Synthetic Control Method

ECON 2400
Applied Econometrics II

Andres Mena

Spring 2024

brown-logo.png

Table of Contents

Content

Comparative Case Study

- Originally proposed in **?** and **?**

Comparative Case Study

- Originally proposed in **?** and **?**
- Studies an aggregate intervention on a large single unit (such as state, region, country)

Comparative Case Study

- Originally proposed in **?** and **?**
- Studies an aggregate intervention on a large single unit (such as state, region, country)
- The aim is to estimate the effect of the intervention on some aggregate outcome of interest.

Comparative Case Study

- Originally proposed in **?** and **?**
- Studies an aggregate intervention on a large single unit (such as state, region, country)
- The aim is to estimate the effect of the intervention on some aggregate outcome of interest.
- Comparative Case Studies: outcome of treated unit vs a group of units that are similar to the exposed.

Comparative Case Study

- Originally proposed in **?** and **?**
- Studies an aggregate intervention on a large single unit (such as state, region, country)
- The aim is to estimate the effect of the intervention on some aggregate outcome of interest.
- Comparative Case Studies: outcome of treated unit vs a group of units that are similar to the exposed.
  - **?** compares the evolution of native unemployment in Miami at the time of the boatlift to the average evolution of native unemployment in four other cities in the United States.

Comparative Case Study

- Originally proposed in **?** and **?**
- Studies an aggregate intervention on a large single unit (such as state, region, country)
- The aim is to estimate the effect of the intervention on some aggregate outcome of interest.
- Comparative Case Studies: outcome of treated unit vs a group of units that are similar to the exposed.
    - **?** compares the evolution of native unemployment in Miami at the time of the boatlift to the average evolution of native unemployment in four other cities in the United States.
    - **?** use Pennsylvania as a comparison to estimate the effects of an increase in the New Jersey minimum wage on employment in fast food restaurants in New Jersey

Comparative Case Study

- Originally proposed in **?** and **?**
- Studies an aggregate intervention on a large single unit (such as state, region, country)
- The aim is to estimate the effect of the intervention on some aggregate outcome of interest.
- Comparative Case Studies: outcome of treated unit vs a group of units that are similar to the exposed.
  - **?** compares the evolution of native unemployment in Miami at the time of the boatlift to the average evolution of native unemployment in four other cities in the United States.
  - **?** use Pennsylvania as a comparison to estimate the effects of an increase in the New Jersey minimum wage on employment in fast food restaurants in New Jersey
- The synthetic control methodology formalizes the selection of the comparison units using a data driven procedure.

Example: German Reunification

? estimates the effect of the 1990 German reunification on per capita GDP in West Germany.

Figures/germany.jpeg

Example: German Reunification

? estimates the effect of the 1990 German reunification on per capita GDP in West Germany.

Figures/germany.jpeg

Setting

Consider a complete panel with non-staggered binary treatment assignment.
WLOG, assume the first unit is treated only in the last period.

	t=1	t=2	...	t=T_0	t=T
j=1					
j=2					
...					
j=J					
j=$J+1$					

- Donor Pool: $j = 2, ..., J+1$ units not affected by the intervention.

Setting

Consider a complete panel with non-staggered binary treatment assignment.
WLOG, assume the first unit is treated only in the last period.

	t=1	t=2	...	t=T_0	t=T
j=1					
j=2					
...					
j=J					
j=$J+1$					

- Donor Pool: $j = 2, ..., J + 1$ units not affected by the intervention.
- T_0 periods before the intervention.

Setting

Consider a complete panel with non-staggered binary treatment assignment.
WLOG, assume the first unit is treated only in the last period.

	t=1	t=2	...	t=T_0	t=T
j=1					
j=2					
...					
j=J					
j=$J+1$					

- Donor Pool: $j = 2, ..., J + 1$ units not affected by the intervention.
- T_0 periods before the intervention.
- Outcome of interest: Y_{jt}

Setting

Consider a complete panel with non-staggered binary treatment assignment.
WLOG, assume the first unit is treated only in the last period.

	t=1	t=2	...	t=T_0	t=T
j=1					
j=2					
...					
j=J					
j=$J+1$					

- Donor Pool: $j = 2, ..., J + 1$ units not affected by the intervention.
- T_0 periods before the intervention.
- Outcome of interest: Y_{jt}
- Predictors: For each j, we observe k predictors of the outcome X_{1j},...,X_{kj}. i)May include untreated Y_{jt} ii) Not affected by the treatment.

Setting

- Potential Outcomes: $Y_{jt}(0)$ and $Y_{jt}(1)$

Setting

- Potential Outcomes: $Y_{jt}(0)$ and $Y_{jt}(1)$
- Parameter of Interest:

$$\tau_{1T} = Y_{1T}(1) - Y_{1T}(0)$$

Setting

- Potential Outcomes: $Y_{jt}(0)$ and $Y_{jt}(1)$
- Parameter of Interest:

$$\tau_{1T} = Y_{1T}(1) - Y_{1T}(0)$$

- Fundamental problem of Causal Inference: $Y_{1T}(1) = Y_{1T}$ is observed but $Y_{1T}(0)$ is missing.

Setting

- Potential Outcomes: $Y_{jt}(0)$ and $Y_{jt}(1)$
- Parameter of Interest:

$$\tau_{1T} = Y_{1T}(1) - Y_{1T}(0)$$

- Fundamental problem of Causal Inference: $Y_{1T}(1) = Y_{1T}$ is observed but $Y_{1T}(0)$ is missing.
- SCM: Proposes to estimate $\hat{Y}_{1T}(0)$ as a weighted average of the donors' outcome and compute:

$$\hat{\tau}_{1T} = Y_{1T} - \hat{Y}_{1T}(0)$$

DiD vs. Synthetic Control Method (SCM)

- **DiD (Difference-in-Differences)**: uses the simple average of untreated units ("donors") as the control:

$$\hat{\tau}_{1T} = Y_{1T} - \frac{1}{J}\sum_{j=2}^{J+1} Y_{jT} - \frac{1}{T_0}\sum_{t=1}^{T_0}\left(Y_{1t} - \frac{1}{J}\sum_{j=2}^{J+1} Y_{jt}\right)$$
$$\equiv Y_{1T} - \hat{Y}_{1T}^{\text{DiD}}(0)$$

DiD vs. Synthetic Control Method (SCM)

