

No.

Date

② Strict Exogeneity

Parallel trends

No anticipation

No unobserved time varying confounders

Past treatment don't affect current outcome (no carryover)
Past outcome don't affect current treatment (no feedback)

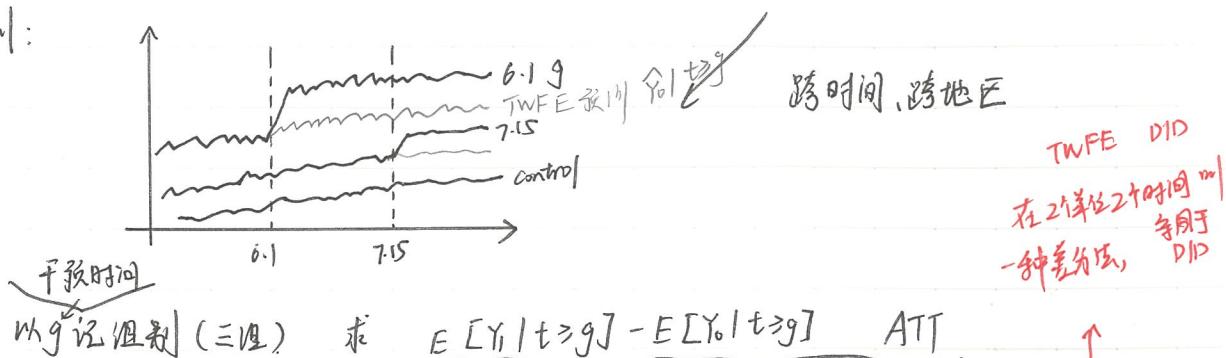
Campus

24. The Difference - Difference Saga

面板数据

跨时间重复观察一个单位

例：



模型：

$$\text{Installit} = \gamma D_{it} + \gamma_i + \theta_t + e_{it}$$

用时停控判所有特征

noise

若用 OLS 估计，则为 TWEE：双向
(γ_i 为组别， θ_t 为时间) 固定效应
模型
 γ_i 与 t 相关， θ_t 与 t 相关)

数学上等价处理： $\tilde{\text{Installit}} = \tilde{\gamma} \tilde{D}_{it} + e_{it}$

$$\text{where: } \left\{ \begin{array}{l} \tilde{\text{Installit}} = \text{Installit} - \frac{1}{T} \sum_{t=0}^T \text{Installit} - \frac{1}{N} \sum_{i=0}^N \text{Installit} \\ \text{time average} \quad \text{unit average} \end{array} \right.$$

$$\tilde{D}_{it} = D_{it} - \frac{1}{T} \sum_{t=0}^T D_{it} - \frac{1}{N} \sum_{i=0}^N D_{it}$$

Death: Failures over Effect Heterogeneity 效果异质性

TWEE 经常有 Bias

没有时间效应的 FE 模型：

$$y_{it} = \gamma D_{it} + \gamma_i + e_{it}$$

设模型假设：① Functional Form Assumptions:

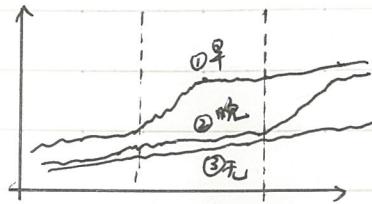
No heterogeneous effect in time (constant effects);

Linearity in the covariates;

Additive fixed effects.

接上：治疗效果在时间上的异质性：

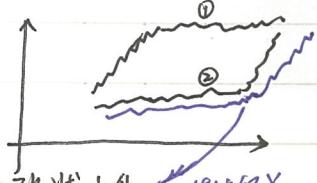
若治疗为渐进式的（从0.1、0.2、...、1共十天）
传说 TWFE 估计的 ATT 有 Bias.



当 TWFE 被分解为多个 2×2 DID:

对照 ↓	$\begin{cases} ① - ③ & \checkmark \\ ② - ③ & \checkmark \\ ① - ② & \checkmark \\ ② - ① & ?? \end{cases}$
---------	--

当以早期①为对照，模型认为 y_t 会以之前已变平缓的形状上升



Goodman-Bacon, 2019:

$$\text{plim}_{x \rightarrow \infty} \hat{\gamma}^{\text{TWFE}} = VWATT - \Delta ATT$$

Bias: ATT 随时间变化的程度

当效果随时间而↑（如例），Bias 向下

当效果为负，随时间递减，估计甚至正负号错误！

相反成立。

Event study design 事件研究设计：

重写 TWFE:

$$y_{it} = \tau^{-k} D_{ik}^{<-k} + \sum_{k=K}^{-2} \tau^{\text{lead}} D_{it}^k + \sum_{k=0}^L \tau^{\text{lag}} D_{it}^k + \tau_k^{L+1} D_{it}^{>L} + e_{it}$$

where $D_{it}^k = 1 \{t - Cohort_i = k\} \Rightarrow$ 若周期为 k 时为 1 否则为 0 加入一个衡量单位治疗开始时期有多远的量

