双重差分模型

慧航

2025年10月

双重差分模型

违背Unconfoundedness假设，但是有面板数据的条件下：双重差分模型（Difference-in-differences）。

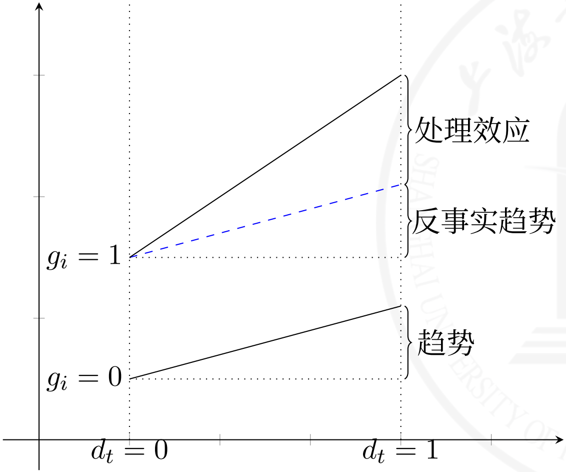
一般设定：

1. 分组变量：$g_i = 1$：处理组；$g_i = 0$：控制组。
   - 比如实施最低工资
   - unilateral divorce law
2. 时间变量：$d_t = 0/1$
3. 处理$w_{it} = d_t \cdot g_i$
4. Outcome：y_{it}

w_{it}	$g_i = 0$	$g_i = 1$
$d_t = 0$	0	0
$d_t = 1$	0	1

共同趋势假设

关键假设：共同趋势（Common trend）



双重差分模型

双重差分模型：

1. 第一次差分：对于两个不同的分组，分别计算：

$$\Delta_1 = \mathbb{E}\left(y_{it}|g_i = 1, d_t = 1\right) - \mathbb{E}\left(y_{it}|g_i = 1, d_t = 0\right)$$

$$\Delta_0 = \mathbb{E}\left(y_{it}|g_i = 0, d_t = 1\right) - \mathbb{E}\left(y_{it}|g_i = 0, d_t = 0\right)$$

2. 第二次差分：处理效应：

$$\tau = \Delta_1 - \Delta_0$$

双重差分模型

实践中，等价于使用回归：

$$y_{it} = c + \lambda \cdot d_t + \gamma \cdot g_i + \tau \cdot d_t \cdot g_i + u_{it}$$

即第一次差分：

$$\Delta_1 = \mathbb{E}\left(y_{it}|g_i = 1, d_t = 1\right) - \mathbb{E}\left(y_{it}|g_i = 1, d_t = 0\right) = \lambda + \tau$$

$$\Delta_0 = \mathbb{E}\left(y_{it}|g_i = 0, d_t = 1\right) - \mathbb{E}\left(y_{it}|g_i = 0, d_t = 0\right) = \lambda$$

第二次差分：

$$\tau = \Delta_1 - \Delta_0$$

双重差分模型中的控制变量

- 使用反事实的框架，处理效应为：$y_{i1}(1) - y_{i1}(0) = \tau_i$，相应的处理组平均处理效应为：

$$\tau_{\text{ATT}} = \mathbb{E}\left(y_{i1}(1) - y_{i1}(0) \,|g_i = 1\right)$$

问题：$\mathbb{E}\left(y_{i1}(0) \,|g_i = 1\right)$不可观测！

- （平行趋势）假设：

$$\mathbb{E}\left(y_{i1}(0) - y_{i0}(0) \,|g_i = 1\right) = \mathbb{E}\left(y_{i1}(0) - y_{i0}(0) \,|g_i = 0\right)$$

- 那么：

$$\begin{aligned}\mathbb{E}\left(y_{i1}(0) \,|g_i = 1\right) = &\ \mathbb{E}\left(y_{i0}(0) \,|g_i = 1\right) \\ &+ \mathbb{E}\left(y_{i1}(0) - y_{i0}(0) \,|g_i = 0\right)\end{aligned}$$

双重差分模型

- 假设

$$\mathbb{E}\left(y_{it}\left(0\right)|x_{it}\right) = \alpha_i + \gamma_t + x_{it}'\beta$$

  - 加入了时间和个体的虚拟变量（TWFE）
  - 并没有给定分组，因而控制组和处理组都满足以上方程。
- 由于：$\tau_{ATT} = \mathbb{E}\left(y_{i1}\left(1\right) - y_{i1}\left(0\right)|g_i = 1, x_{it}\right)$，从而：

$$\mathbb{E}\left(y_{i1}\left(1\right)|g_i = 1, x_{it}\right) = \tau_{ATT} + \mathbb{E}\left(y_{i1}\left(0\right)|g_i = 1, x_{it}\right)$$

从而设定：

$$\begin{aligned} y_{it} &= \alpha_i + \gamma_t + x_{it}'\beta + \tau_{ATT} w_{it} + u_{it} \\ &= \alpha_i + \gamma_t + x_{it}'\beta + \tau_{ATT} \cdot d_t \cdot g_i + u_{it} \end{aligned}$$