- **DiD (Difference-in-Differences)**: uses the simple average of untreated units ("donors") as the control:

$$\begin{aligned}\hat{\tau}_{1T} &= Y_{1T} - \frac{1}{J}\sum_{j=2}^{J+1} Y_{jT} - \frac{1}{T_0}\sum_{t=1}^{T_0}\left(Y_{1t} - \frac{1}{J}\sum_{j=2}^{J+1} Y_{jt}\right) \\ &\equiv Y_{1T} - \hat{Y}_{1T}^{\text{DiD}}(0)\end{aligned}$$

- **Synthetic Control Method (SCM)**:

DiD vs. Synthetic Control Method (SCM)

- **DiD (Difference-in-Differences)**: uses the simple average of untreated units ("donors") as the control:

$$\begin{aligned}\hat{\tau}_{1T} &= Y_{1T} - \frac{1}{J}\sum_{j=2}^{J+1} Y_{jT} - \frac{1}{T_0}\sum_{t=1}^{T_0}\left(Y_{1t} - \frac{1}{J}\sum_{j=2}^{J+1} Y_{jt}\right) \\ &\equiv Y_{1T} - \hat{Y}_{1T}^{\text{DiD}}(0)\end{aligned}$$

- **Synthetic Control Method (SCM)**:
  - What if we found a weighted average of donors that closely traced the pre-treatment path of Y_{1t}: a "synthetic control" unit?

DiD vs. Synthetic Control Method (SCM)

- **DiD (Difference-in-Differences)**: uses the simple average of untreated units ("donors") as the control:

$$\begin{aligned}\hat{\tau}_{1T} &= Y_{1T} - \frac{1}{J}\sum_{j=2}^{J+1} Y_{jT} - \frac{1}{T_0}\sum_{t=1}^{T_0}\left(Y_{1t} - \frac{1}{J}\sum_{j=2}^{J+1} Y_{jt}\right) \\ &\equiv Y_{1T} - \hat{Y}_{1T}^{\text{DiD}}(0)\end{aligned}$$

- **Synthetic Control Method (SCM)**:
  - What if we found a weighted average of donors that closely traced the pre-treatment path of Y_{1t}: a "synthetic control" unit?
  - If the same relationship continues into $t = T$, we can use

$$\hat{Y}_{1T}(0) = \sum_{j=2}^{J+1} \omega_{jT} Y_{jT}$$

Content

Estimations SCM

- SC is defined as a weighted average of units in the donor pool. Formally:

$$W = (\omega_2, \omega_3, \ldots, \omega_{j+1})$$

Estimations SCM

- SC is defined as a weighted average of units in the donor pool. Formally:

$$W = (\omega_2, \omega_3, \ldots, \omega_{j+1})$$

- Abadie et al. (2010) propose to choose weights so that the SCM minimizes the distance of pre-treatment predictors under the Euclidean norm for a given vector of positive constants $V = [v_1, v_2, \ldots, v_k]$:

$$W^*(V) = \underset{\omega_2, \omega_3, \ldots, \omega_{J+1}}{\arg\min} \left(\sum_{h=1}^{k} v_h \left(X_{h1} - w_2 X_{h2} - \cdots - w_{J+1} X_{hJ+1} \right)^2 \right)^{1/2}$$

subject to:

$$\omega_j \geq 0, \quad \sum_{j=2}^{J+1} \omega_j = 1$$

Estimations SCM

- Predictors can contain some or all pre-treatment outcomes and a set of covariates Z_j:
$$X_j = (Y_{j1}, ..., Y_{jT_0}, Z_j)$$
- "Simplex constraints" ensure that the counterfactual is a convex combination of the donors (avoids extrapolation).

Estimations SCM

- Predictors can contain some or all pre-treatment outcomes and a set of covariates Z_j:
$$X_j = (Y_{j1}, ..., Y_{jT_0}, Z_j)$$
- "Simplex constraints" ensure that the counterfactual is a convex combination of the donors (avoids extrapolation).
- Also, it is a form of regularization:

Estimations SCM

- Predictors can contain some or all pre-treatment outcomes and a set of covariates Z_j:
$$X_j = (Y_{j1}, ..., Y_{jT_0}, Z_j)$$
- "Simplex constraints" ensure that the counterfactual is a convex combination of the donors (avoids extrapolation).
- Also, it is a form of regularization:
  - Prevents overfitting and improves prediction properties.

Estimations SCM

- Predictors can contain some or all pre-treatment outcomes and a set of covariates Z_j:
$$X_j = (Y_{j1}, ..., Y_{jT_0}, Z_j)$$
- "Simplex constraints" ensure that the counterfactual is a convex combination of the donors (avoids extrapolation).
- Also, it is a form of regularization:
    - Prevents overfitting and improves prediction properties.
    - Typically produces a unique solution.

Estimations SCM

- Predictors can contain some or all pre-treatment outcomes and a set of covariates Z_j:

$$X_j = (Y_{j1}, ..., Y_{jT_0}, Z_j)$$

- "Simplex constraints" ensure that the counterfactual is a convex combination of the donors (avoids extrapolation).
- Also, it is a form of regularization:
  - Prevents overfitting and improves prediction properties.
  - Typically produces a unique solution.
  - Ensures a sparse solution.

Estimations SCM

- Predictors can contain some or all pre-treatment outcomes and a set of covariates Z_j:

$$X_j = (Y_{j1}, ..., Y_{jT_0}, Z_j)$$

- "Simplex constraints" ensure that the counterfactual is a convex combination of the donors (avoids extrapolation).
- Also, it is a form of regularization:
  - Prevents overfitting and improves prediction properties.
  - Typically produces a unique solution.
  - Ensures a sparse solution.
- Sparsity makes the counterfactual more interpretable.

Estimations SCM

- Predictors can contain some or all pre-treatment outcomes and a set of covariates Z_j:
$$X_j = (Y_{j1}, ..., Y_{jT_0}, Z_j)$$
- "Simplex constraints" ensure that the counterfactual is a convex combination of the donors (avoids extrapolation).
- Also, it is a form of regularization:
  - Prevents overfitting and improves prediction properties.
  - Typically produces a unique solution.
  - Ensures a sparse solution.
- Sparsity makes the counterfactual more interpretable.
- $W^*(V)$ is a function of some positive constants V that indicates the relative importance of each predictor.

