# Section: Module 10

## Difference-in-Difference / Synthetic Control Method

Kentaro Nakamura

GOV 2002

December 5th, 2025

# Logistics

- Important Dates
  - Problem Set 10: Due December 8th
  - Review Session: December 8th (CGIS K354)
  - Final Exam: December 11th

- Today's agenda
  - Difference-in-Difference (DiD)
  - Synthetic Control Methods (SCM)

# Overview: Panel Data / TSCS Data

- Two identification regimes for panel data analysis
- **Strict Exogeneity**: $\{Y_{it}(1), Y_{it}(0)\} \perp\!\!\!\perp D_{it'} \mid \boldsymbol{X}_i^{1:T}, \alpha_i, \boldsymbol{f}^{1:T}$
  - $\alpha_i$ is unit fixed effects, $\boldsymbol{f}^{1:T}$ is time fixed effects, and $\boldsymbol{X}_i^{1:T}$ are the entire histories of covariates
  - Implication
    - No time-varying confounder exists
    - No direct effects from past outcomes to the current outcomes
    - No feedback from past outcomes to current and future treatment status
    - No carryover effects from current treatments to future treatments (i.e., no arrows from $D_{t-1}$ to $Y_t$ or $Y_{t+1}$)
  - Strict exogeneity implies parallel trend assumption

$$\mathbb{E}[Y_{it}(0) - Y_{is}(0) \mid \boldsymbol{X}_i^{1:T}] = \mathbb{E}[Y_{jt}(0) - Y_{js}(0) \mid \boldsymbol{X}_i^{1:T}]$$

- **Sequential ignorability**: $\{Y_{it}(1), Y_{it}(0)\} \perp\!\!\!\perp D_{it'} \mid \boldsymbol{X}_i^{1:T}, \boldsymbol{Y}_i^{1:(t-1)}$
  - This regime allows the inclusion of time-varying confounder, but they must be **observed**
  - For identification and estimation under sequential ignorability, see previous module's slide (e.g., Marginal Structural Models)

# Overview: Panel Data / TSCS Data

FIGURE 2. DAG FOR DGPS UNDER STRICT EXOGENEITY

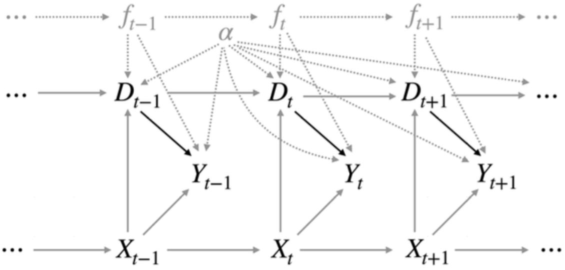
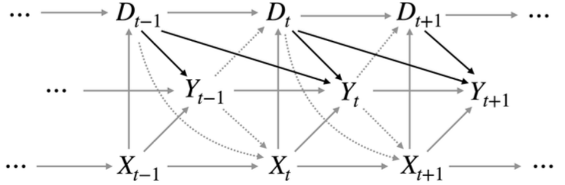



FIGURE 3. DAG FOR DGPS UNDER SEQUENTIAL IGNORABILITY



# DiD for Two Time Periods (1)

- Notation:
  - $G_i$: treatment indicator ($G_i = 1$ for treatment group)
  - $D_{it} = tG_i$: treatment assignment indicator
  - $Y_{it}$: observed outcome for unit $i$ at time $t$
  - $Y_{it}(d)$: potential outcome for unit $i$ at time $t$
- Estimand: Average treatment effect for the treated (ATT)

$$\tau = \mathbb{E}[Y_{i1}(1) - Y_{i1}(0) \mid G_i = 1]$$

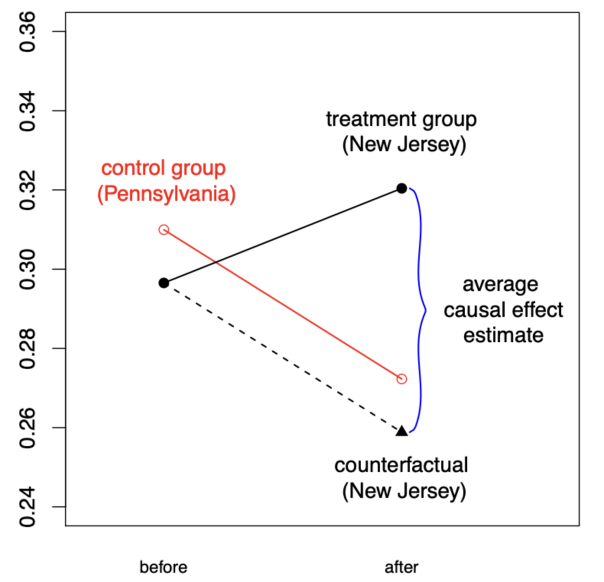
- Assumption: **Parallel trend**

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 1] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 0]$$

