

体育经济分析: 原理与应用

DiD

周正卿

04 November 2023

大纲

大纲

- Level 1
 - 一个例子
- Level 2
 - 基本概念
- Level 3
 - 具体实战

潜在结果框架帮助理解“真相”

因果识别方法

- 随着计量经济学“可信性革命”（credibility revolution）席卷经济学的各个领域，基于潜在因果模型的因果效应识别策略，如匹配法（matching）、工具变量法（instrumental variable）、双重差分法（difference-in-differences）和断点回归设计（regression discontinuity design）等，逐渐成为了经济学等社会科学领域实证研究的通行研究范式。

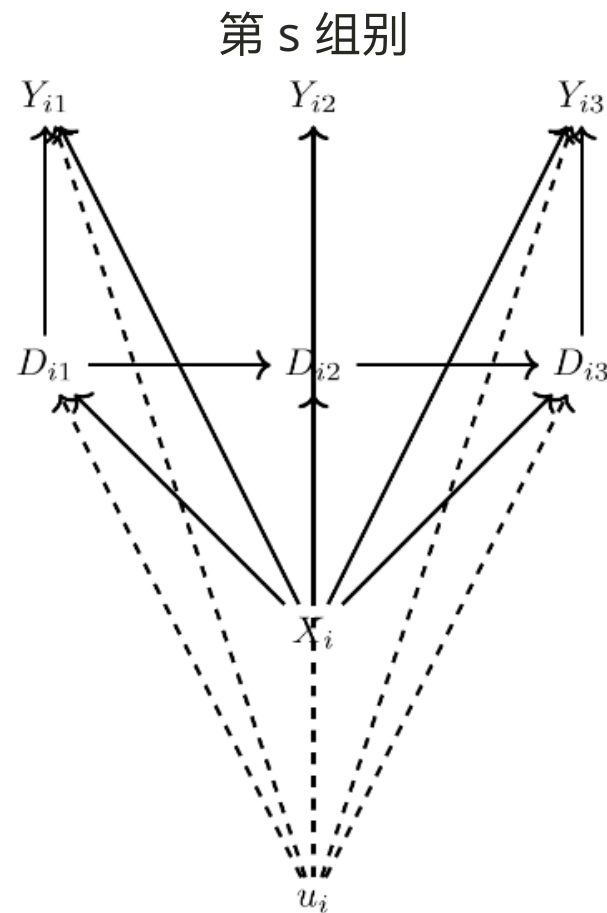
面板数据框架

面板数据框架

- 因果推断工具包中最重要的工具之一就是**面板数据估计量** (panel data estimator)
 - 该估计量为**纵贯数据** (longitudinal data) 专门设计
 - 不同时间点上对个体的重复观测
- 面板数据的优点:
 1. 更多的变化 (横截面数据和时间序列变化)
 2. 可以处理**特定类型数据带来的遗漏变量偏差** (不可观测的且不随时间推移而变化的因素)
- 示例:
 1. 个体 (球员) i 在年份 t 的收入
 2. 公司 (球队) i 在年份 t 的业绩
 3. 国家 i 在年份 t 的GDP
 4. 学校 i 在年份 t 的升学率