Estimations SCM

- Predictors can contain some or all pre-treatment outcomes and a set of covariates Z_j:
$$X_j = (Y_{j1}, ..., Y_{jT_0}, Z_j)$$
- "Simplex constraints" ensure that the counterfactual is a convex combination of the donors (avoids extrapolation).
- Also, it is a form of regularization:
  - Prevents overfitting and improves prediction properties.
  - Typically produces a unique solution.
  - Ensures a sparse solution.
- Sparsity makes the counterfactual more interpretable.
- $W^*(V)$ is a function of some positive constants V that indicates the relative importance of each predictor.

How do we choose V?

Importance of Covariates

How to pick weights on predictors, V?

- Simple strategy: v_h is the inverse of the variance of $X_{h1}, \ldots, X_{hJ+1}$.
  - Rescales the rows of X to have unit variance.

Importance of Covariates

How to pick weights on predictors, V?

- Simple strategy: v_h is the inverse of the variance of $X_{h1}, \ldots, X_{hJ+1}$.
  - Rescales the rows of X to have unit variance.
- **?** propose to choose V to minimize the mean square prediction error of the SCM with respect to the pre-treatment outcome $Y_{1t}(0)$ for $t < T_0$.

$$V^* = \arg\min_{V \in \mathbb{V}} \sum_{t=1}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j^*(V) Y_{jt} \right)^2$$

Importance of Covariates

How to pick weights on predictors, V?

- Simple strategy: v_h is the inverse of the variance of $X_{h1}, \ldots, X_{hJ+1}$.
  - Rescales the rows of X to have unit variance.
- **?** propose to choose V to minimize the mean square prediction error of the SCM with respect to the pre-treatment outcome $Y_{1t}(0)$ for $t < T_0$.

$$V^* = \arg\min_{V \in \mathbb{V}} \sum_{t=1}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j^*(V) Y_{jt} \right)^2$$

- **?** proposes out-of-sample validation.

Importance of Covariates

How to pick weights on predictors, V?

- Simple strategy: v_h is the inverse of the variance of $X_{h1}, \ldots, X_{hJ+1}$.
  - Rescales the rows of X to have unit variance.
- **?** propose to choose V to minimize the mean square prediction error of the SCM with respect to the pre-treatment outcome $Y_{1t}(0)$ for $t < T_0$.

$$V^* = \arg\min_{V \in \mathbb{V}} \sum_{t=1}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j^*(V) Y_{jt} \right)^2$$

- **?** proposes out-of-sample validation.
  - We want to assess the relative importance of $X_{11}, \ldots, X_{k1}$ as predictors of $Y_{1t}(0)$ for $t > T_0$.

Importance of Covariates

How to pick weights on predictors, V?

- Simple strategy: v_h is the inverse of the variance of $X_{h1}, \dots, X_{hJ+1}$.
  - Rescales the rows of X to have unit variance.
- **?** propose to choose V to minimize the mean square prediction error of the SCM with respect to the pre-treatment outcome $Y_{1t}(0)$ for $t < T_0$.

$$V^* = \arg\min_{V \in \mathbb{V}} \sum_{t=1}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j^*(V) Y_{jt} \right)^2$$

- **?** proposes out-of-sample validation.
  - We want to assess the relative importance of $X_{11}, \dots, X_{k1}$ as predictors of $Y_{1t}(0)$ for $t > T_0$.
  - $Y_{1t}(0)$ only observed for $t \leq T_0$.

Importance of Covariates

How to pick weights on predictors, V?

- Simple strategy: v_h is the inverse of the variance of $X_{h1}, \dots, X_{hJ+1}$.
    - Rescales the rows of X to have unit variance.
- **?** propose to choose V to minimize the mean square prediction error of the SCM with respect to the pre-treatment outcome $Y_{1t}(0)$ for $t < T_0$.

$$V^* = \arg\min_{V \in \mathbb{V}} \sum_{t=1}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j^*(V) Y_{jt} \right)^2$$

- **?** proposes out-of-sample validation.
    - We want to assess the relative importance of $X_{11}, \dots, X_{k1}$ as predictors of $Y_{1t}(0)$ for $t > T_0$.
    - $Y_{1t}(0)$ only observed for $t \leq T_0$.
    - Assess predictive power of X on a subset of the pre-intervention data $Y_{1t}(0)$.

Out of Sample validation of V

- Split the pre-treatment period into two parts: Training period ($t < T_v$)and Validation period ($t \geq T_v$).

Out of Sample validation of V

- Split the pre-treatment period into two parts: Training period ($t < T_v$)and Validation period ($t \geq T_v$).
- For every $V \in \mathbb{V}$, estimate the SCM $W^*(V)$ using the training period and compute the prediction error in the validation period.

$$\sum_{t=T_v}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j(V) Y_{jt} \right)^2$$

Out of Sample validation of V

- Split the pre-treatment period into two parts: Training period ($t < T_v$)and Validation period ($t \geq T_v$).
- For every $V \in \mathbb{V}$, estimate the SCM $W^*(V)$ using the training period and compute the prediction error in the validation period.

$$\sum_{t=T_v}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j(V) Y_{jt} \right)^2$$

- Choose V^* that minimizes the prediction error in the validation period.

Out of Sample validation of V

- Split the pre-treatment period into two parts: Training period ($t < T_v$)and Validation period ($t \geq T_v$).
- For every $V \in \mathbb{V}$, estimate the SCM $W^*(V)$ using the training period and compute the prediction error in the validation period.

$$\sum_{t=T_v}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j(V) Y_{jt} \right)^2$$

- Choose V^* that minimizes the prediction error in the validation period.
- In some cases V^* may not be unique:

$$\sum_{t=T_v}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j(V) Y_{jt} \right)^2 + \gamma V' V$$

Out of Sample validation of V

- Split the pre-treatment period into two parts: Training period ($t < T_v$)and Validation period ($t \geq T_v$).
- For every $V \in \mathbb{V}$, estimate the SCM $W^*(V)$ using the training period and compute the prediction error in the validation period.

$$\sum_{t=T_v}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j(V) Y_{jt} \right)^2$$

- Choose V^* that minimizes the prediction error in the validation period.
- In some cases V^* may not be unique:
  - Modify loss function to favor a dense set of weights

$$\sum_{t=T_v}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j(V) Y_{jt} \right)^2 + \gamma V' V$$

Considerations on V^*

- In practice, researchers should aim to demonstrate that results are robust to different choices of V.