  - We also assume **no anticipation** assumption

$$Y_{i0}(1) = Y_{i0}(0)$$

# DiD for Two Time Periods (2)



# DiD for Two Time Periods (3): Identification

Under the parallel trends assumption,

$$\tau = \Big\{\mathbb{E}[Y_{i1} \mid G_i = 1] - \mathbb{E}[Y_{i1} \mid G_i = 0]\Big\} - \Big\{\mathbb{E}[Y_{i0} \mid G_i = 1] - \mathbb{E}[Y_{i0} \mid G_i = 0]\Big\}$$

**Proof:**

$$\begin{aligned}
&\{\mathbb{E}[Y_{i1} \mid G_i = 1] - \mathbb{E}[Y_{i1} \mid G_i = 0]\} - \{\mathbb{E}[Y_{i0} \mid G_i = 1] - \mathbb{E}[Y_{i0} \mid G_i = 0]\} \\
&= \{\mathbb{E}[Y_{i1}(1) \mid G_i = 1] - \mathbb{E}[Y_{i1}(0) \mid G_i = 0]\} \\
&\qquad - \{\mathbb{E}[Y_{i0}(0) \mid G_i = 1] - \mathbb{E}[Y_{i0}(0) \mid G_i = 0]\} \\
&= \underbrace{\mathbb{E}[Y_{i1}(1) \mid G_i = 1] - \mathbb{E}[Y_{i1}(0) \mid G_i = 1]}_{=\,\tau_{ATT}} + \mathbb{E}[Y_{i1}(0) \mid G_i = 1] \\
&\qquad - \mathbb{E}[Y_{i1}(0) \mid G_i = 0] - \mathbb{E}[Y_{i0}(0) \mid G_i = 1] + \mathbb{E}[Y_{i0}(0) \mid G_i = 0] \\
&= \tau_{ATT} + \underbrace{\Big(\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 1] - \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 0]\Big)}_{=0 \text{ under parallel trends}} \\
&= \tau_{ATT}.
\end{aligned}$$

# Two-way Fixed Effects for Two Time Periods

- Consider the two-way fixed effects model

$$Y_{it} = \alpha_i + \beta_t + \tau D_{it} + \epsilon_{it}$$

- Now, since $D_{it} = tG_i$,

$$Y_{i1} - Y_{i0} = (\beta_1 - \beta_0) + \tau G_i + \epsilon_{i1} - \epsilon_{i0}$$

  which means that there are two ways to estimate $\tau$
  - Running two-way fixed effects using the entire data
  - Taking the first difference and run regression
- Since $G_i$ is binary, $\tau$ is written as

$$\tau = \mathbb{E}[Y_{i1} - Y_{i0} \mid G_i = 1] - \mathbb{E}[Y_{i1} - Y_{i0} \mid G_i = 0]$$

  which corresponds to DiD estimator![1]
- **Takeaway**: Two-way fixed effects and DiD are equivalent if there are only two time periods

[1] The parallel trend assumption implies $\mathbb{E}[\epsilon_{i1} - \epsilon_{i0} \mid G_i] = \mathbb{E}[\epsilon_{i1} - \epsilon_{i0}] = 0$

# Covariate in DiD for Two Time Periods

- You can use time-invariant covariates $\boldsymbol{X}_i$ to make **conditional parallel trend assumption**

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 1, \boldsymbol{X}_i] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 0, \boldsymbol{X}_i]$$

- Estimation strategy
  - **Outcome regression**
    - 2 regressions with $\mathbb{E}[Y_{i1} - Y_{i0} \mid G_i = g, \boldsymbol{X}_i]$ for $g \in \{0, 1\}$
    - 4 regressions with $\mathbb{E}[Y_{it} \mid G_i = g, \boldsymbol{X}_i]$ for $g \in \{0, 1\}$ and $t \in \{0, 1\}$
  - **Propensity score weighting** (Abadie 2005)

$$\mathbb{E}\left[\frac{Y_{i1} - Y_{i0}}{\mathbb{P}(G_i = 1)} \cdot \frac{G_i - \pi(\boldsymbol{X}_i)}{1 - \pi(\boldsymbol{X}_i)}\right]$$

    where $\pi(\boldsymbol{X}_i) = \mathbb{P}(G_i = 1 \mid \boldsymbol{X}_i)$ is propensity score
  - **Doubly robust estimation** (Callaway and Sant'Anna 2021)

$$\mathbb{E}\left[\frac{Y_{i1} - Y_{i0} - m(\boldsymbol{X}_i)}{\mathbb{P}(G_i = 1)} \cdot \frac{G_i - \pi(\boldsymbol{X}_i)}{1 - \pi(\boldsymbol{X}_i)}\right]$$

    where $m(\boldsymbol{X}_i) = \mathbb{E}[Y_{i1} - Y_{i0} \mid G_i = 0, \boldsymbol{X}_i]$.