- 以上为标准设定：加入控制变量、双向固定效应使得共同趋势假设成立。

标准误的问题

- 在DID模型中，（广义的）自相关严重影响推断，cluster标准误（聚类标准误）是必须的！
- 原因：在DID模型中，同一个体的误差项之间存在着时间上的相关性，这种相关性会严重影响推断。
    - 系数问题不大
    - 但是标准误是错的！

双重差分：一定要面板数据吗？

- 传统上，双重差分模型一般需要面板数据
- 但是实际上，不使用面板数据有时也可以使用双重差分模型：
  - Duflo(2001) 关于教育回报的研究中，比较cohorts
  - Archibong and Annan(2017)关于疾病与人力资本投资
    - 设定为：

$$\text{edu} = \beta \cdot \text{MEMIN} \cdot \text{female} + \cdots + u$$

    - MEMIN为流行性脑膜炎的严重程度，为实际上的双重差分变量
    - DID_disease.do

平行趋势检验

检验：

1. Placebo test：提前冲击发生的时间
2. 平行趋势假设检验：若冲击发生时间为$t=\tilde{T}$，设定：

$$y_{it}=\alpha_i+\lambda_t+\sum_{\ell=-p,\ell\neq-1}^{q}\tau_l\cdot\mathbb{1}\left\{t=\tilde{T}+\ell\right\}\cdot g_i+x_{it}'\beta+u_{it}$$

3. 标准：
   1. τ_ℓ当$\ell<0$时(leads)不显著
   2. τ_ℓ当$\ell\geq 0$时(lag)显著

平行趋势检验

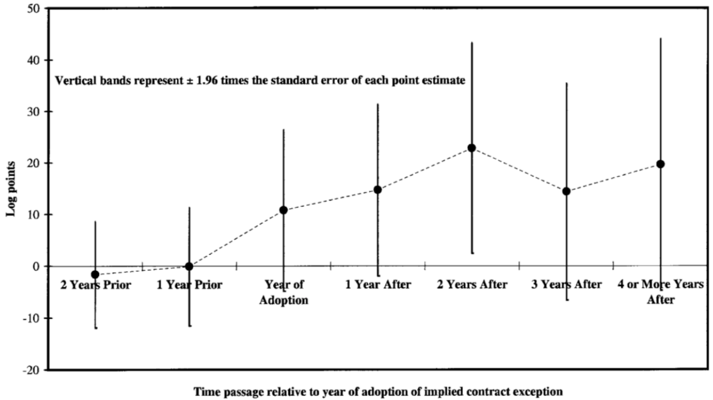


FIG. 3.—Estimated impact of implied contract exception on log state temporary help supply industry employment for years before, during, and after adoption, 1979–95.

多期DID的静态设定

- 传统做法：将$d_t \cdot G_i$替换为w_{it}，其中w_{it}指第t期第i个个体是否接受处理:

$$w_{it} = \begin{cases} 1 & \text{个体}i\text{在时间}t\text{接受处理} \\ 0 & else \end{cases}$$

设定模型（TWFE）：

$$y_{it} = \alpha_i + \lambda_t + \tau \cdot w_{it} + x'_{it}\beta + u_{it}$$

静态设定的问题

一般的DID使用如上的静态估计（双向固定效应）：

$$y_{it} = \alpha_i + \gamma_t + x'_{it}\beta + \tau_{ATT}w_{it} + u_{it}$$

Goodman-Bacon(2021) 指出，以上设定：可以写成所有2×2的DID结果的加权平均，但是会有负的权重，从而识别失败

Goodman-Bacon分解

- 首先考虑没有控制变量的简单情况，2×2的情况即最简单的只有处理组、对照组以及事前事后的设定：

$$y_{it}=c+\lambda\cdot d_t+\gamma\cdot g_i+\tau^{2\times2}\cdot d_t\cdot g_i+u_{it}$$

从而

$$\hat{\tau}^{2\times2}=\left(\bar{y}^{\text{Post}}_{\text{Treat}}-\bar{y}^{\text{Pre}}_{\text{Treat}}\right)-\left(\bar{y}^{\text{Post}}_{Control}-\bar{y}^{\text{Pre}}_{Control}\right)$$

- 然而对于多期DID，每个个体接受处理的时间t_i是不同的，TWFE估计量：

$$y_{it}=\alpha_i+\lambda_t+\tau\cdot w_{it}+u_{it}$$

分步回归得到：

$$\hat{\tau}=\frac{\hat{\mathbb{C}}\left(y_{it},\tilde{w}_{it}\right)}{\hat{\mathbb{V}}\left(\tilde{w}\right)}=\frac{\frac{1}{NT}\sum_i\sum_t y_{it}\tilde{w}_{it}}{\frac{1}{NT}\sum_i\sum_t\tilde{w}_{it}^2}$$

其中$\tilde{w}_{it}=(w_{it}-\bar{w}_i)-\left(\bar{w}_t-\bar{\bar{w}}\right)$，即使用$w$对$\alpha_i,\lambda_t$双向固定效应回归得到的残差。