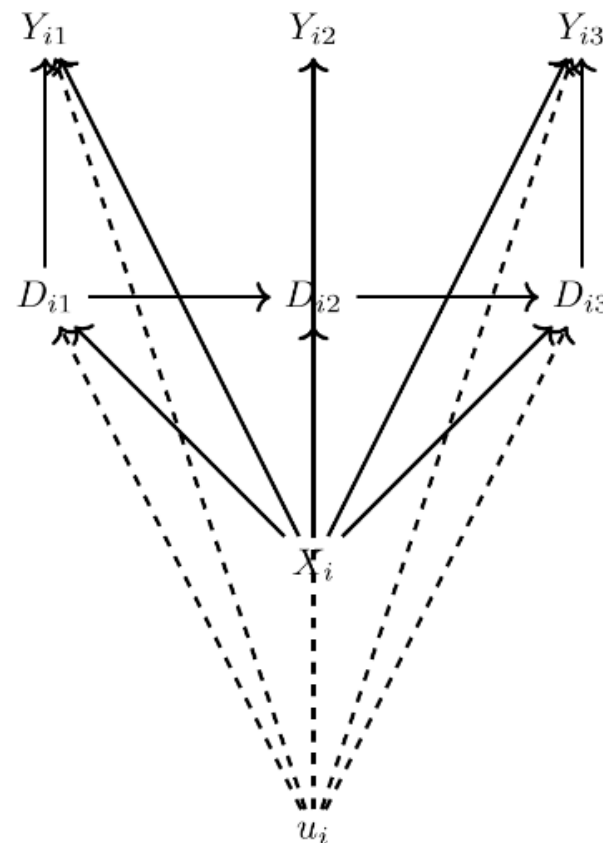
纵贯数据的有向无环图 (Imai 和 Kim, 2017)

- 结果变量 Y_i
 - 包含 3 期, 分别是 Y_{i1}, Y_{i2}, Y_{i3}
- 协变量矩阵 D_i 随时间推移而变化 D_{i1}, D_{i2}, D_{i3}
- 随个体变化但不随时间变化, 可被观测的 X_i
- 随个体变化但不随时间变化, 但无法被观测到的 u_i
 - 无法观测
 - 随个体而变动
 - 给定个体 i , 不随时间推移而变化



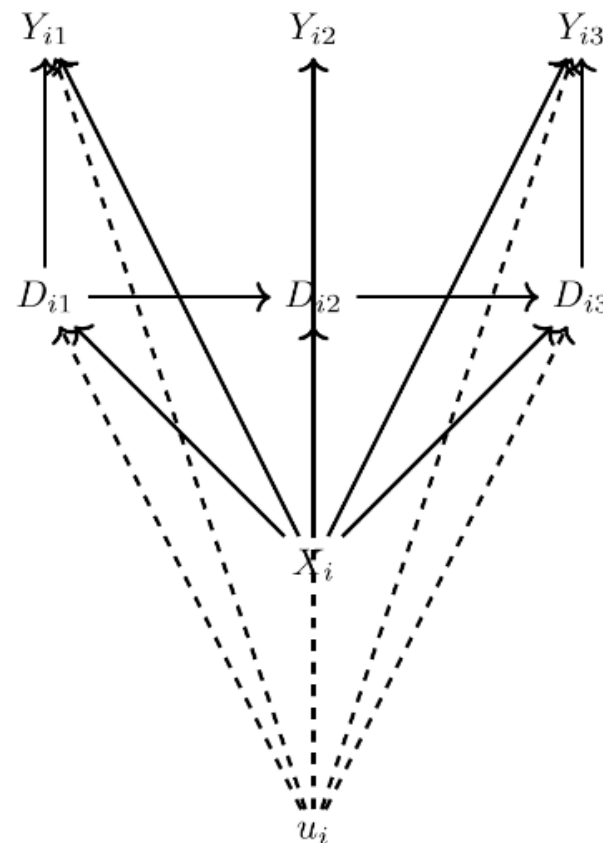
纵贯数据的有向无环图 (Imai 和 Kim, 2017)

- 第1层: Y_{i1} 与 Y_{i2} 和 Y_{i3} 相互不影响
- 第2层: D_{i1} 影响第1层 Y_{i1} 和同层 D_{i2}
- 第3层: X_i 影响第2层 $D_{i1} - D_{i3}$ 和第1层 $Y_{i1} - Y_{i3}$
- 第4层: 不可观测的混杂因子 u_i 影响第3层 X_i 和 第2层 $D_{i1} - D_{i3}$ 以及第1层 $Y_{i1} - Y_{i3}$
- 处理变量 D 产生了**内生性**问题 → 直接估计存在偏误
- **注意几点**
 - **不存在**与第2层 D_{it} 相关的、不能被观测到且**随时间的推移而变化的其他混杂因素**
 - 第1层**过去结果** Y_{it-1} **不直接影响当前结果** Y_{it}
 - 1→2**过去结果** Y_{it-1} **不直接影响当前处理变量** D_{it}
 - 2→1**过去处理变量** D_{it-1} **不直接影响当前结果** Y_{it}