Enlightenment: A Flexible Functional Form

新模型：

对每个个体都有一个治疗效果：

$$\text{Installsit} = \sum_{i=0}^N \sum_{t=0}^T z_{it} D_{it} + \gamma_i + \theta_t + e_{it} \quad (\text{参数太多})$$

对每个个体(g)：

$$\text{Installsit} = \sum_{g=0}^G \sum_{t=0}^T z_{gt} D_{it} + \gamma_i + \theta_t + e_{it} \quad \checkmark \text{ (优优)}$$

再减少参数量：删去治疗前、删去对照。

$$\text{Installsit} = \sum_{g=0}^G \sum_{t=g}^T z_{gt} D_{it} + \gamma_i + \theta_t + e_{it}$$

其中 $g < q$ 之前的组群被定义为控制组

需要创建组群假设 $(0,1)(1,0) \dots$

缺陷：① 治疗效果会随时间的推移而改变 (之前未考虑)

② DID 与 TWFE 必须满足 平行趋势假设 → 限制 很大很大

25. Synthetic Difference-in-Differences 合成差分法

不同时间、多个单位、治疗非治疗混杂：

$$D = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{array}{l} \leftarrow \text{行为时间} \\ \uparrow \uparrow \\ \text{成为单位} \end{array}$$

统一处理

设计中唯一要求：一旦一个单位被处理，其就

不会返回到未处理的状态。

数据表示方法1： synthetic Control

将 D 化为

$$D = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

则：

$$Y = \begin{bmatrix} Y_{pre,co} & Y_{pre,tr} \\ Y_{post,co} & Y_{post,tr} \end{bmatrix} \quad \begin{array}{l} \text{pre 治疗前, post 治疗后} \\ \text{co 控制组, tr 治疗组.} \end{array}$$

方法2：列表： 单位 时间 结果 是否为处理单位 是否已被处理
DID | | ; | | ; |

对于 Y 的观测量：

$$Y = \begin{bmatrix} Y_{10} \text{ pre,co} & Y_{10} \text{ pre,tr} \\ Y_{10} \text{ post,co} & Y_{10} \text{ post,tr} \end{bmatrix}$$

$$\text{ATT} = Y_{10} \text{ post,tr} - \underbrace{Y_{10} \text{ post,co}}_{\text{反事实}}$$

重新审视 DID 与 Synthetic Controls

DID: $Y_{it} = \beta_0 + \beta_1 Post_t + \beta_2 Treated_i + \beta_3 Treated_i Post_t + e_{it}$

DID 前提: $E[Y_{10}^{(0)} post, co] - E[Y_{10}^{(0)} pre, co] = E[Y_{10}^{(0)} post, tr] - E[Y_{10}^{(0)} pre, tr]$

即 对照组与治疗组在非治疗下趋势相同

用以下方式重写, 与 DID 估计器原理与结果相同

$$\hat{\gamma}_{DID} = \underset{\mu, \alpha, \beta, \gamma}{\operatorname{argmin}} \left\{ \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - (\mu + \alpha_i + \beta t + \gamma D_{it}))^2 \right\}$$

$$\hat{Y}_{it} = Y_{it} - \bar{Y}_i - \bar{Y}_t = Y_{it} - T^{-1} \sum_{t=0}^{T-1} Y_{it} - N^{-1} \sum_{i=0}^{N-1} Y_{it}$$

$$\hat{D}_{it} = D_{it} - \bar{D}_i - \bar{D}_t = D_{it} - T^{-1} \sum_{t=0}^{T-1} D_{it} - N^{-1} \sum_{i=0}^{N-1} D_{it}$$

这个公式与 Synthetic Controls 相似, 但缺少权重;

Synthetic Controls

$$\hat{w}^{SC}$$

前沿 - 前拦截 崩塌

$$\underset{w}{\operatorname{argmin}} \| \bar{y}_{pre, tr} - Y_{pre, co} w \|^2$$

$$\text{s.t. } \sum w_i = 1 \text{ and } w_i > 0 \forall i$$

有时将合成控制
称为水平回归

where the outcome $Y_{pre, co}$ is a T_{pre} by N_{co} matrix, 列是单位, 行是时间段

w_{co} is a N_{co} by 1 column vector 每个条目是治疗前的治疗

得之后:
控. 权

单位的时间

$$ysc = Y_{co} \hat{w}^{SC}$$

平均值

RJ

ATT 为 $\hat{\tau} = \bar{y}_{post, tr} - \bar{y}_{post, sc}$

\hat{w}^{SC} 为逆估计的, co 组的权重

查书:

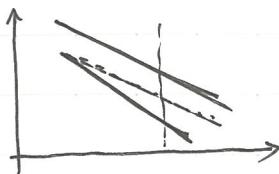
$$\hat{\gamma}_{sc} = \underset{\beta, \gamma}{\operatorname{argmin}} \left\{ \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \beta t - \gamma D_{it})^2 \hat{w}_i^{SC} \right\}$$

$\frac{1}{N_{tr}}$ 为 参照组 权重
的加权

无 μ 总体截距
无 d_i 固定效应

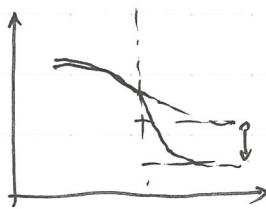
KOKUYO

DID



被叫线与对照平行，放大了效果

SC



用前半生成权重，组成虚拟未治疗治疗组。
用权重乘后半均值
后半均值之差为估计



Synthetic Diff in Diff

已有：

$$\hat{\gamma}^{sc} = \underset{\beta, z}{\operatorname{argmin}} \left\{ \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \beta_t - z D_{it})^2 \hat{w}_i^{sc} \right\}$$

$$\hat{\gamma}^{did} = \underset{\mu, d, \beta, z}{\operatorname{argmin}} \left\{ \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - (\mu + d_i + \beta_t + z D_{it}))^2 \right\}$$

\Rightarrow 合并为：

$$\hat{\gamma}^{sdid} = \underset{\mu, d, \beta, z}{\operatorname{argmin}} \left\{ \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - (\mu + d_i + \beta_t + z D_{it}))^2 \frac{\hat{w}_i^{solid}}{\hat{\gamma}_t^{solid}} \right\}$$

单位权重 时间权重

同理：

$$\hat{\gamma}^{sdid} = \underset{\lambda}{\operatorname{argmin}} \| \bar{y}_{post,co} - (\lambda \bar{y}_{pre,co} + \lambda_0) \|_2^2$$

s.t. $\lambda t = 1$ and $\lambda t > 0 \forall t$

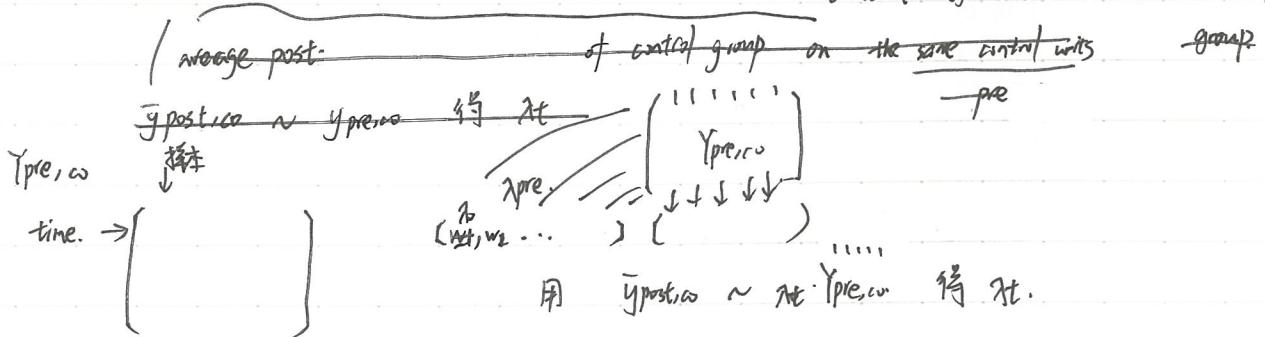
$Y_{pre,co}$ is a T_{pre} by N_{co} matrix 结果矩阵，行为时间，列为单位

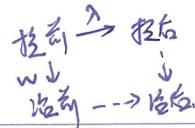
$\bar{y}_{post,co}$ is a 1 by N_{co} row vector 每条是控制单元在治疗后的平均结果

\bar{y}_{pre} is a 1 by T_{pre} row vector 每条是治疗前的时期，时间的权。

入为时间单位

w_后 Y_{pre,co} w_{co} : average outcome for each unit of the treated group on outcome of the units in the control group