Goodman-Bacon分解

- 考虑有三个cohort:
  - $t_i = k$ (early group)
  - $t_i = l$ (late group)
  - $t_i = \infty$ (control/untreat/never group)
- 几种不同的比较:
  - Treat v.s. Untreat:

$$\hat{\tau}_{j\infty}^{2\times 2} = \left(\bar{y}_j^{\text{Post}(j)} - \bar{y}_j^{\text{Pre}(j)}\right) - \left(\bar{y}_U^{\text{Post}(j)} - \bar{y}_U^{\text{Pre}(j)}\right), j = k, l$$

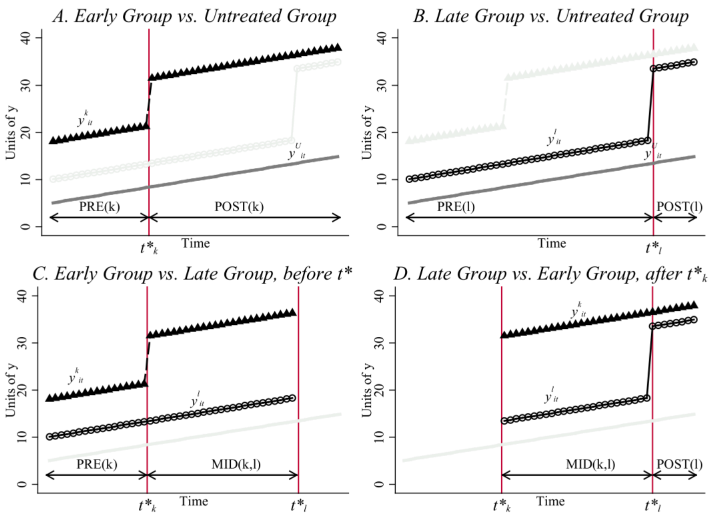
  - Early v.s. Late, Before l:

$$\hat{\tau}_{kl}^{2\times 2,k} = \left(\bar{y}_k^{\text{Mid}(k,l)} - \bar{y}_k^{\text{Pre}(k)}\right) - \left(\bar{y}_l^{\text{Mid}(k,l)} - \bar{y}_l^{\text{Pre}(k)}\right)$$

  - Late v.s. Early, After k:

$$\hat{\tau}_{kl}^{2\times 2,l} = \left(\bar{y}_l^{\text{Post}(l)} - \bar{y}_l^{\text{Mid}(k,l)}\right) - \left(\bar{y}_k^{\text{Post}(l)} - \bar{y}_k^{\text{Mid}(k,l)}\right)$$

双向固定效应的问题


A. Early Group vs. Untreated Group
B. Late Group vs. Untreated Group
C. Early Group vs. Late Group, before t*
D. Late Group vs. Early Group, after t*k
Units of y
Time
PRE(k)
POST(k)
PRE(l)
POST(l)
MID(k,l)
t*k
t*l
0
10
20
30
40

Goodman-Bacon分解

DID分解定理

双向固定效应的DID估计量可以被分解为:

$$\hat{\tau} = \sum_{j \neq u} s_{j\infty} \hat{\tau}_{j\infty}^{2\times 2} + \sum_{k \neq \infty} \sum_{l > k} \left[s_{kl}^{k} \hat{\tau}_{kl}^{2\times 2,k} + s_{kl}^{l} \hat{\tau}_{kl}^{2\times 2,l} \right]$$

其中

$$s_{j\infty} = \frac{\left(n_j + n_\infty\right)^2 n_{j\infty} \left(1 - n_{jU}\right) \bar{w}_j \left(1 - \bar{w}_j\right)}{\hat{\mathbb{V}}\left(w\right)}$$

$$s_{kl}^{k} = \frac{\left[\left(n_k + n_l\right)\left(1 - \bar{w}_l\right)\right]^2 \hat{\mathbb{V}}_{kl}^{w,k}}{\hat{\mathbb{V}}\left(w\right)}$$

$$s_{kl}^{l} = \frac{\left[\left(n_k + n_l\right)\bar{w}_k\right]^2 \hat{\mathbb{V}}_{kl}^{w,l}}{\hat{\mathbb{V}}\left(w\right)}$$

Goodman-Bacon分解

DID分解定理（续）

其中$n_k = \sum_i \mathbb{1}\{t_i = k\}/N, \bar{w}_k = \sum_t \mathbb{1}\{t \geq k\}/T$，此外

$$\hat{\mathbb{V}}_{kl}^{w,k} = n_{kl}\left(1 - n_{kl}\right) \frac{\bar{w}_k - \bar{w}_l}{1 - \bar{w}_l} \frac{1 - \bar{w}_k}{1 - \bar{w}_l}$$

$$\hat{\mathbb{V}}_{kl}^{w,l} = n_{kl}\left(1 - n_{kl}\right) \frac{\bar{w}_l}{\bar{w}_k} \frac{\bar{w}_k - \bar{w}_l}{\bar{w}_k}$$