纵贯数据的有向无环图 (Imai 和 Kim, 2017)

- 基于数据的特殊结构和假设：可以使用**固定效应** (fixed effects) 的特定面板方法来分离 D 对 Y 的因果效应
- 若固定效应中包含**年份固定效应**，该估计量通常称为**双向固定效应估计量** (Two-Way Fixed Effects Estimator, 简称TWFE)
- 回归方程为： $Y_{it} = \alpha_0 + \delta D_{it} + \alpha_i + \alpha_t + X_{it} + \varepsilon_{it}$
 - $u_i = \alpha_i + \alpha_t$
 - u_i 称**未观测到的异质性**(unobserved heterogeneity)
 - ε_{it} 称**特异性误差** (idiosyncratic error)



- 在 Y_{it} 对 D_{it} 回归时会发生什么？

估计方法 1：合并普通最小二乘 (pooled OLS)

- 考虑**综合误差项** $\eta_{it} \equiv c_i + \varepsilon_{it}$
 - 这里的 c_i **视具体研究情境**而定。若以双向固定效应模型视角，则写作 $c_i = \alpha_i + \alpha_t$
- 则TWFE 可以改写为 $Y_{it} = \delta D_{it} + \eta_{it}; \quad t = 1, 2, \dots, T$
- 如果

$$E[\eta_{it} \mid D_{i1}, D_{i2}, \dots, D_{iT}] = E[\eta_{it} \mid D_{it}] = 0 \quad \text{对于所有的 } t = 1, 2, \dots, T$$

用 pooled OLS 估计该模型可以获得一致的 δ 估计值

- 这相当假设**未观测到的异质性任何时期的 D_{it} 都不相关**，但前面的 DAG 并不支持该假设 → 存在**内生性问题** → **遗漏变量偏差**
- **序列相关问题**： u_i 在每个时期的 η_{it} 中，序列相关问题导致**异方差稳健标准误偏小** → **系数不显著**

估计方法 2：固定效应（组内估计量）[fixed effects (within estimator)]

- 假设没有年份固定效应，因此 TWFE 变为 $Y_{it} = \delta D_{it} + u_i + \varepsilon_{it}$
- 对该式进行**时间中心化**处理

$$\begin{aligned}Y_{it} &= \delta D_{it} + u_i + \varepsilon_{it} \\ \bar{Y}_i &= \delta \bar{D}_i + u_i + \bar{\varepsilon}_i \\ (Y_{it} - \bar{Y}_i) &= (\delta D_{it} - \delta \bar{D}_i) + (u_i - u_i) + (\varepsilon_{it} - \bar{\varepsilon}_i) \\ \ddot{Y}_{it} &= \delta \ddot{D}_{it} + \ddot{\varepsilon}_{it}\end{aligned}$$

- 注意：
 - “组内”估计量=“固定效应”估计量=“双向固定效应”估计量（当含年份固定效应时）
 - 由于每个作者的遵循的命名传统不同，若文献中出现时，要识别的方程具体形式
- 操作：
 - 对数据时间中心化处理后，进行 \ddot{Y}_{it} 对 \ddot{D}_{it} 的回归（需要矫正自由度）
 - 对原始方程添加个体虚拟变量来估计，**最小二乘虚拟变量 (LSDV) 估计量**
 - 在 STATA 固定效应模型中选择 fe 选项

识别假设1: 所有“右手变量”严格外生

- $E[\varepsilon_{it} \mid D_{i1}, D_{i2}, \dots, D_{iT}, u_i] = 0$
- 比一般OLS中的假设更宽松, 允许 D_{it} 与 u_i 任意相关
 - 含义: 所有不可观测的影响因素中, 那些不随时间推移而变化的固定效应全部都可以被 u_i 捕获, 不存在遗漏。在这种情况下, 回归过程中只限定与 ε_{it} 的相关性, 不涉及与 u_i 的相关性