# Proof: Propensity score weighting (1)

- Let's prove the identification of weighting estimator.
- Now,

$$
\begin{aligned}
&\mathbb{E}\left[\frac{Y_{i1}-Y_{i0}}{\mathbb{P}(G_i=1)}\cdot\frac{G_i-\pi(\boldsymbol{X}_i)}{1-\pi(\boldsymbol{X}_i)}\right] \\
&=\mathbb{E}\left\{\mathbb{E}\left[\frac{Y_{i1}-Y_{i0}}{\mathbb{P}(G_i=1)}\cdot\frac{G_i-\pi(\boldsymbol{X}_i)}{1-\pi(\boldsymbol{X}_i)}\mid \boldsymbol{X}_i\right]\right\} \\
&=\mathbb{E}\left\{\frac{\mathbb{E}[(Y_{i1}-Y_{i0})G_i\mid \boldsymbol{X}_i]-\pi(\boldsymbol{X}_i)\mathbb{E}[Y_{i1}-Y_{i0}\mid \boldsymbol{X}_i]}{\mathbb{P}(G_i=1)(1-\pi(\boldsymbol{X}_i))}\right\}
\end{aligned}
$$

- Then, notice that by the law of total probability,

$$
\mathbb{E}[(Y_{i1}-Y_{i0})G_i\mid \boldsymbol{X}_i]=\mathbb{E}[Y_{i1}-Y_{i0}\mid G_i=1,\boldsymbol{X}_i]\pi(X_i)
$$

and

$$
\begin{aligned}
&\mathbb{E}[Y_{i1}-Y_{i0}\mid \boldsymbol{X}_i] \\
&=\mathbb{E}[Y_{i1}-Y_{i0}\mid \boldsymbol{X}_i, G_i=1]\pi(\boldsymbol{X}_i) \\
&\qquad+\mathbb{E}[Y_{i1}-Y_{i0}\mid \boldsymbol{X}_i, G_i=0](1-\pi(\boldsymbol{X}_i))
\end{aligned}
$$

# Proof: Propensity score weighting (2)

- Let $m_g(\boldsymbol{X}_i) = \mathbb{E}[Y_{i1} - Y_{i0} \mid \boldsymbol{X}_i, G_i = g]$. Then,

$$
\begin{aligned}
&\mathbb{E}\left[\frac{Y_{i1} - Y_{i0}}{\mathbb{P}(G_i = 1)} \cdot \frac{G_i - \pi(\boldsymbol{X}_i)}{1 - \pi(\boldsymbol{X}_i)}\right] \\
&= \mathbb{E}\left\{\frac{m_1(\boldsymbol{X}_i)\pi(\boldsymbol{X}_i) - m_1(\boldsymbol{X}_i)\pi(X_i)^2 - m_0(\boldsymbol{X}_i)(1 - \pi(\boldsymbol{X}_i))\pi(\boldsymbol{X}_i)}{\mathbb{P}(G_i = 1)(1 - \pi(\boldsymbol{X}_i))}\right\} \\
&= \mathbb{E}\left\{\frac{(1 - \pi(\boldsymbol{X}_i))\pi(\boldsymbol{X}_i)\{m_1(\boldsymbol{X}_i) - m_0(\boldsymbol{X}_i)\}}{\mathbb{P}(G_i = 1)(1 - \pi(\boldsymbol{X}_i))}\right\} \\
&= \mathbb{E}\left\{\frac{\pi(\boldsymbol{X}_i)\{m_1(\boldsymbol{X}_i) - m_0(\boldsymbol{X}_i)\}}{\mathbb{P}(G_i = 1)}\right\}
\end{aligned}
$$

- Under conditional parallel trend assumption,

$$
m_1(\boldsymbol{X}_i) - m_0(\boldsymbol{X}_i) = \mathbb{E}[Y_{i1}(1) - Y_{i1}(0) \mid G_i = 1, \boldsymbol{X}_i] := \tau(\boldsymbol{X}_i)
$$