且$\sum_{j \neq u} s_{k\infty} + \sum_{k \neq \infty} \sum_{l > k} \left[s_{kl}^k + s_{kl}^l\right] = 1$。

Goodman-Bacon分解

- 如果有K个不同的处理组cohort：
  - 有K个$\hat{\tau}_{j\infty}^{2\times 2}$
  - 有K^2-K个$\hat{\tau}_{kl}^{2\times 2,k}$以及$\hat{\tau}_{kl}^{2\times 2,l}$
  - 总共有K^2个不同的2×2DID估计量。
- 问题：这样做可以识别什么？

TWFE的识别问题

- 记潜在结果$y_{it}(k)$为一个属于k的cohort的个体i在第t期如果被处理的处理效应，从而$y_{it}(t_i)$为实际看到的结果，记没有受到处理的潜在结果为$y_{it}(\infty)$，如果$t<t_i$，那么$y_{it}(t_i)=y_{it}(\infty)$。
- 观察到的结果为：

$$y_{it}=w_{it}y_{it}(t_i)+(1-w_{it})\,y_{it}(\infty)$$

- 定义cohort k在$t\geq k$的ATT：

$$\mathrm{ATT}_k(t)=\mathbb{E}\left[y_{it}(k)-y_{it}(\infty)\,|t_i=k\right]$$

- 对于某个时间段$\mathcal{T}$，定义这一时间段的平均处理效应：

$$\mathrm{ATT}_k(\mathcal{T})=\frac{1}{|\mathcal{T}|}\sum_{t\in\mathcal{T}}\mathbb{E}\left[y_{it}(k)-y_{it}(\infty)\,|t_i=k\right]$$

TWFE的识别问题

- 记两个时间段假设未处理的潜在结果之差:

$$\Delta y_k^{\infty}\left(\mathcal{T}_1, \mathcal{T}_0\right) = \frac{1}{\left|\mathcal{T}_1\right|} \sum_{t \in \mathcal{T}_1} \mathbb{E}\left[y_{it}\left(\infty\right) | t_i = k\right] - \frac{1}{\left|\mathcal{T}_0\right|} \sum_{t \in \mathcal{T}_0} \mathbb{E}\left[y_{it}\left(\infty\right) | t_i = k\right]$$

- 那么对于$j = k, l$:

$$\begin{aligned}
\beta_{j\infty}^{2\times 2} &= \left(\frac{1}{\left|\text{Post}\left(j\right)\right|} \sum_{t \in \text{Post}(j)} \mathbb{E}\left(y_{it}\left(k\right) | t_i = k\right) - \frac{1}{\left|\text{Pre}\left(j\right)\right|} \sum_{t \in \text{Pre}(j)} \mathbb{E}\left(y_{it}\left(\infty\right) | t_i = k\right)\right) \\
&\quad - \left(\frac{1}{\left|\text{Post}\left(j\right)\right|} \sum_{t \in \text{Post}(j)} \mathbb{E}\left(y_{it}\left(\infty\right) | t_i = \infty\right) - \frac{1}{\left|\text{Pre}\left(j\right)\right|} \sum_{t \in \text{Pre}(j)} \mathbb{E}\left(y_{it}\left(\infty\right) | t_i = \infty\right)\right) \\
&= \text{ATT}_k\left(\text{Post}\left(j\right)\right) \\
&\quad + \left[\Delta y_k^{\infty}\left(\text{Post}\left(j\right), \text{Pre}\left(j\right)\right) - \Delta y_{\infty}^{\infty}\left(\text{Post}\left(j\right), \text{Pre}\left(j\right)\right)\right]
\end{aligned}$$

TWFE的识别问题

- 同理:

$$
\begin{aligned}
\beta_{kl}^{2\times 2,k} &= ATT_k\left(\text{Mid}\left(k,l\right)\right) \\
&+ \left[\Delta y_k^{\infty}\left(\text{Mid}\left(k,l\right),\text{Pre}\left(k\right)\right) - \Delta y_l^{\infty}\left(\text{Mid}\left(k,l\right),\text{Pre}\left(k\right)\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
\beta_{kl}^{2\times 2,l} &= ATT_k\left(\text{Post}\left(l\right)\right) \\
&+ \left[\Delta y_l^{\infty}\left(\text{Post}\left(l\right),\text{Mid}\left(k,l\right)\right) - \Delta y_k^{\infty}\left(\text{Post}\left(l\right),\text{Mid}\left(k,l\right)\right)\right] \\
&- \left[ATT_k\left(\text{Post}\left(l\right)\right) - ATT_k\left(\text{Mid}\left(k,l\right)\right)\right]
\end{aligned}
$$

TWFE的识别问题

最终，TWFE估计量识别了：

$$\hat{\tau} \xrightarrow{p} \text{VWATT} + \text{VWCT} - \Delta\text{ATT}$$

其中：

- 第一项：方差加权的ATT

$$\begin{aligned}\text{VWATT} = &\sum_{j \neq u} \sigma_{k\infty} \text{ATT}_k\left(\text{Post}\left(j\right)\right) \\ &+ \sum_{k \neq \infty} \sum_{l > k} \left[\sigma_{kl}^{k} \text{ATT}_k\left(\text{Mid}\left(k, l\right)\right) + \sigma_{kl}^{l} \text{ATT}_k\left(\text{Post}\left(l\right)\right)\right]\end{aligned}$$