Considerations on V^*

- In practice, researchers should aim to demonstrate that results are robust to different choices of V.
- V^* is useful as long as it produces:

$$Y_{it} \approx \sum_{j=2}^{J+1} w_j(V) Y_{jt} \quad \text{for } t \in [T_v, T_0]$$

and

$$X_1 \approx X_0 W^*(V)$$

Considerations on V^*

- In practice, researchers should aim to demonstrate that results are robust to different choices of V.
- V^* is useful as long as it produces:

$$Y_{it} \approx \sum_{j=2}^{J+1} w_j(V) Y_{jt} \quad \text{for } t \in [T_v, T_0]$$

and

$$X_1 \approx X_0 W^*(V)$$

- If the size $X_1 - X_0 W^*(V)$ is large, **?** suggest that the SCM may not be reliable.

German Reunification Example (cont.)

As predictors, use pre-treatment GDP and covariates (average for the 1981-1990 period).

Figures/balans.jpeg

German Reunification Example (cont.)

The counterfactual for West Germany is given by a weighted average of Austria (0.42), Japan (0.16), the Netherlands (0.09), Switzerland (0.11), and the United States (0.22).

Figures/sparsity.jpeg

Content

Context requirements

Context requirements for SCM:

- "Large" treatment effect: $\frac{|Y_{it}(1) - Y_{it}(0)|}{var(Y_{it}(0))} > C$

Context requirements

Context requirements for SCM:

- "Large" treatment effect: $\frac{|Y_{it}(1) - Y_{it}(0)|}{var(Y_{it}(0))} > C$
  - Outcome variables with substantial random noise elevate the risk of overfitting.

Context requirements

Context requirements for SCM:

- "Large" treatment effect: $\frac{|Y_{it}(1) - Y_{it}(0)|}{var(Y_{it}(0))} > C$
    - Outcome variables with substantial random noise elevate the risk of overfitting.
    - Volatility of the outcome can be reduced by averaging over time or filtering. (**?**)

Context requirements

Context requirements for SCM:

- "Large" treatment effect: $\frac{|Y_{it}(1)-Y_{it}(0)|}{var(Y_{it}(0))} > C$
  - Outcome variables with substantial random noise elevate the risk of overfitting.
  - Volatility of the outcome can be reduced by averaging over time or filtering. (**?**)
- "Nice" Donor Pool:

Context requirements

Context requirements for SCM:

- "Large" treatment effect: $\frac{|Y_{it}(1)-Y_{it}(0)|}{var(Y_{it}(0))} > C$
  - Outcome variables with substantial random noise elevate the risk of overfitting.
  - Volatility of the outcome can be reduced by averaging over time or filtering. (**?**)
- "Nice" Donor Pool:
  - Unaffected by the treatment

Context requirements

Context requirements for SCM:

- "Large" treatment effect: $\frac{|Y_{it}(1) - Y_{it}(0)|}{var(Y_{it}(0))} > C$
  - Outcome variables with substantial random noise elevate the risk of overfitting.
  - Volatility of the outcome can be reduced by averaging over time or filtering. (**?**)
- "Nice" Donor Pool:
  - Unaffected by the treatment
  - SUTVA/No spillovers. (Austria wasn't affected by the German reunification?)

Context requirements

Context requirements for SCM:

- "Large" treatment effect: $\frac{|Y_{it}(1) - Y_{it}(0)|}{var(Y_{it}(0))} > C$
  - Outcome variables with substantial random noise elevate the risk of overfitting.
  - Volatility of the outcome can be reduced by averaging over time or filtering. (**?**)
- "Nice" Donor Pool:
  - Unaffected by the treatment
  - SUTVA/No spillovers. (Austria wasn't affected by the German reunification?)
  - Similar to the treated unit in terms of pre-treatment outcomes and covariates.

Context requirements

Context requirements for SCM:

- "Large" treatment effect: $\frac{|Y_{it}(1)-Y_{it}(0)|}{var(Y_{it}(0))} > C$
  - Outcome variables with substantial random noise elevate the risk of overfitting.
  - Volatility of the outcome can be reduced by averaging over time or filtering. (**?**)
- "Nice" Donor Pool:
  - Unaffected by the treatment
  - SUTVA/No spillovers. (Austria wasn't affected by the German reunification?)
  - Similar to the treated unit in terms of pre-treatment outcomes and covariates.
  - Not affected by "large" idiosyncratic shocks (e.g. natural disasters)

Context requirements

Context requirements for SCM:

- "Large" treatment effect: $\frac{|Y_{it}(1) - Y_{it}(0)|}{var(Y_{it}(0))} > C$
  - Outcome variables with substantial random noise elevate the risk of overfitting.
  - Volatility of the outcome can be reduced by averaging over time or filtering. (**?**)
- "Nice" Donor Pool:
  - Unaffected by the treatment
  - SUTVA/No spillovers. (Austria wasn't affected by the German reunification?)
  - Similar to the treated unit in terms of pre-treatment outcomes and covariates.
  - Not affected by "large" idiosyncratic shocks (e.g. natural disasters)
- No anticipation:

Context requirements

Context requirements for SCM:

- "Large" treatment effect: $\frac{|Y_{it}(1) - Y_{it}(0)|}{var(Y_{it}(0))} > C$
    - Outcome variables with substantial random noise elevate the risk of overfitting.
    - Volatility of the outcome can be reduced by averaging over time or filtering. (**?**)
- "Nice" Donor Pool:
    - Unaffected by the treatment
    - SUTVA/No spillovers. (Austria wasn't affected by the German reunification?)
    - Similar to the treated unit in terms of pre-treatment outcomes and covariates.
    - Not affected by "large" idiosyncratic shocks (e.g. natural disasters)
- No anticipation:
    - This can be solved by backdating the date of the intervention

Context requirements (cont)

- Convex Hull Condition

$$\sum_{h=1}^{k} v_h \left(X_{h1} - w_2 X_{h2} - \cdots - w_{J+1} X_{hJ+1} \right) \approx 0$$

Context requirements (cont)

- Convex Hull Condition

$$\sum_{h=1}^{k} v_h \left(X_{h1} - w_2 X_{h2} - \cdots - w_{J+1} X_{hJ+1} \right) \approx 0$$

  - Predictors of the treated unit should be well approximated by a convex combination of the predictors of the donors.