识别假设2: 无完全多重共线性与秩条件

- 考虑模型 $wage_{it} = \beta_0 + \beta_1 experience_{it} + \beta_2 male_i + \beta_3 white_i + u_i + \varepsilon_{it}$
 - 由于 $male_i$ 和 $white_i$, 存在多重共线性问题。
 - 不能一致性估计系数 β_2 和 β_3 , 因为这些时间恒定的不可观测因素固定效应 u_i 捕获
- $\text{rank}\left(\sum_{t=1}^T E\left[\ddot{D}_{it}' \ddot{D}_{it}\right]\right) = K$
 - 估计系数 $\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$ 总是一个缩放的 (scaled) 协方差, 而缩放要用到方差项
→ 进入回归的变量至少有部分的个体 i 是要随时间推移而变化的, 并且不共线, 才能确保 $\hat{\delta} \approx \delta$

推断与要点

- 该数据框架中的标准误必须以面板中的个体进行聚类，以允许同一个个体 i 的 ε_{it} 可以随时间相关
- 只要聚类的数量足够大，就可以产生有效推断
 - 经验法则：聚类数量要大于 30 个
 - 比如省级面板
- **要点 #1:** Fixed effects cannot solve reverse causality
 - 当存在反向因果关系时，DAG 就不再是前面提供的样子，会有 $Y \rightarrow D$ 的渠道出现
- **要点 #2:** Fixed effects cannot address time-variant unobserved heterogeneity
 - 变量中心化处理只是简单去除了一个未被观察到的随时间推移而变化的变量的均值
 - 当 $u_i \rightarrow u_{it}$ 时，中心化后的 \ddot{u}_{it} 仍然与 \ddot{D}_{it} 相关，所以后者仍然产生内生性

面板数据的总结

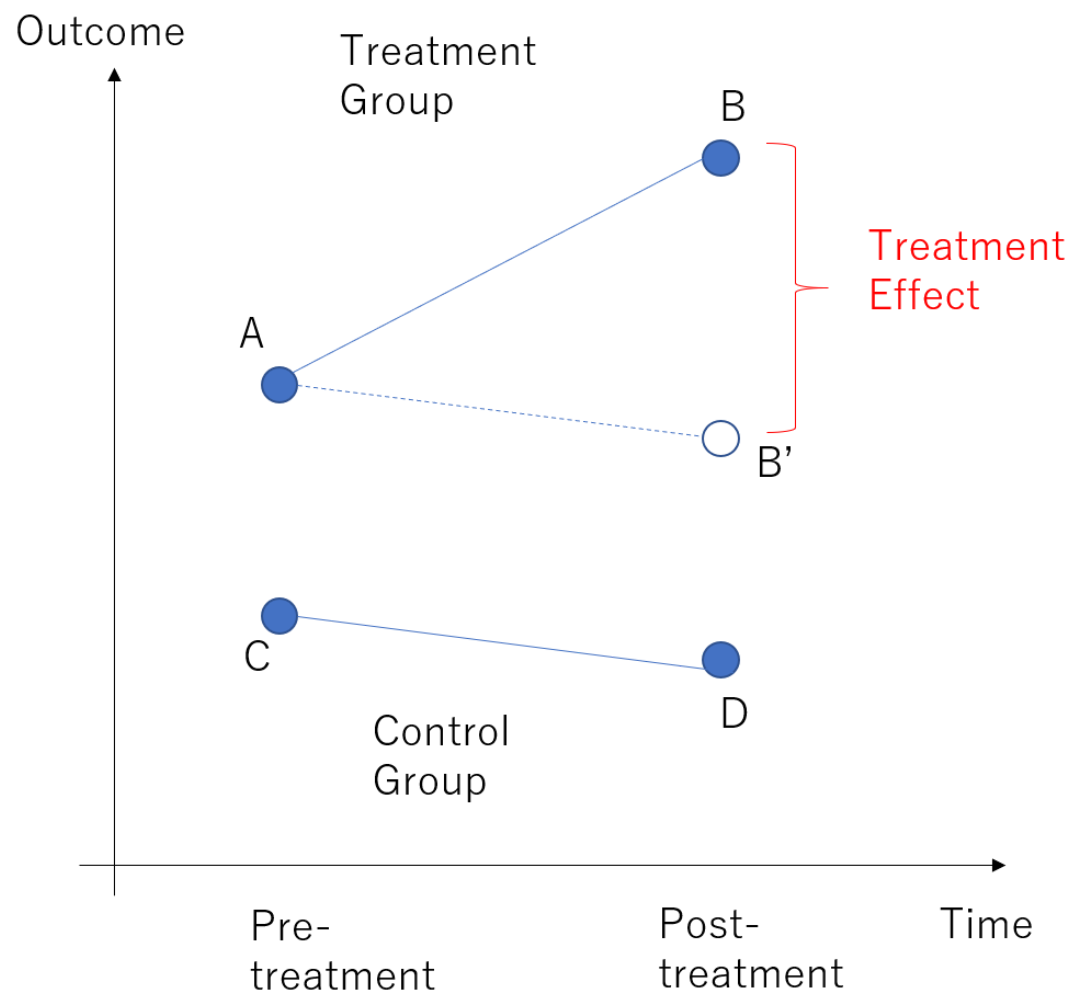
- 作用：
 - 固定效应（组内）能够消除所有与处理变量相关的**不随时间推移而变化的协变量**，无论该这些变量是否被观测到
 - 只要处理和结果随时间推移而变化，并且存在严格的外生性，那么固定效应（组内）估计量就能识别处理对结果的因果效应
- 限制：
 1. 不能处理**随时间推移而变化的未被观测到的异质性**
 2. 不能处理**强反向因果关系**
- 当存在上述2点限制时，固定效应（组内）不能估计一致因果效应，必须专项其他框架