TWFE的识别问题

- 第二项：方差加权的平行趋势

$$
\begin{aligned}
\mathrm{VWCT} = & \sum_{j \neq u} \sigma_{k\infty} \left[\Delta y_k^{\infty}\left(\mathrm{Post}\left(j\right), \mathrm{Pre}\left(j\right)\right) - \Delta y_{\infty}^{\infty}\left(\mathrm{Post}\left(j\right), \mathrm{Pre}\left(j\right)\right)\right] \\
& + \sum_{k \neq \infty} \sum_{l > k} \{\sigma_{kl}^{k} \left[\Delta y_k^{\infty}\left(\mathrm{Mid}\left(k, l\right), \mathrm{Pre}\left(k\right)\right) - \Delta y_l^{\infty}\left(\mathrm{Mid}\left(k, l\right), \mathrm{Pre}\left(k\right)\right)\right] \\
& + \sigma_{kl}^{l} \left[\Delta y_l^{\infty}\left(\mathrm{Post}\left(l\right), \mathrm{Mid}\left(k, l\right)\right) - \Delta y_k^{\infty}\left(\mathrm{Post}\left(l\right), \mathrm{Mid}\left(k, l\right)\right)\right]\}
\end{aligned}
$$

  - 接下来假设平行趋势满足，从而$\mathrm{VWCT} = 0$。
- 第三项：处理效应异质性：

$$
\Delta \mathrm{ATT} = \sum_{k \neq \infty} \sum_{l > k} \sigma_{kl}^{l} \left[\mathrm{ATT}_k\left(Post\left(l\right)\right) - \mathrm{ATT}_k\left(Mid\left(k, l\right)\right)\right]
$$

识别：个体异质性但无时间异质性的处理效应

- 第一种情况：个体有异质性，但没有时间异质性，从而$\text{ATT}_k(\mathcal{T}) = \text{ATT}_k$
  - 此时$\Delta\text{ATT} = 0$
  - 识别了：

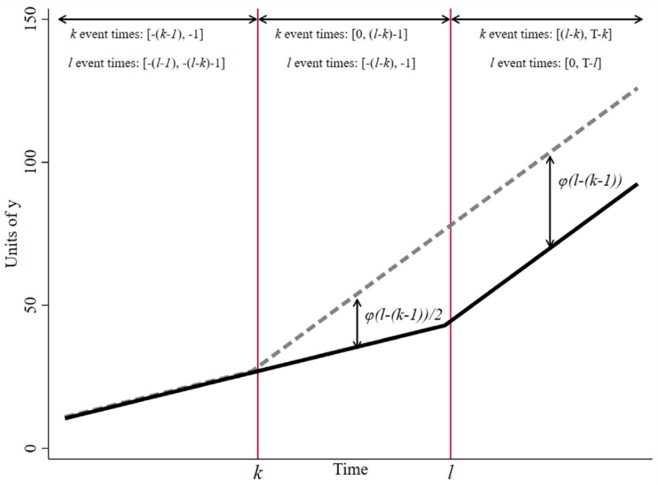
$$\text{VWATT} = \sum_{k \neq \infty} \text{ATT}_k \overbrace{\left[\sigma_{k\infty} + \sum_{j=1}^{k-1} \sigma_{jk}^{k} + \sum_{j=k+1}^{K} \sigma_{kj}^{k}\right]}^{\omega_k^T}$$

  - 一般$\omega_k^T \neq n_k$，从而没有识别ATT。与ATT的差距取决于：
    - 处理效应异质性的大小
    - 处理的timing
    - 研究权重ω_k^T或者直接加权（如Sun和Abraham，2021等）

识别：个体同质性、时间异质性的处理效应

- 第二种情况：个体处理效应同质，但是在时间上有异质性
  - 由于不同的$\beta^{2\times 2}$都在时间上计算平均，所以也会有异质性，从而VWATT与ATT不相等
  - $\Delta \text{ATT} \neq 0$!
  - 由于ΔATT前面的系数是负的，可能导致估计出的符号都是错的!

时间异质性的处理效应



150
100
50
0
Units of y
k event times: [-(k-1), -1]
l event times: [-(l-1), -(l-k)-1]
k event times: [0, (l-k)-1]
l event times: [-(l-k), -1]
k event times: [(l-k), T-k]
l event times: [0, T-l]
φ(l-(k-1))
φ(l-(k-1))/2
k
Time
l

动态设定

- 动态设定:

$$y_{it}=\alpha_i+\lambda_t+\sum_{\ell=-p,\ell\neq -1}^{q}\tau_l\cdot \mathbb{1}\left\{t=t_i+\ell\right\}+u_{it}$$