L_2 penalty: 使非0权重更广泛地分布

治疗前 治疗后

$$\hat{w}^{\text{SDID}} = \underset{w}{\operatorname{argmin}} \| \bar{y}_{\text{pre}, \text{tr}} - (y_{\text{pre}, 00} w_{00} + w_0) \|_2^2 + \gamma^2 T_{\text{pre}} \| w \|^2$$

s.t. $w_i = 1$ and $w_i > 0 \forall i$

其中 $\gamma = (\lambda \nu * T_{\text{post}})^{1/4} \delta(\Delta t)$

Δt 是 $y_{it} - \bar{y}_{i(t-1)}$ 的各个差异

而 $\delta(\Delta t)$ 为 Δt 标准偏差

有截距

同理，在 $\bar{y}_{\text{pre}, \text{tr}}$ 前加一行 1 作为截距，并回归

与之前的 SC 相比，①更贴近治疗

② L_2 罚则

当处理 治疗组的单位权重 和 后治疗的时间权重，直接采用治疗组 $\frac{1}{T_{\text{post}}}$, $\lambda \nu$
SDID 相比 SC，只关注治疗近期起效果的数据；

Time Effect Heterogeneity and Staggered Adoption 时间效应的异质性和交错采用

SDID 估计每个时间段的一个效应很容易，只需多次运行

例：估斤

$$D = \begin{pmatrix} 0_1 & 0_1 & 0_1 \\ 0_2 & 0_2 & 0_2 \\ 0_3 & 0_3 & 1_3 \\ 0_4 & 0_4 & 1_4 \end{pmatrix}$$

SDID 得三、四期平均 ATT

时间效应，异质性

分割估计三、四期

$$D_1 = \begin{pmatrix} 0_1 & 0_1 & 0_1 \\ 0_2 & 0_2 & 0_2 \\ 0_3 & 0_3 & 1_3 \end{pmatrix}$$

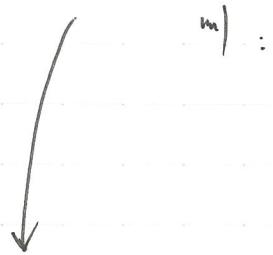
$$D_2 = \begin{pmatrix} 0_1 & 0_1 & 0_1 \\ 0_2 & 0_2 & 0_2 \\ 0_4 & 0_4 & 1_4 \end{pmatrix}$$

Placebo Variance Estimation 安慰剂差异估计

与SDID相比，SDID有更低的 error bars； SDID 捕捉时间、单位降低了结果的方差

设置SDID的置信区间：(适用于单一治疗单位)

对控制组内每个单位，假设其受到了治疗，用SDID估计该安慰剂的ATT。



$$\hat{V}_2^{\text{placebo}} = B^{-1} \sum_{b=1}^B (\hat{\tau}^{(b)} - \bar{\tau}^{(b)})^2$$

$$\bar{\tau} \in \hat{\tau}^{\text{sdid}} \pm z_{\alpha/2} \sqrt{\hat{V}_2}$$



(方差的ES值)

这个数目的方差为SDID估计的安慰剂方差，用其构建置信区间

APPENDIX:

解决非随机数据
{| 倾向得分
| 正交化}

1. Debiasing with Orthogonalization 用正交法去除偏移

同22章

所有特征在X中：

(step1):

$$y_i = \theta_0 + \theta_1 X_{ii} + e_i \quad \text{取第一组特征.}$$

取残差 $y^* = y_i - (\hat{\theta}_0 + \hat{\theta}_1 X_{ii})$

(step2):

$$X_{2i} = \gamma_0 + \gamma_1 X_{ii} + e_i \quad \text{取第二组特征估计第一组特征}$$

取残差: $X_{2i}^* = X_{2i} - (\hat{\gamma}_0 + \hat{\gamma}_1 X_{ii})$

(step3):

$$y_i^* = \beta_0 + \beta_2 X_{2i}^* + e_i$$

2. Debiasing with Propensity Score

医疗保险

对每个单位加权。

$$\text{高极: } w_i = \frac{T}{P(T=1|X)} + \frac{1-T}{1-P(T=1|X)}$$

即: 受治疗概率高但未受治疗; 其本身係治疗组但观测为控制组, 可用来估计
 $y_0 | T=1$

$$\text{进阶: } w_i = \sum_{k=0}^K \frac{1_{k \neq i}}{P(T=k|X)}$$

当没有倾向性得分时, 以其受治疗概率缺为基准做估计: ML

2%~0.2%的人应收山电子邮件 | ...

倾向性得分的弱点:

① 倾向性得分过低的样本 ($p < 0.05$) 权重为 1, 出现太多次, 会对去偏的数据集影响很大 (人工裁剪) — 高方差

② 积极性或共同支持 (Positivity or Common Support)

被处理单位和控制单位的特征之间必须有足够的重叠

(当重叠过小, 用倾向性得分消除相关性效果不佳) 负人工手动不考虑积极

不可外推主极性不成立的人口子集上

性过偏的治疗

3. When Prediction Fails. 预测与因果