DiD

介绍

- **差分法** (DiD) 是在非实验环境中估计因果效应的最欢迎的策略之一。
 - 在NBER WPs中使用率超过20%
- 近几年，关于DiD的计量经济学研究激增
- 出现时间比随机实验还要早大约 85 年
- 区分为经典模型（一组个体同时受处理）和常见模型（一组个体在不同时间接受处理）

DiD 图示



经典 2×2 DiD 模型

- Goodman-Bacon(2019)
 - 处理组和控制组各有2个时期:
 - 处理组有一个前期 $\text{pre}(k)$ 和一个后期 $\text{post}(k)$
 - 控制组有一个前期 $\text{pre}(U)$ 和一个后期 $\text{post}(U)$

经典 2×2 DiD 模型

$$\begin{aligned}
 \hat{\delta}_{kU}^{2 \times 2} (1) &= (\bar{y}_k^{\text{post}(k)} - \bar{y}_k^{\text{pre}(k)}) - (\bar{y}_U^{\text{post}(k)} - \bar{y}_U^{\text{pre}(k)}) \\
 (2) &= (E[Y_k \mid \text{Post}] - E[Y_k \mid \text{Pre}]) - (E[Y_U \mid \text{Post}] - E[Y_U \mid \text{Pre}]) \\
 (3) &= \underbrace{(E[Y_k^1 \mid \text{Post}] - E[Y_k^0 \mid \text{Pre}]) - (E[Y_U^0 \mid \text{Post}] - E[Y_U^0 \mid \text{Pre}])}_{\text{潜在结果转换方程}} \\
 &\quad + \underbrace{E[Y_k^0 \mid \text{Post}] - E[Y_k^0 \mid \text{Post}]}_{\text{加上一个“0”}} \\
 (4) &= \underbrace{E[Y_k^1 \mid \text{Post}] - \overbrace{E[Y_k^0 \mid \text{Post}]}^{\text{反事实}}}_{\text{ATT}} \\
 &\quad + \underbrace{[E[Y_k^0 \mid \text{Post}] - E[Y_k^0 \mid \text{Pre}]] - [E[Y_U^0 \mid \text{Post}] - E[Y_U^0 \mid \text{Pre}]]}_{2 \times 2 \text{ 情况下的非平行趋势偏差}}
 \end{aligned}$$