- Sun和Abraham（2021）对这一设定进行了分析，同样发现这一设定在异质性存在的情况下识别可能是有问题的
- 定义cohort的ATT（CATT）：

$$\mathrm{CATT}_{k,\ell}=\mathbb{E}\left[y_{i,k+\ell}(k)-y_{i,k+\ell}(0)\,|t_i=k\right]$$

动态设定的潜在假定

平行趋势假定

对于$s \neq t$，对于所有的w，有$\mathbb{E}\left[y_{it}(\infty) - y_{is}(\infty) \mid t_i = k\right]$都相等。

- 实证中：$w = \infty$，即控制组（never treated）可能不满足这一假定，那么需要控制组排除
- Goodman-Bacon：加权版本的平行趋势，比这里的条件弱

动态设定的潜在假定

处理前不可预见假设（no anticipatory behavior prior to treatment）

在处理之前没有处理效应，即对于所有的k和$\ell < 0$，
有$\mathbb{E}\left[y_{i,k+\ell}(k) - y_{i,k+\ell}(\infty) \,|\, t_i = k\right] = 0$

- 不会根据对未来的预见调整行为
- 最好是个体根本不能预见整个的路径
- Goodman-Bacon：假定如果$t < t_i$，那么$y_{it}(t_i) = y_{it}(\infty)$，更强

动态设定的潜在假定

处理效应相对时间同质性

对于每个ℓ，$\mathrm{CATT}_{k,\ell}$与k无关，从而$\mathrm{CATT}_{k,\ell} = \mathrm{ATT}_{\ell}$。

- 每个个体都有相同的处理效应路径（而不管是哪个cohort）

静态与动态设定

- 令$w_{it}^{\ell} = \mathbb{1}\{t = t_i + \ell\}$，从而

$$w_{it} = \sum_{\ell \geq 0} w_{it}^{\ell}$$

- 那么静态设定为：

$$y_{it} = \alpha_i + \lambda_t + \tau \sum_{\ell \geq 0} w_{it}^{\ell} + u_{it}$$

- 动态设定为：

$$y_{it} = \alpha_i + \lambda_t + \sum_{\ell=-P}^{-2} \tau_\ell w_{it}^{\ell} + \sum_{\ell=0}^{L} \tau_\ell w_{it}^{\ell} + u_{it}$$

- 完全动态设定，考虑事前超过P期、事后超过L期的情况：

$$y_{it} = \alpha_i + \lambda_t + \tau_{-Pb} \sum_{\ell < -P} w_{it}^{\ell} + \sum_{\ell=-P}^{-2} \tau_\ell w_{it}^{\ell} + \sum_{\ell=0}^{L} \tau_\ell w_{it}^{\ell} + \tau_{Lf} \sum_{\ell > L} w_{it}^{\ell} + u_{it}$$

统一设定

将以上设定统一为

$$y_{it} = \alpha_i + \lambda_t + \sum_{g \in \mathcal{G}} \tau_g \cdot \mathbb{1}\{t - t_i \in g\} + u_{it}$$

其中需要有一个组被排除，记为$g^{\text{excl}} = \{\ell : \ell \notin \mathcal{G}\}$。

- 静态设定：$g = [0, T], g^{\text{excl}} = [-T, -1]$
- 动态设定：$g^{\text{excl}} = \{-T, ..., -K-1, -1, L+1, ..., T\}$
- 完全动态：$g^{\text{excl}} = \{-1\}$

识别

回归系数的识别

使用如上设定，系数识别了：

$$\tau_g = \sum_{\ell\in g}\sum_{k\neq\infty}\omega^g_{k,\ell}\left[\mathbb{E}\left(y_{i,k+\ell}-y_{i,0}(\infty)\,|t_i=k\right)-\mathbb{E}\left(y_{i,k+\ell}(\infty)-y_{i,0}(\infty)\right)\right]$$
$$+\sum_{g'\neq g,g'\in\mathcal{G}}\sum_{\ell\in g'}\sum_{k\neq\infty}\omega^g_{k,\ell}\left[\mathbb{E}\left(y_{i,k+\ell}-y_{i,0}(\infty)\,|t_i=k\right)-\mathbb{E}\left(y_{i,k+\ell}(\infty)-y_{i,0}(\infty)\right)\right]$$
$$+\sum_{\ell\in g^{excl}}\sum_{k\neq\infty}\omega^g_{k,\ell}\left[\mathbb{E}\left(y_{i,k+\ell}-y_{i,0}(\infty)\,|t_i=k\right)-\mathbb{E}\left(y_{i,k+\ell}(\infty)-y_{i,0}(\infty)\right)\right]$$
$$+\sum_{t}\omega^g_{\infty,t}\left[\mathbb{E}\left(y_{it}-y_{i,0}(\infty)\,|t_i=\infty\right)-\mathbb{E}\left(y_{it}(\infty)-y_{i,0}(\infty)\right)\right]$$

- τ_g不仅仅含有$\ell\in g$的部分，还有$\ell\in\mathcal{G}-g$的部分以及$\ell\in g^{excl}$的部分！
- $\omega^g_{w,\ell}$可以通过估计：

$$w^\ell_{it}\cdot\mathbb{1}\{t_i=k\}=\alpha_i+\lambda_t+\sum\omega^g_{w,\ell}\cdot\mathbb{1}\{t-t_i\in g\}+v_{it}$$