# Proof: Propensity score weighting (3)

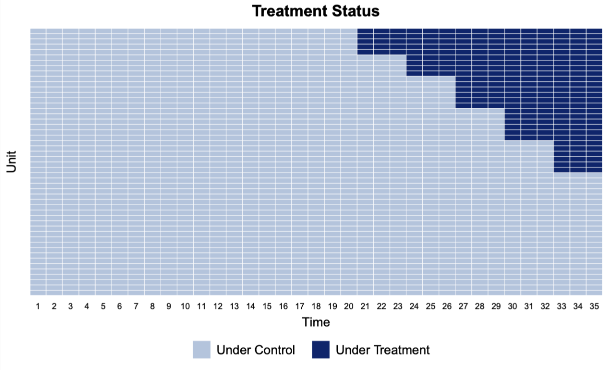
- Notice that $\pi(\boldsymbol{X}_i) = \mathbb{E}[T_i \mid \boldsymbol{X}_i]$

$$
\begin{aligned}
\mathbb{E}\left[\frac{Y_{i1} - Y_{i0}}{\mathbb{P}(G_i = 1)} \cdot \frac{G_i - \pi(\boldsymbol{X}_i)}{1 - \pi(\boldsymbol{X}_i)}\right] &= \mathbb{E}\left\{\frac{\mathbb{E}[G_i \mid \boldsymbol{X}_i]\tau(\boldsymbol{X}_i)}{\mathbb{P}(G_i = 1)}\right\} \\
&= \mathbb{E}\left\{\frac{\mathbb{E}[G_i \tau(\boldsymbol{X}_i) \mid \boldsymbol{X}_i]}{\mathbb{P}(G_i = 1)}\right\} \\
&= \mathbb{E}\left\{\frac{G_i \tau(\boldsymbol{X}_i)}{\mathbb{P}(G_i = 1)}\right\} \quad (\because \text{L.I.E.}) \\
&= \mathbb{E}\left\{\frac{1 \cdot \tau(\boldsymbol{X}_i)}{\mathbb{P}(G_i = 1)} \mid G_i = 1\right\}\mathbb{P}(G_i = 1) \\
&= \mathbb{E}[\tau(\boldsymbol{X}_i) \mid G_i = 1] = \mathrm{ATT}.
\end{aligned}
$$

- We can do the similar proof for the doubly robust estimator.

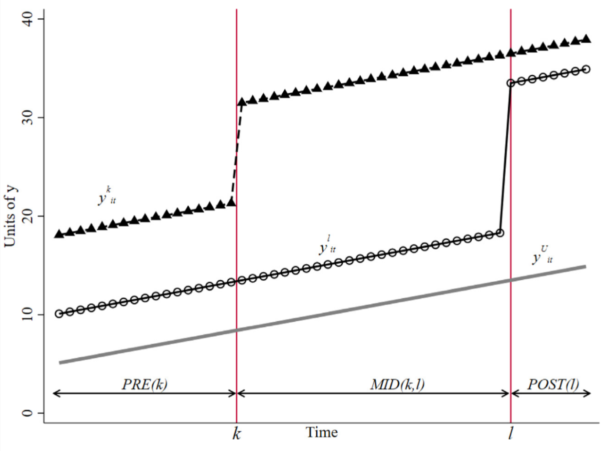
# Staggered Adoption

- In some applications, however, different units receive treatment at the different timing
- This setting is called **staggered adoption**
  - Importantly, staggered adoption still assumes no switchback in treatment status



# Two-way fixed effects under staggered adoption (1)

- Under staggered adoption, two-way fixed effects has a lot of troubles.
  - For simplicity, let's consider the following case with two different treatment timing



# Two-way fixed effects under staggered adoption (2)

- Under staggered adoption, two-way fixed effects model gives you a **weighted average** of all possible 2x2 DiD estimator
- Specifically, let

1. Early vs Never $\beta_{kU} = (y_k^{\text{Post(k)}} - y_k^{\text{Pre(k)}}) - (y_U^{\text{Post(U)}} - y_U^{\text{Pre(U)}})$
2. Late vs Never $\beta_{lU} = (y_l^{\text{Post(l)}} - y_k^{\text{Pre(l)}}) - (y_U^{\text{Post(U)}} - y_U^{\text{Pre(U)}})$
3. Early vs Late $\beta_{kl}^k = (y_k^{\text{Mid}} - y_k^{\text{Pre(k)}}) - (y_l^{\text{Mid}} - y_l^{\text{Pre(l)}})$
4. Early vs Late $\beta_{kl}^l = (y_l^{\text{Post(l)}} - y_l^{\text{Mid}}) - (y_k^{\text{Post(l)}} - y_k^{\text{Mid}})$

- Then, Theorem 1 of Goodman-Bacon (2021) shows that $\beta$ in TWFE model represents