- 等号 (1) : $\hat{\delta}_{kU}^{2 \times 2}$ 为 k 组估计的 ATT, 右边第一项为处理组 k 处理后减去处理前的差值
- 等号 (2) : 重新写成条件期望形式

经典 2×2 DiD 模型

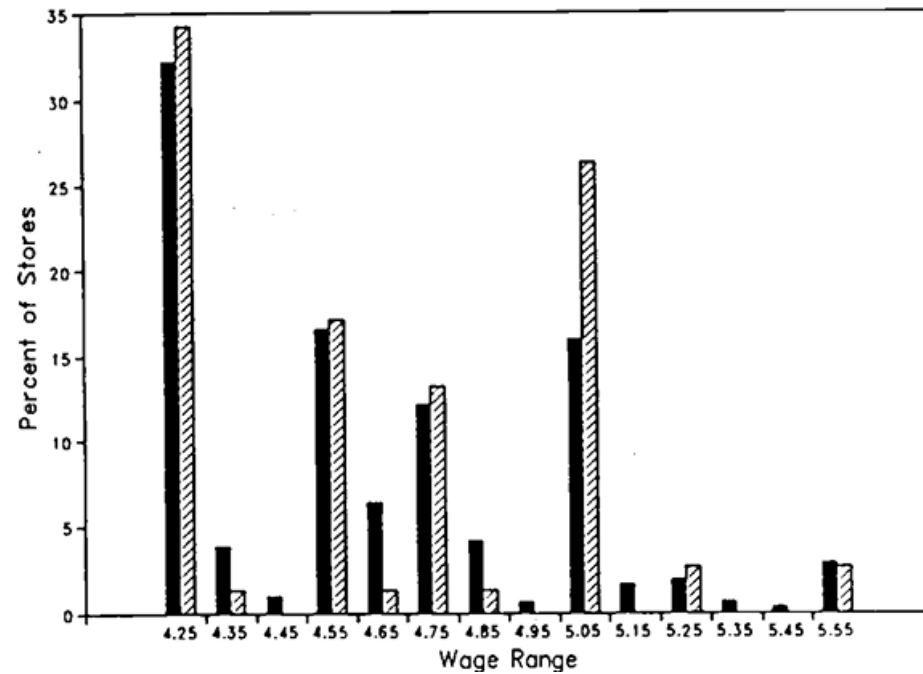
$$\begin{aligned}
 \hat{\delta}_{kU}^{2 \times 2} (1) &= (\bar{y}_k^{\text{post}(k)} - \bar{y}_k^{\text{pre}(k)}) - (\bar{y}_U^{\text{post}(k)} - \bar{y}_U^{\text{pre}(k)}) \\
 (2) &= (E[Y_k \mid \text{Post}] - E[Y_k \mid \text{Pre}]) - (E[Y_U \mid \text{Post}] - E[Y_U \mid \text{Pre}]) \\
 (3) &= \underbrace{(E[Y_k^1 \mid \text{Post}] - E[Y_k^0 \mid \text{Pre}]) - (E[Y_U^0 \mid \text{Post}] - E[Y_U^0 \mid \text{Pre}])}_{\text{潜在结果转换方程}} \\
 &\quad + \underbrace{E[Y_k^0 \mid \text{Post}] - E[Y_k^0 \mid \text{Post}]}_{\text{加上一个“0”}} \\
 (4) &= \underbrace{E[Y_k^1 \mid \text{Post}] - \overbrace{E[Y_k^0 \mid \text{Post}]}^{\text{反事实}}}_{\text{ATT}} \\
 &\quad + \underbrace{[E[Y_k^0 \mid \text{Post}] - E[Y_k^0 \mid \text{Pre}]] - [E[Y_U^0 \mid \text{Post}] - E[Y_U^0 \mid \text{Pre}]]}_{2 \times 2 \text{ 情况下的非平行趋势偏差}}
 \end{aligned}$$