Context requirements (cont)

- Convex Hull Condition

$$\sum_{h=1}^{k} v_h \left(X_{h1} - w_2 X_{h2} - \cdots - w_{J+1} X_{hJ+1} \right) \approx 0$$

  - Predictors of the treated unit should be well approximated by a convex combination of the predictors of the donors.
  - Less problematic if at least the outcome is well approximated (e.g. Inflation in West Germany)

Context requirements (cont)

- Convex Hull Condition

$$\sum_{h=1}^{k} v_h \left(X_{h1} - w_2 X_{h2} - \cdots - w_{J+1} X_{hJ+1} \right) \approx 0$$

  - Predictors of the treated unit should be well approximated by a convex combination of the predictors of the donors.
  - Less problematic if at least the outcome is well approximated (e.g. Inflation in West Germany)
  - If the outcome is not well approximated, consider taking first-differences (similar to cointegration)

Context requirements (cont)

- Convex Hull Condition

$$\sum_{h=1}^{k} v_h \left(X_{h1} - w_2 X_{h2} - \cdots - w_{J+1} X_{hJ+1}\right) \approx 0$$

  - Predictors of the treated unit should be well approximated by a convex combination of the predictors of the donors.
  - Less problematic if at least the outcome is well approximated (e.g. Inflation in West Germany)
  - If the outcome is not well approximated, consider taking first-differences (similar to cointegration)
- Pre-Intervention Data

Context requirements (cont)

- Convex Hull Condition

$$\sum_{h=1}^{k} v_h \left(X_{h1} - w_2 X_{h2} - \cdots - w_{J+1} X_{hJ+1} \right) \approx 0$$

  - Predictors of the treated unit should be well approximated by a convex combination of the predictors of the donors.
  - Less problematic if at least the outcome is well approximated (e.g. Inflation in West Germany)
  - If the outcome is not well approximated, consider taking first-differences (similar to cointegration)
- Pre-Intervention Data
  - **?** shows that bias is inversely proportional to the length of the pre-intervention period.

Context requirements (cont)

- Convex Hull Condition

$$\sum_{h=1}^{k} v_h \left(X_{h1} - w_2 X_{h2} - \cdots - w_{J+1} X_{hJ+1}\right) \approx 0$$

  - Predictors of the treated unit should be well approximated by a convex combination of the predictors of the donors.
  - Less problematic if at least the outcome is well approximated (e.g. Inflation in West Germany)
  - If the outcome is not well approximated, consider taking first-differences (similar to cointegration)
- Pre-Intervention Data
  - **?** shows that bias is inversely proportional to the length of the pre-intervention period.
  - Structural breaks can be problematic.

Context requirements (cont)

- Convex Hull Condition

$$\sum_{h=1}^{k} v_h \left(X_{h1} - w_2 X_{h2} - \cdots - w_{J+1} X_{hJ+1} \right) \approx 0$$

  - Predictors of the treated unit should be well approximated by a convex combination of the predictors of the donors.
  - Less problematic if at least the outcome is well approximated (e.g. Inflation in West Germany)
  - If the outcome is not well approximated, consider taking first-differences (similar to cointegration)
- Pre-Intervention Data
  - **?** shows that bias is inversely proportional to the length of the pre-intervention period.
  - Structural breaks can be problematic.
  - Choosing V to favour closer predictors can help.

Content

Inference

How to determine if the observed effects are statistically significant? (We have only two observations per year)

- **?** propose an "*exact p-value*" test (similar to Fisher's test).

Inference

How to determine if the observed effects are statistically significant? (We have only two observations per year)

- **?** propose an "*exact p-value*" test (similar to Fisher's test).
- H_0: "*no effect whatsoever*" **?**.

$$H_0 : \quad Y_{jt}(1) = Y_{jt}(0) \quad \text{for each } j, t$$

Inference

How to determine if the observed effects are statistically significant? (We have only two observations per year)

- **?** propose an "*exact p-value*" test (similar to Fisher's test).
- H_0: "*no effect whatsoever*" **?**.

$$H_0 : \quad Y_{jt}(1) = Y_{jt}(0) \quad \text{for each } j, t$$

- Algorithm: For each j in the donors' sample

Inference

How to determine if the observed effects are statistically significant? (We have only two observations per year)

- **?** propose an "*exact p-value*" test (similar to Fisher's test).
- H_0: "*no effect whatsoever*" **?**.

$$H_0 : \quad Y_{jt}(1) = Y_{jt}(0) \quad \text{for each } j, t$$

- Algorithm: For each j in the donors' sample
  - Iteratively compute SCM using all units "$-j$" as donors (also $j = 1$).

Inference

How to determine if the observed effects are statistically significant? (We have only two observations per year)

- **?** propose an "*exact p-value*" test (similar to Fisher's test).
- H_0: "*no effect whatsoever*" **?**.

$$H_0 : \quad Y_{jt}(1) = Y_{jt}(0) \quad \text{for each } j, t$$

- Algorithm: For each j in the donors' sample
  - Iteratively compute SCM using all units "$-j$" as donors (also $j = 1$).
  - Compute the treatment effect $\hat{\tau}_{jT}$.

Inference

How to determine if the observed effects are statistically significant? (We have only two observations per year)

- **?** propose an "*exact p-value*" test (similar to Fisher's test).
- H_0: "*no effect whatsoever*" **?**.

$$H_0 : \quad Y_{jt}(1) = Y_{jt}(0) \quad \text{for each } j, t$$

- Algorithm: For each j in the donors' sample
  - Iteratively compute SCM using all units "$-j$" as donors (also $j = 1$).
  - Compute the treatment effect $\hat{\tau}_{jT}$.
  - Compute the pre-period prediction error.

$$RMSPE_j = \left[\frac{1}{T_0} \sum_{t=1}^{T_0} \left(Y_{jt} - \hat{Y}_{jt}(0) \right)^2 \right]^{1/2}$$

Inference

How to determine if the observed effects are statistically significant? (We have only two observations per year)

- **?** propose an "*exact p-value*" test (similar to Fisher's test).
- H_0: "*no effect whatsoever*" **?**.

$$H_0 : \quad Y_{jt}(1) = Y_{jt}(0) \quad \text{for each } j, t$$

- Algorithm: For each j in the donors' sample
  - Iteratively compute SCM using all units "$-j$" as donors (also $j = 1$).
  - Compute the treatment effect $\hat{\tau}_{jT}$.
  - Compute the pre-period prediction error.

$$RMSPE_j = \left[\frac{1}{T_0} \sum_{t=1}^{T_0} \left(Y_{jt} - \hat{Y}_{jt}(0) \right)^2 \right]^{1/2}$$

  - Compute the test statistic as the ratio: $r_j = \frac{\hat{\tau}_{jT}}{RMSPE_j}$

German Reunification Example (cont.)