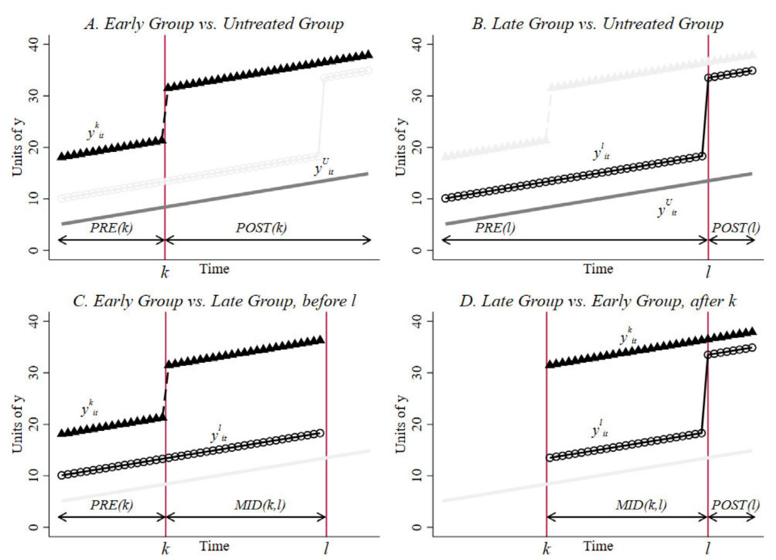
$$\beta = w_{kU}\beta_{kU} + w_{lU}\beta_{lU} + w_{kl}^k\beta_{kl}^k + w_{kl}^l\beta_{kl}^l$$

where the weight sums up to 1 ($w_{kU} + w_{lU} + w_{kl}^k + w_{kl}^l = 1$) and

$$w_g \propto (\text{Share of Group } g)^2 \times \text{Variance of } D_{it} \text{ in that group}$$

# Two-way fixed effects under staggered adoption (3)

- Four possible comparisons



# Two-way fixed effects under staggered adoption (4)

- However, you see that $\beta^l_{kl}$ does not give you treatment effect under parallel trend
  - Comparison group is already treated (forbidden comparison!)
- Specifically, $\beta^l_{kl}$ is written as

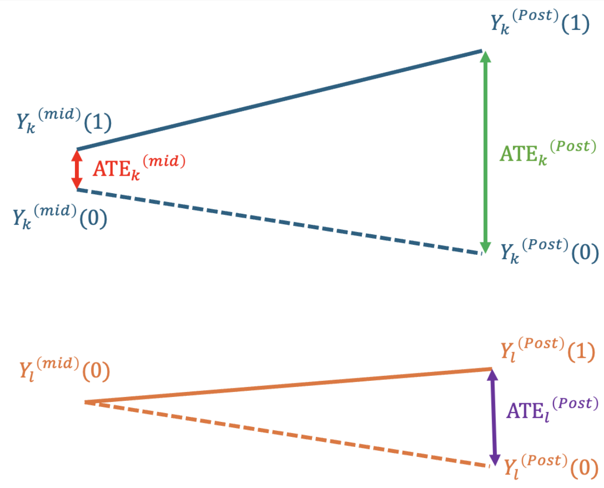
$$\begin{aligned}\beta^l_{kl} &= (y_l^{\text{Post(l)}} - y_l^{\text{Mid}}) - (y_k^{\text{Post(l)}} - y_k^{\text{Mid}}) \\ &= (y_l^{\text{Post(l)}} - y_l^{\text{Mid}}) \\ &\quad - \left( (Y_k^{\text{Post}(l)}(0) + \mathrm{ATT}_k^{\text{Post}}) - (Y_k^{\text{Mid}}(0) - \mathrm{ATT}_k^{\text{Mid}}) \right) \\ &= \mathrm{ATT}_l^{\text{Post}} - (\mathrm{ATT}_k^{\text{Post}} - \mathrm{ATT}_k^{\text{Mid}})\end{aligned}$$

  so it can be negative if $\mathrm{ATT}_k^{\text{Post}} > \mathrm{ATT}_k^{\text{Mid}}$
  - This means that it is possible that even if all ATT is positive, the coefficient can be negative

- **Takeaway**: In staggered adoption setting, TWFE does not recover the interpretable causal parameter in general
  - This is because TWFE uses already-treated units as controls

# Two-way fixed effects under staggered adoption (5)

- Graphical illustration of $\beta_{kl}^{l}$



# Extra: Negative weight problem

- The problem of TWFE can be seen as **negative weight problem** (de Chaisemartin and D'Haultfoeuille 2020)
- Let $\tau_{it}$ be the treatment effect for unit $i$ at time $t$. de Chaisemartin and D'Haultfoeuille (2020) showed that

$$\hat{\beta}_{\mathrm{TWFE}} \xrightarrow{p} \sum_{i,t:\ D_{it}=1} \omega_{it}\tau_{it}, \quad \omega_{it} = \frac{\hat{\epsilon}_{it}}{\sum_{i,t:\ D_{it}=1} \hat{\epsilon}_{it}}$$

  where $\hat{\epsilon}_{it}$ is the residuals from running $D_{it}$ on the fixed effects