权重

- 对于$\ell \in g$，相应的权重：

$$\sum_{\ell \in g} \sum_{k} \omega_{k,\ell}^{g} = 1$$

- 对于$\ell \in g', \ell \in \mathcal{G} - g$，权重：

$$\sum_{\ell \in g'} \sum_{k} \omega_{k,\ell}^{g} = 0$$

- 对于排除组，权重：

$$\sum_{\ell \in g^{excl}} \sum_{k} \omega_{k,\ell}^{g} = -1$$

识别+共同趋势假设

- 如果假设共同趋势，那么以上所有的

$$
\begin{aligned}
&\mathbb{E}\left(y_{i,k+\ell}-y_{i,0}(\infty)\,|t_i=k\right)-\mathbb{E}\left(y_{i,k+\ell}(\infty)-y_{i,0}(\infty)\right)\\
=&\mathbb{E}\left(y_{i,k+\ell}-y_{i,0}(\infty)\,|t_i=k\right)-\mathbb{E}\left(y_{i,k+\ell}(\infty)\,|t_i=k\right)+\mathbb{E}\left(y_{i,k+\ell}(\infty)\,|t_i=k\right)\\
&-\mathbb{E}\left(y_{i,k+\ell}(\infty)-y_{i,0}(\infty)\right)\\
=&\mathrm{CATT}_{w,\ell}+\mathbb{E}\left(y_{i,k+\ell}(\infty)-y_{i,0}(\infty)\,|t_i=k\right)-\mathbb{E}\left(y_{i,k+\ell}(\infty)-y_{i,0}(\infty)\right)\\
=&\mathrm{CATT}_{w,\ell}
\end{aligned}
$$

- 从而

$$
\begin{aligned}
\tau_g=&\sum_{\ell\in g}\sum_{k}\omega_{k,\ell}^{g}\mathrm{CATT}_{w,\ell}\\
&+\sum_{g'\neq g,g'\in\mathcal{G}}\sum_{\ell\in g'}\sum_{k}\omega_{k,\ell}^{g}\mathrm{CATT}_{w,\ell}\\
&+\sum_{\ell\in g^{excl}}\sum_{k\neq\infty}\omega_{k,\ell}^{g}\mathrm{CATT}_{w,\ell}
\end{aligned}
$$

识别+共同趋势假设+其他假设

- 如果额外假设无预见，那么对于$\ell < 0$，$\mathbb{E}\left(y_{i,k+\ell} - y_{i,k+\ell}(\infty) \mid t_i = k\right) = 0$，从而：

$$\tau_g = \sum_{\ell \in g, \ell \geq 0} \sum_k \omega^g_{k,\ell} \mathrm{CATT}_{w,\ell} + \sum_{g' \neq g, g' \in \mathcal{G}} \sum_{\ell \in g', \ell \geq 0} \sum_k \omega^g_{k,\ell} \mathrm{CATT}_{w,\ell} + \sum_{\ell \in g^{excl}, \ell \geq 0} \sum_{k \neq \infty} \omega^g_{k,\ell} \mathrm{CATT}_{w,\ell}$$

- 如果额外假设处理效应相对时间同质性假设：

$$\tau_g = \sum_{\ell \in g} \omega^g_\ell \mathrm{CATT}_\ell + \sum_{g' \neq g, g' \in \mathcal{G}} \sum_{\ell \in g'} \omega^g_\ell \mathrm{CATT}_\ell + \sum_{\ell \in g^{excl}} \omega^g_\ell \mathrm{CATT}_\ell$$

其中$\omega^g_\ell = \sum_k \omega^g_{k,\ell}$，可以通过估计：

$$w^\ell_{it} = \alpha_i + \lambda_t + \sum_{g \in \mathcal{G}} \omega^g_\ell \cdot \mathbb{1}\left\{t - t_i \in g\right\} + v_{it}$$

得到（eventstudyweights）

- 完全动态设定：

$$\tau_\ell = \mathrm{ATT}_\ell + \sum_{\ell' \in g^{excl}} \omega^\ell_{\ell'} \mathrm{ATT}_\ell$$

即使完全动态设定，得到的结果也会被排除组所污染（contaminated）

异质性处理效应

放弃处理效应的相对时间同质性假设，Sun和Abraham（2021）提出使用如下回归：

$$Y_{it} = \alpha_i + \lambda_t + \sum_{k \notin C} \sum_{\ell \neq -1} \delta_{k,\ell} \cdot \mathbb{1}\{t_i = k\} \cdot \mathbb{1}\{t = t_i + \ell\} + u_{it}$$

其中：

- $\mathbb{1}\{t_i = k\}$为cohort虚拟变量
- $\mathbb{1}\{t = t_i + \ell\}$为政策发生时间之后期数的虚拟变量
- C为控制组（从未受到处理的组），或者最晚受到处理的组