- Reject H_0 if $p = \frac{1}{J+1} \sum_{J+1}^{j=1} \mathbf{1}\{r_j > r_1\} < c$, for some pre-specified c (e.g. $c = 0.05$).

Figures/inference.jpeg

Content

Robustness Check - Placebo Tests

- Inference is realized using "Placebo interventions" with modified treated units.

Robustness Check - Placebo Tests

- Inference is realized using "Placebo interventions" with modified treated units.
- Alternatively, use the same treated unit but with a different intervention date.

Figures/placebo.jpeg

Robustness Check - Leave-One-Out

- We may want to evaluate how sensitive are estimations to the set of controls

Robustness Check - Leave-One-Out

- We may want to evaluate how sensitive are estimations to the set of controls
- Leave-one-out: Sequentially remove non-zero weights from the donor pool.

Figures/leaveone.jpeg

Robustness Check - Weighted Inference

- Benchmark rejection rule on the p-value formula weights all units equally.

Robustness Check - Weighted Inference

- Benchmark rejection rule on the p-value formula weights all units equally.
- We may want to evaluate how sensitive is inference to the equal-weights choices

Robustness Check - Weighted Inference

- Benchmark rejection rule on the p-value formula weights all units equally.
- We may want to evaluate how sensitive is inference to the equal-weights choices
- **?** propose to evaluate robustness by changing the weights using the parametric model:

$$p = \sum_{J+1}^{j=1} \frac{exp(\phi v_j)}{\sum_{j'=1}^{J+1} exp(\phi v'_j)} \mathbf{1}\{r_j > r_1\}$$

Robustness Check - Weighted Inference

- ϕ is the sensitivity to heterogeneity in weights.
- $\phi = 0$ implies equal weights
- If p changes too abruptly with ϕ, the inference is not robust.

Figures/firpo.jpeg

Content

Factor Model

- The SCM can be seen as a linear factor model.

$$Y_{jt}(0) = \delta_t + \lambda_t \mu_j + \theta_t Z_j + \epsilon_{jt}$$

Factor Model

- The SCM can be seen as a linear factor model.

$$Y_{jt}(0) = \delta_t + \lambda_t \mu_j + \theta_t Z_j + \epsilon_{jt}$$

- In time series literature: λ_t and θ_t are the common factors, Z_j and μ_j are the factor loadings. δ_t is a common factor with constant loadings

Factor Model

- The SCM can be seen as a linear factor model.

$$Y_{jt}(0) = \delta_t + \lambda_t \mu_j + \theta_t Z_j + \epsilon_{jt}$$

- In time series literature: λ_t and θ_t are the common factors, Z_j and μ_j are the factor loadings. δ_t is a common factor with constant loadings
- Diff-in-Diff: $\lambda_t = \lambda$. PTA implies $\delta_t + \theta_t z + \lambda \mu_j$ for all j such that $Z_j = z$

Factor Model

- The SCM can be seen as a linear factor model.

$$Y_{jt}(0) = \delta_t + \lambda_t \mu_j + \theta_t Z_j + \epsilon_{jt}$$

- In time series literature: λ_t and θ_t are the common factors, Z_j and μ_j are the factor loadings. δ_t is a common factor with constant loadings
- Diff-in-Diff: $\lambda_t = \lambda$. PTA implies $\delta_t + \theta_t z + \lambda \mu_j$ for all j such that $Z_j = z$
- *SCM* relax PTA

Factor Model

- The SCM can be seen as a linear factor model.

$$Y_{jt}(0) = \delta_t + \lambda_t \mu_j + \theta_t Z_j + \epsilon_{jt}$$

- In time series literature: λ_t and θ_t are the common factors, Z_j and μ_j are the factor loadings. δ_t is a common factor with constant loadings
- Diff-in-Diff: $\lambda_t = \lambda$. PTA implies $\delta_t + \theta_t z + \lambda \mu_j$ for all j such that $Z_j = z$
- *SCM* relax PTA
- The *SC* is a weighted average of the donors' outcomes:

$$\sum_{j=2}^{w_j} w_j Y_{jt}(0) = \delta_t + \theta_t \sum_{j=2}^{J+1} w_j Z_j + \lambda_t \sum_{j=2}^{J+1} w_j \mu_j + \sum_{j=2}^{J+1} w_j \epsilon_{jt}.$$

Bias Bound

- If the SCM is a good approximation of the factor model, we can use the factor model to bound the bias of the SCM.

Bias Bound

- If the SCM is a good approximation of the factor model, we can use the factor model to bound the bias of the SCM.
- **?** shows that if $\sum_{j=2}^{J+1} w_j^* Y_{jt}(0) = Y_{1t}(0)$ and $\sum_{j=2}^{J+1} w_j^* Z_j = Z_1$ then:

$$Bias_t = \sum_{j=2}^{J+1} w_j^* \sum_{s=1}^{T_0} \frac{(\lambda_s \lambda_t)(\epsilon_{js} - \epsilon_{1s})}{\sum_{n=1}^{T_0} \lambda_n^2} - \sum_{j=2}^{J+1} w_j^* (\epsilon_{jt} - \epsilon_{1t})$$

Bias Bound

- If the SCM is a good approximation of the factor model, we can use the factor model to bound the bias of the SCM.
- **?** shows that if $\sum_{j=2}^{J+1} w_j^* Y_{jt}(0) = Y_{1t}(0)$ and $\sum_{j=2}^{J+1} w_j^* Z_j = Z_1$ then:

$$Bias_t = \sum_{j=2}^{J+1} w_j^* \sum_{s=1}^{T_0} \frac{(\lambda_s \lambda_t)(\epsilon_{js} - \epsilon_{1s})}{\sum_{n=1}^{T_0} \lambda_n^2} - \sum_{j=2}^{J+1} w_j^* (\epsilon_{jt} - \epsilon_{1t})$$

- As $T_0 \to \infty$ we have that $Bias_t \to 0$

Bias Bound

- If the SCM is a good approximation of the factor model, we can use the factor model to bound the bias of the SCM.
- **?** shows that if $\sum_{j=2}^{J+1} w_j^* Y_{jt}(0) = Y_{1t}(0)$ and $\sum_{j=2}^{J+1} w_j^* Z_j = Z_1$ then:

$$Bias_t = \sum_{j=2}^{J+1} w_j^* \sum_{s=1}^{T_0} \frac{(\lambda_s \lambda_t)(\epsilon_{js} - \epsilon_{1s})}{\sum_{n=1}^{T_0} \lambda_n^2} - \sum_{j=2}^{J+1} w_j^* (\epsilon_{jt} - \epsilon_{1t})$$

- As $T_0 \to \infty$ we have that $Bias_t \to 0$
- For short pre-treatment, the bias depends on the variability of idiosyncratic errors.