  - The proof uses FWL theorem

- Importantly, later period can have smaller, and even **negative weights**
  - As a result, even if all $\tau_{it}$ are positive, $\hat{\beta}_{\mathrm{TWFE}}$ can be negative (the same conclusion from Goodman-Bacon decomposition)
  - This is known as the **negative weight problem**
  - Sun and Abraham (2021) shows that this happens for event-study design (i.e., TWFE including lead and lag variable) as well

# Callaway and Sant'Anna Estimator

- Goodman-Bacon pointed out that the problem is due to the use of already-treated units as control
  - But in staggered adoption, we can estimate the effect of each period's ATT under parallel trend assumption
- Callaway and Sant'Anna (2021) proposed the following strategy:
  - First, estimate **the group-specific ATT** using either (i) never-treated or (ii) not-yet-treated group

    $$\mathrm{ATT}(g,t) = \mathbb{E}[Y_t(1) - Y_t(0) \mid G = g]$$

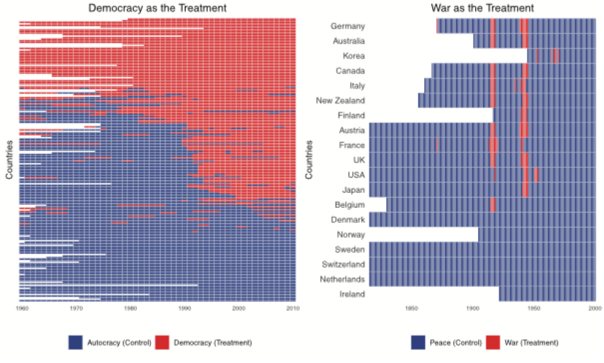
    where $G = g$ refers to the timing of treatment reception (group) and $T = t$ is the time that the treatment effect is measured
    - Pre-treatmennt covariates can be included for covariate-specific parallel trend in doubly-robust way
  - After estimating the group-specific ATTs, aggregate them according to the question, including

$$\text{Dynamic Effect}: \quad \mathrm{ATT}(k) = \text{average of } \mathrm{ATT}(g, g+k)$$

$$\text{Overall treatment effect}: \quad \text{ATT} = \sum_{g=1}^{G} \sum_{t=2}^{T} \omega_{g,t} \mathrm{ATT}(g,t)$$

# Panel Match (1)

- In many cases, there is a treatment reversal.
  - Recall that staggered adoption assumes no treatment reversal




- Panel match overcomes this limitation by combining DiD with matching
  - Assuming conditional parallel trend assumption

# Panel Match (2)

- Estimand: Average Treatment Effect of Policy Change for the Treated

$$\mathbb{E}\Big\{Y_{i,t+F}\Big(X_{it}=1, X_{i,t-1}=0, \underbrace{\{X_{i,t-l}\}_{l=2}^{L}}_{\text{Treatment History}}\Big) - Y_{i,t+F}\Big(X_{it}=0, X_{i,t-1}=0, \underbrace{\{X_{i,t-l}\}_{l=2}^{L}}_{\text{Treatment History}}\Big) \mid \underbrace{X_{it}=1, X_{i,t-1}=0}_{\text{for treated unit}}\Big\}$$

- Procedure
  1. Find the treated observations with $X_{i,t-1}=0$ and $X_{it}=1$
  2. For each treated observation, form a matched set $\mathcal{M}_{it}$ of control observations with identical treatment history from $t-1$ to $t-L$
  3. Refine $\mathcal{M}_{it}$ via matching / weighting to adjust time-invariant / time-varying confounders
  4. Compute DiD estimator

$$\frac{1}{\sum_{i=1}^{N}\sum_{t=L+1}^{T-F} D_{it}} \sum_{i=1}^{N}\sum_{t=L+1}^{T-F} D_{it}\Big\{\Big(Y_{i,t+F}-Y_{i,t-1}\Big) - \sum_{i'\in\mathcal{M}_{it}} w_{it}^{i'}\Big(Y_{i',t+F}-Y_{i',t-1}\Big)\Big\}$$

# Synthetic Control Methods (SCM)

- **Setting**: Suppose that we have $N$ units and $T$ time periods, and there is only one treated unit ($i = N$), which receives the treatment at time $T$.
- **Estimand**:

$$Y_{NT}(1) - Y_{NT}(0) = Y_{NT} - Y_{NT}(0)$$

and thus the goal is to impute $Y_{NT}(0)$.