如此，$\delta_{k,\ell}$估计了$\mathrm{CATT}_{k,l}$。接下来，使用$\mathrm{CATT}_{k,l}$的加权平均：

$$\hat{v}_g = \frac{1}{|g|} \sum_{\ell \in g} \sum_k \hat{\delta}_{k,l} \times w_k$$

其中w_k为cohort k的比例。以上成为交互加权估计量（interaction-weighted estimator）

Callaway和San’t Anna（2021）

Callaway和San’t Anna（2021）使用了另外的处理方法，但是极其类似：

- 非参数识别（使用never group）：

$$\mathrm{ATT}(k,t;\delta)=\mathbb{E}\left[\left(\frac{w_{it}^{g}}{\mathbb{E}\left[w_{it}^{g}\right]}-\frac{\frac{p_k(X)C}{1-p_k(X)}}{\mathbb{E}\left[\frac{p_k(X)C}{1-p_k(X)}\right]}\right)(y_t-y_{g-\delta-1})\right]$$

- 可以扩展到doubly robust
- 命令：csdid

双重差分模型：扩展

其他方法:

1. Triple differences: e.g.，某个地区g的某个特定组别r受到处理：加入两两交互项，三个变量的交互项的系数为处理效应。
2. DID+Matching:
   1. 先匹配出子样本
   2. 先差分，再匹配
3. 非线性DID：Changes-in-Changes：common trend假设的满足取决于y的函数形式

CIC模型

非线性DID：CIC（Athey and Imbens, 2006）
假设：

1. $Y_i(0) = h(U_i, T_i)$，其中$h(\cdot,\cdot)$对U_i为单调的函数，T_i为时间
2. $U \perp\!\!\!\perp T|G$，即给定分组，U的分布与时间独立
3. support假设

关键结论：

$$F_{Y^N,11}(y) = F_{Y,10}\left(F^{-1}_{Y,00}\left(F_{Y,01}(y)\right)\right)$$

其中$F_{Y,gt}$为分布函数。进而处理效应：

$$\tau_{CIC} = \mathbb{E}(Y_{11}) - \mathbb{E}\left(F^{-1}_{01}\left(F_{00}(Y_{10})\right)\right)$$

合成控制法

- 在一些比较宏观的问题上，处理组个体非常少（甚至只有一个）——DID失效。
- 可能的解决办法：通过控制组个体的凸组合来构造反事实。
- 假设$i = 0, 1, ..., N$个个体，$i = 0$被处理，其他未被处理。
- 假设$T = 1, ..., T$，$i = 0$被处理的时间为T_0
- 通过因子模型理解：如果未经过处理的个体数据生成过程为：

$$Y_{it}(0) = F_t' \lambda_i + Z_i' \zeta_t + u_{it}$$

那么可以使用控制组的Y预测处理组的Y

合成控制法

- 令X_0为T_0之前的$i=0$个体的k个特征（如Y_1以及Z_1）；相应的，$X_{J\times k}$为T_0之前的其他J个个体的特征
- 为J个个体赋予不同权重：$w_1,...,w_J$，$w_1+\cdots+w_{J=1}$且$w_j\geq 0$（凸组合）
- 给定一个半正定矩阵V，选取权重：

$$w^*=\arg\min_{w}\left(X_0-X'w\right)'V\left(X_0-X'w\right)$$

  - 特殊情况，如果$X_0=[Y_{00},...,Y_{0,T_0-1}]$，那么即使用处理组$Y_0$对控制组$Y_j$做时间序列回归
- 则处理效应：

$$\tau_{0t}=Y_{0t}-\sum_{j=1}^{J}w_j^*Y_{jt}$$

合成控制法

- Stata：synth命令
- V的选取：交叉验证等方法都可以使用，保证预测准确
  - 通常：

$$V^{*}=\arg\min_{V}\sum_{t=1}^{T_0-1}\left(Y_{0t}-\sum_{j=1}^{J}w_jY_{jt}\right)^2$$

  - nested选项
- 推断：placebo test构建$p-value$
  - 假装未处理的个体被处理，估计政策效应。

HCW方法

- Hsiao, Ching and Wan(2012)提出，可以直接使用回归，而非构造凸组合的方法构建反事实。
- 如果假设在没有政策冲击的时候，未被处理的结果y由因子模型决定：

$$Y_{it}(0) = F_t'\lambda_i + \alpha_i + \epsilon_{it}$$

  其中F_t为K个不可观测的因子。
- 关键假设：控制组的个体对处理组没有影响。
- 可以通过构造null space的方法构造反事实。
- 实际中：
  1. 在政策发生之前，用Y_{0t}对$Y_t = \{Y_{it}, i = 1, ..., N\}$做时间序列回归
  2. 预测$Y_{0t}(1), t > T_0$
  3. 政策效应$\tau = Y_{0t} - Y_{0t}(1), t > T_0$

HCW方法中的Lasso

Li和Bell(2017)推荐使用Lasso方法作为HCW方法中的模型选择

- 实例：中欧班列开通对于陕西省出口的影响：hcw_lasso.do