- 等号 (3)：潜在结果转换方程讲 Y 的历史取值转换成潜在结果形式
- 等号 (4)：重新派丽，根据潜在结果的条件期望对 2×2 DiD 进行分解 → 第二项包含两个差值，第一个差值（涉及处理组 k，是反事实），第二个差值（涉及控制组）

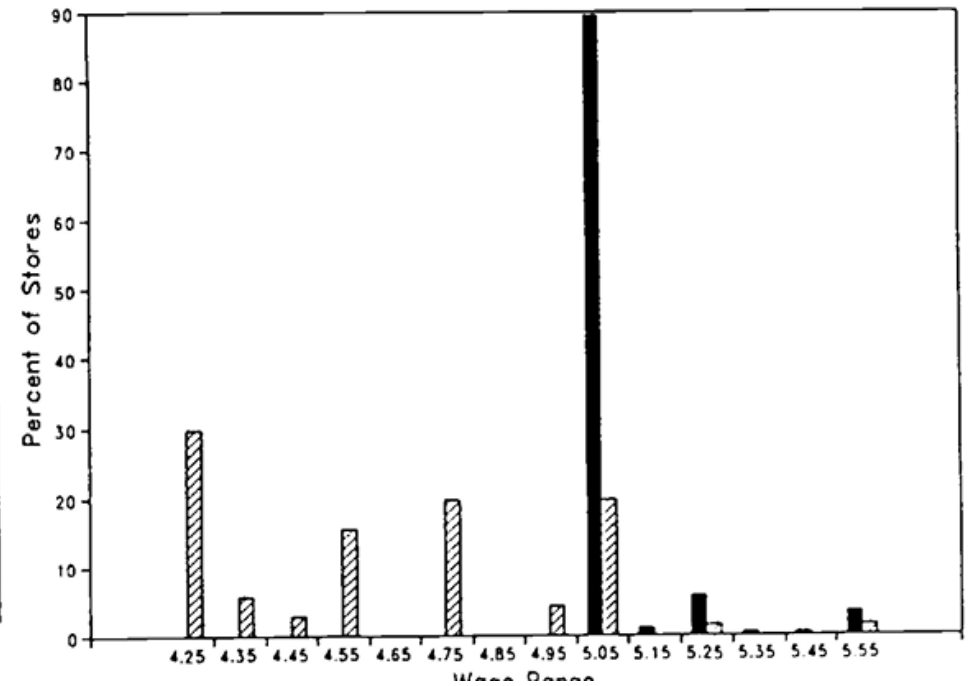
应用: Card and Kruger (1994, AER)

- 问题：最低工资增加对就业的影响
- 竞争性的理论
 - 在完全竞争市场中，最低工资提高会使得向右下方倾斜的需求曲线左移 → 就业率下降
 - 垄断市场中，最低工资提高导致就业增加
- 政策背景
 - 1992年 11 月，新泽西州的最低工资从4.25美元上涨到5.05美元
 - 相邻的宾夕法尼亚州最低工资仍然保持在 4.25 美元
- 比较新泽西州（处理组）和宾夕法尼亚东部（控制组）的快餐店在涨薪前后的情况

February 1992



November 1992



估计方法 1: 样本均值之差

Variable	Stores by state		
	PA (i)	NJ (ii)	Difference, NJ – PA (iii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	– 2.89 (1.44)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	– 0.14 (1.07)
3. Change in mean FTE employment	– 2.16 (1.25)	0.59 (0.54)	2.76 (1.36)