- **Synthetic Control Method** proposed to use the weighted average

$$\widehat{Y_{NT}(0)} = \sum_{i=1}^{N-1} \hat{w}_i Y_{iT}$$

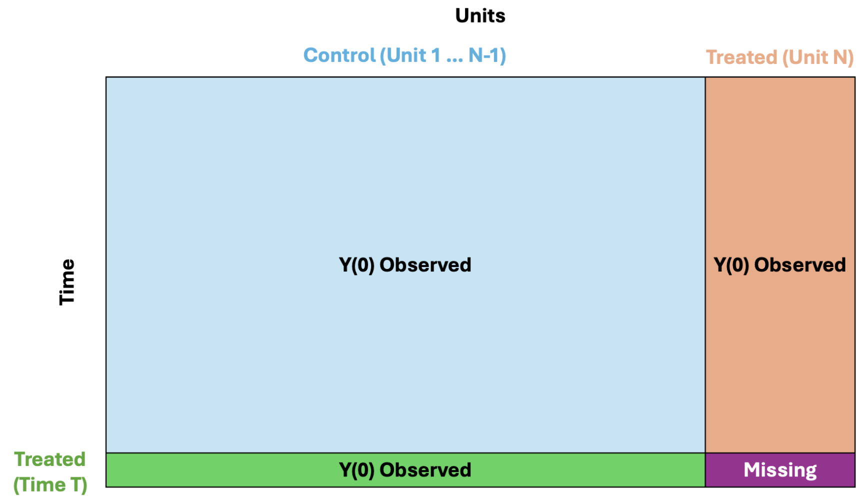
$$\text{where} \quad \hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \sum_{i=1}^{T-1} \left( Y_{Nt} - \sum_{i=1}^{N-1} w_i Y_{it} \right)^2$$

$$\sum_{i=1}^{N-1} \hat{w}_i = 1, \quad \hat{w}_i \geq 0$$

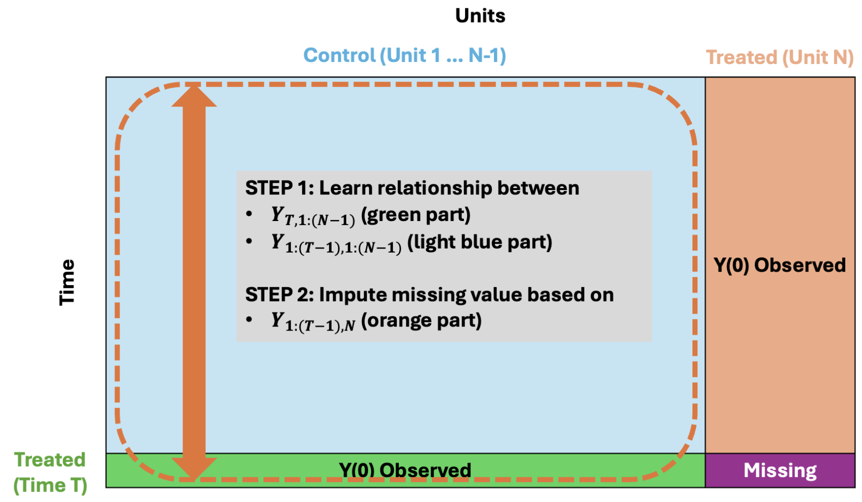
  - This can be seen as a regression problem with some constraints

# Vertical Regression and Horizontal Regression

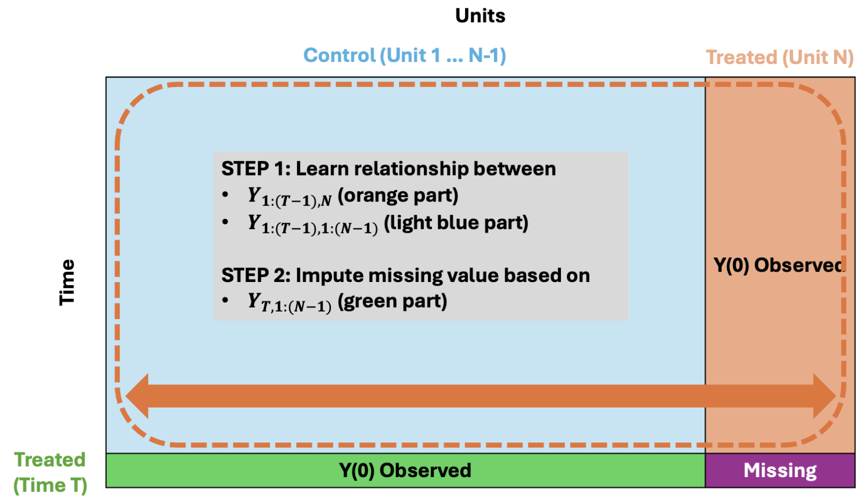
- **Intuition of SCM**: Impute the missing values $\widehat{Y_{NT}(0)}$



