{\underline{X}}^t \underline{\underline{W}}\,\underline{Y} = \left(\sum_{i=1}^n w_i Y_{i,T_0+1}, \ldots, \sum_{i=1}^n w_i Y_{iT}, \sum_{t=T_0+1}^T w_n Y_{nt}\right)^t$$

The result follows noticing that $\hat{\Delta}$ is the last element in the vector $(\underline{\underline{X}}^t \underline{\underline{W}}\,\underline{\underline{X}})^{-1} \underline{\underline{X}}^t \underline{\underline{W}}\,\underline{\underline{Y}}$.