Bias Bound

- If the SCM is a good approximation of the factor model, we can use the factor model to bound the bias of the SCM.
- **?** shows that if $\sum_{j=2}^{J+1} w_j^* Y_{jt}(0) = Y_{1t}(0)$ and $\sum_{j=2}^{J+1} w_j^* Z_j = Z_1$ then:

$$Bias_t = \sum_{j=2}^{J+1} w_j^* \sum_{s=1}^{T_0} \frac{(\lambda_s \lambda_t)(\epsilon_{js} - \epsilon_{1s})}{\sum_{n=1}^{T_0} \lambda_n^2} - \sum_{j=2}^{J+1} w_j^* (\epsilon_{jt} - \epsilon_{1t})$$

- As $T_0 \to \infty$ we have that $Bias_t \to 0$
- For short pre-treatment, the bias depends on the variability of idiosyncratic errors.
- If $X_1 - X_0 W^*$ is significant, bias may be large even for $T_0 \to \infty$

Bias Bound

- If the SCM is a good approximation of the factor model, we can use the factor model to bound the bias of the SCM.
- **?** shows that if $\sum_{j=2}^{J+1} w_j^* Y_{jt}(0) = Y_{1t}(0)$ and $\sum_{j=2}^{J+1} w_j^* Z_j = Z_1$ then:

$$Bias_t = \sum_{j=2}^{J+1} w_j^* \sum_{s=1}^{T_0} \frac{(\lambda_s \lambda_t)(\epsilon_{js} - \epsilon_{1s})}{\sum_{n=1}^{T_0} \lambda_n^2} - \sum_{j=2}^{J+1} w_j^* (\epsilon_{jt} - \epsilon_{1t})$$

- As $T_0 \to \infty$ we have that $Bias_t \to 0$
- For short pre-treatment, the bias depends on the variability of idiosyncratic errors.
- If $X_1 - X_0 W^*$ is significant, bias may be large even for $T_0 \to \infty$
- For fixed T_0, the bias is increasing in J

Extensions: Multiple Treated Units

- Conceptually equivalent to the single treated unit case.

Extensions: Multiple Treated Units

- Conceptually equivalent to the single treated unit case.
- Repeat the SCM for each treated unit and average the results.

Extensions: Multiple Treated Units

- Conceptually equivalent to the single treated unit case.
- Repeat the SCM for each treated unit and average the results.
- However the optimal weights may not be unique and inference may be more challenging.

Extensions: Multiple Treated Units

- Conceptually equivalent to the single treated unit case.
- Repeat the SCM for each treated unit and average the results.
- However the optimal weights may not be unique and inference may be more challenging.
    - **?** proposes solutions to these challenges.

Extensions: Multiple Treated Units

- Conceptually equivalent to the single treated unit case.
- Repeat the SCM for each treated unit and average the results.
- However the optimal weights may not be unique and inference may be more challenging.
  - **?** proposes solutions to these challenges.
  - Multiple solutions: Penalized SCM (next)

Extensions: Multiple Treated Units

- Conceptually equivalent to the single treated unit case.
- Repeat the SCM for each treated unit and average the results.
- However the optimal weights may not be unique and inference may be more challenging.
  - **?** proposes solutions to these challenges.
  - Multiple solutions: Penalized SCM (next)
  - Inference: Generalized permutation test. I treated units reasigned among $I + J$ units and for each permutation $b = 1, 2, .., B$ the rank statistics $r_{b1}, ..., r_{bI}$ are computed. $t_b = sum_{i=1}^{I} r_{bi}$

$$p = \frac{1}{B+1} \sum_{b=2}^{B} \mathbf{1}\{t_b \geq t_1\}$$

Extensions: Multiple Treated Units

- Conceptually equivalent to the single treated unit case.
- Repeat the SCM for each treated unit and average the results.
- However the optimal weights may not be unique and inference may be more challenging.
  - **?** proposes solutions to these challenges.
  - Multiple solutions: Penalized SCM (next)
  - Inference: Generalized permutation test. I treated units reasigned among $I + J$ units and for each permutation $b = 1, 2, .., B$ the rank statistics $r_{b1}, ..., r_{bI}$ are computed. $t_b = sum_{i=1}^{I} r_{bi}$

$$p = \frac{1}{B+1} \sum_{b=2}^{B} \mathbf{1}\{t_b \geq t_1\}$$

- Alternativley, **?** propose a single SC for all treated units.

Extensions: Penalized SCM

- In many cases, for multiple treated units, the predictors of some treated units may fall in the convex hull of the columns of X_0.

$$X_1 = X_0 W^*$$

Extensions: Penalized SCM

- In many cases, for multiple treated units, the predictors of some treated units may fall in the convex hull of the columns of X_0.

$$X_1 = X_0 W^*$$

- Typically, the number of solutions W^* will be infinite.

Extensions: Penalized SCM

- In many cases, for multiple treated units, the predictors of some treated units may fall in the convex hull of the columns of X_0.

$$X_1 = X_0 W^*$$

- Typically, the number of solutions W^* will be infinite.
- Even if $X_1 \approx X_0 W^*$, X_j may lie far away of X_1 for some j creating interpolation bias.