# Vertical Regression



# Horizontal Regression




- Note that vertical regression and horizontal regression is actually algebraic equivalent (Shen et al. Econometrica)
  - The only difference between horizontal / vertical regression and SCM is that weight can be negative for vertical / horizontal

# Model-based justification of SCM

- **Factor Analytic Model**:

$$Y_{it}(0) = \gamma_t + \underbrace{\delta_t^T \boldsymbol{X}_i + \xi_t^T \boldsymbol{U}_i}_{\text{Time-varying effects}} + \epsilon_{it}$$

  - The assumption is that there exists weights such that

$$\sum_{i=1}^{N-1} w_i \boldsymbol{X}_i = \boldsymbol{X}_N \quad \text{and} \quad \sum_{i=1}^{N-1} w_i \boldsymbol{U}_i = \boldsymbol{U}_{\boldsymbol{N}}$$

    - Not compatible with time-varying covariates

- **Auto-Regressive Model with time-varying covariates**

$$\begin{aligned} Y_{it}(0) &= \rho_t Y_{i,t-1}(0) + \delta_t^T \boldsymbol{X}_{it} + \epsilon_{it} \\ \boldsymbol{X}_{it} &= \lambda_{t-1} Y_{i,t-1}(0) + \Delta_{t-1} \boldsymbol{X}_{i,t-1} + v_{it} \end{aligned}$$

  - It allows that the past outcomes affect the current treatment
  - It assumes no unobserved time-invariant confounders

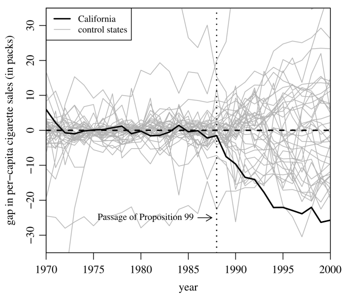
- **Takeaway**: As long as either is correct, SCM gives you the treatment effect

# Permutation Tests for SCM (1)

- Statistical inference for SCM is tricky since there is one single treated unit
  - Recall that SCM only tries to impute $\widehat{Y_{NT}(0)}$. We observe $Y_{NT}(1)$.
  - If we assume $Y_{NT}(1)$ as fixed (design-based perspective / finite-population), then the uncertainty coming from $\widehat{Y_{NT}(0)}$.

- We can see how unusual the treatment unit is (permutation test)
  - **Assumption**: Exchangeability of treated units and control units in the absence of treatment
  - Under this assumption, you can apply SCM to each control unit $\widehat{Y_{iT}(0)}$ for $i = 1, \cdots, N-1$

# Permutation Tests for SCM (2)




- Gray line: imputed outcome difference for $Y_{iT} - \widehat{Y_{iT}(0)}$ for $i = 1, \cdots, N-1$
  - As the imputed outcome difference is significantly different for treated unit, this implies that the effect is statistically significant
- Note that variance of $\widehat{Y_{iT}(0)}$ is big after treated periods in this figure
  - This is because of over-fitting, which is caused by the fact that SCM has too many parameters
  - Therefore, you need to use SCM with regularization to mitigate the overfitting

# Augmented Synthetic Control

- **Goal**: Adjust for possible bias due to poor pre-treatment fit
- Assume the data generating process $Y_{it}(0) = \mu_{it} + \epsilon_{it}$, which encompasses
  - Interactive factor model: $\mu_{it} = \gamma_t + \boldsymbol{\delta}_t^T \boldsymbol{X}_i + \xi_t \boldsymbol{U}_i$
  - Autoregressive model: $\mu_{it} = \rho_t Y_{i,t-1}(0) + \boldsymbol{\delta}_t^T \boldsymbol{X}_{it}$
- Then, impute the counterfactual outcome by

$$\widehat{Y_{it}(0)} = \underbrace{\sum_{i=1}^{N-1} w_i Y_{iT}}_{\text{SCM}} + \underbrace{\left(\hat{\mu}_{NT} - \sum_{i=1}^{N-1} w_i \hat{\mu}_{iT}\right)}_{\text{Imbalance in } \mu_{\cdot T}}$$

$$= \hat{\mu}_{NT} + \underbrace{\sum_{i=1}^{N-1} w_i \left(Y_{iT} - \hat{\mu}_{iT}\right)}_{\text{Residual Balancing}}$$

- Same idea as bias-correction in matching
  - See section 8's slide 12