Extensions: Penalized SCM

- In many cases, for multiple treated units, the predictors of some treated units may fall in the convex hull of the columns of X_0.

$$X_1 = X_0 W^*$$

- Typically, the number of solutions W^* will be infinite.
- Even if $X_1 \approx X_0 W^*$, X_j may lie far away of X_1 for some j creating interpolation bias.
- Penalized SCM: Add a penalty term $\lambda > 0$ to the objective function to restrict the "quality" of donors:

$$||X_i - X_0 W||^2 + \lambda \sum_{j=I+1}^{I+J} w_i ||X_i - X_j||^2$$

Extensions: Penalized SCM

- The penalty term reduces interpolation bias

Extensions: Penalized SCM

- The penalty term reduces interpolation bias
- As $\lambda \to \infty$ SCM converges to Nearest-Neighbour Matching

Extensions: Penalized SCM

- The penalty term reduces interpolation bias
- As $\lambda \to \infty$ SCM converges to Nearest-Neighbour Matching
- $0 < \lambda < \infty$ implies a trade-off between aggregate fit vs pairwise fit.

Extensions: Penalized SCM

- The penalty term reduces interpolation bias
- As $\lambda \to \infty$ SCM converges to Nearest-Neighbour Matching
- $0 < \lambda < \infty$ implies a trade-off between aggregate fit vs pairwise fit.

Theorem

(**?**) *If $\lambda > 0$, $W_i^*(\lambda)$ is unique and sparse, with at most $p + 1$ nonzero components.*

Extensions: Penalized SCM

- The penalty term reduces interpolation bias
- As $\lambda \to \infty$ SCM converges to Nearest-Neighbour Matching
- $0 < \lambda < \infty$ implies a trade-off between aggregate fit vs pairwise fit.

Theorem

*(**?**) If $\lambda > 0$, $W_i^*(\lambda)$ is unique and sparse, with at most $p + 1$ nonzero components.*

Figures/penalized.jpeg

Extensions: Synthetic Diff-in-Diff

- Traditional Diff-in-Diff puts equal weight on all pre-treatment periods and units, but if some periods/units are more useful in prediction, we could improve the estimation.

Extensions: Synthetic Diff-in-Diff

- Traditional Diff-in-Diff puts equal weight on all pre-treatment periods and units, but if some periods/units are more useful in prediction, we could improve the estimation.
- **?** introduced an SCM that weights not only the units but also the pre-intervention time periods.

Extensions: Synthetic Diff-in-Diff

- Traditional Diff-in-Diff puts equal weight on all pre-treatment periods and units, but if some periods/units are more useful in prediction, we could improve the estimation.
- **?** introduced an SCM that weights not only the units but also the pre-intervention time periods.
- The time weights play a similar role as the predictor weights $V = (v_1, \ldots, v_k)$, reflecting the importance of each time period on the prediction of the treated-unit outcome.

$$A\hat{T}T = \bar{Y}_{\text{treated, post}} - \sum_{i=1}^{N_0} \hat{w}_i \bar{Y}_{i,\text{post}} - \sum_{t=1}^{T_0} \hat{v}_t \left(\bar{Y}_{\text{treated, } t} - \sum_{i=1}^{N_0} \hat{w}_i Y_{i,t} \right)$$

Extensions: Synthetic Diff-in-Diff

- Traditional Diff-in-Diff puts equal weight on all pre-treatment periods and units, but if some periods/units are more useful in prediction, we could improve the estimation.
- **?** introduced an SCM that weights not only the units but also the pre-intervention time periods.
- The time weights play a similar role as the predictor weights $V = (v_1, \ldots, v_k)$, reflecting the importance of each time period on the prediction of the treated-unit outcome.

$$A\hat{T}T = \bar{Y}_{\text{treated, post}} - \sum_{i=1}^{N_0} \hat{w}_i \bar{Y}_{i,\text{post}} - \sum_{t=1}^{T_0} \hat{v}_t \left(\bar{Y}_{\text{treated, } t} - \sum_{i=1}^{N_0} \hat{w}_i Y_{i,t} \right)$$

- Implemented TWFE weighted by $\hat{w}_i \cdot \hat{v}_t$ (with $\hat{v}_T = 1$ for $T > T + 0$ and $w_i = 1$ for treated units).

SCM as Online Linear Regression

- **?** recognizes that SCM is an online convex optimization algorithm known as Follow-the-leader. At each period t:
  - An online player chooses a set of weights(e.g. $W_t = w_1, .., w_J$)
  - An adversarial chosses a loss function (e.g. $l_t = ||X_i - X_0 W||^2$)
  - The player suffers a loss $l_t(w_t)$ and the process restarts the next period.

SCM as Online Linear Regression

- **?** recognizes that SCM is an online convex optimization algorithm known as Follow-the-leader. At each period t:
  - An online player chooses a set of weights(e.g. $W_t = w_1, .., w_J$)
  - An adversarial chosses a loss function (e.g. $l_t = ||X_i - X_0 W||^2$)
  - The player suffers a loss $l_t(w_t)$ and the process restarts the next period.
- The idea is that choosing an optimal set of weights would minimize the cumulative loss (Regret). The adversarial can choose any loss function and even the outcomes X_i, X_0

SCM as Online Linear Regression

- **?** recognizes that SCM is an online convex optimization algorithm known as Follow-the-leader. At each period t:
  - An online player chooses a set of weights(e.g. $W_t = w_1, .., w_J$)
  - An adversarial chosses a loss function (e.g. $l_t = ||X_i - X_0 W||^2$)
  - The player suffers a loss $l_t(w_t)$ and the process restarts the next period.
- The idea is that choosing an optimal set of weights would minimize the cumulative loss (Regret). The adversarial can choose any loss function and even the outcomes X_i, X_0

Theorem

With bounded outcomes $||Y||_\infty \leq 1$, the SCM satisfies a regret bound of:

$$Regret_T(Y) \leq 16N(log(\sqrt[root]{N}T) + 1)) = O(NlogT)$$

SCM as Online Linear Regression

- **?** recognizes that SCM is an online convex optimization algorithm known as Follow-the-leader. At each period t:
  - An online player chooses a set of weights(e.g. $W_t = w_1, .., w_J$)
  - An adversarial chosses a loss function (e.g. $l_t = ||X_i - X_0 W||^2$)
  - The player suffers a loss $l_t(w_t)$ and the process restarts the next period.
- The idea is that choosing an optimal set of weights would minimize the cumulative loss (Regret). The adversarial can choose any loss function and even the outcomes X_i, X_0

Theorem

With bounded outcomes $||Y||_\infty \leq 1$, the SCM satisfies a regret bound of:

$$Regret_T(Y) \leq 16N(log(\sqrt[root]{N}T) + 1)) = O(NlogT)$$

- This theorem shows that *SCM* performs better or equal than any other weighted average of the donors (including Diff-in-Diff and Synthetic Diff-in-Diff)

References I