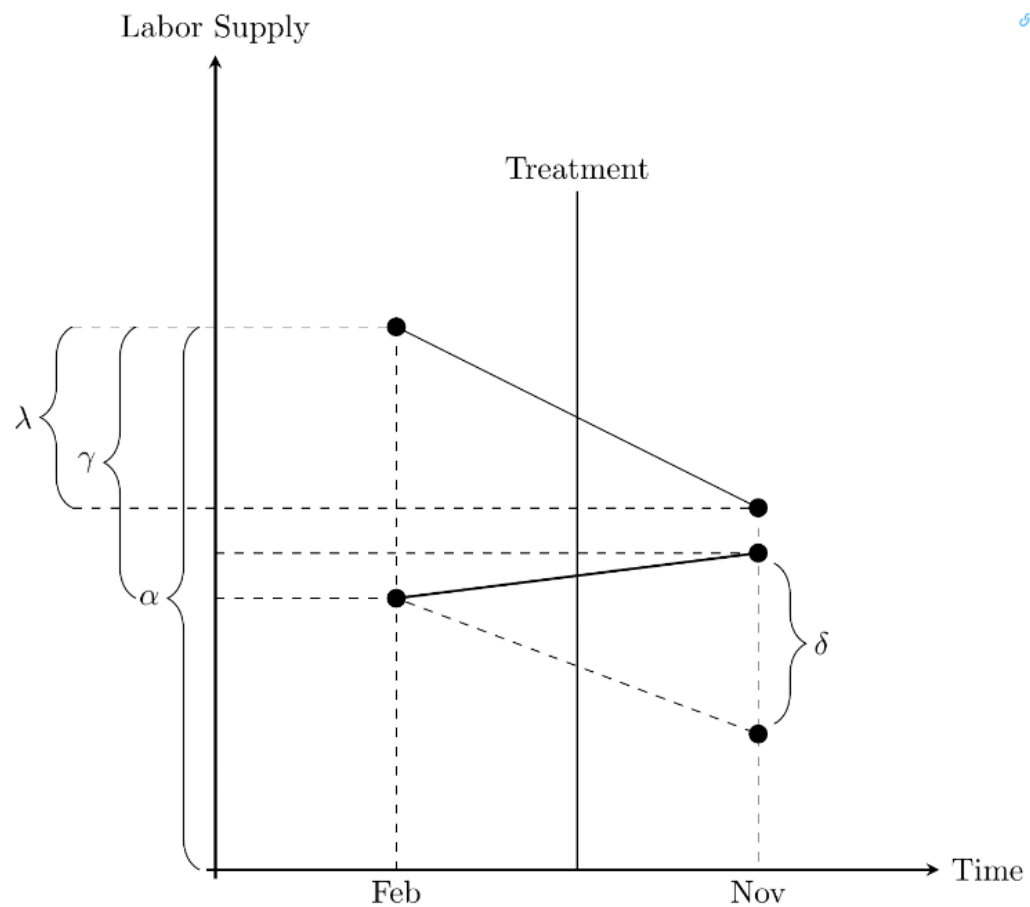
a	PA	NJ	NJ-PA
Pre	$E[Y_{PA}^0 \text{Pre}]$	$E[Y_{NJ}^0 \text{Pre}]$	$E[Y_{NJ}^0 \text{Pre}] - E[Y_{PA}^0 \text{Pre}]$
Post	$E[Y_{PA}^0 \text{Post}]$	$E[Y_{NJ}^1 \text{Post}]$	$E[Y_{NJ}^1 \text{Post}] - E[Y_{PA}^0 \text{Post}]$
变动均值 ?		?	作差后, 为什么要强调满足平行趋势假设?

方法设计：样本均值之差

- **共同趋势假设**：在没有事件影响的情况下，处理组和控制组在事件发生前后的平均潜在结果变化程度是相同的
 - 意味着其他因素（非事件造成）对处理组和控制组在处理时期前后Y变化的影响是相同
 - 也意味着，在事件没有发生或政策没有实施的情况下，处理组和控制组的Y有相同的时间趋势

$$\hat{\delta}_{NJ,PA}^{2 \times 2} = \underbrace{E[Y_{NJ}^1 | \text{Post}] - E[Y_{NJ}^0 | \text{Post}]}_{\text{ATT}} + \underbrace{[E[Y_{NJ}^0 | \text{Post}] - E[Y_{NJ}^0 | \text{Pre}]] - [E[Y_{PA}^0 | \text{Post}] - E[Y_{PA}^0 | \text{Pre}]]}_{\text{不满足共同趋势假设产生的偏误}}$$

DiD回归的理论图示



$$\delta = E \left[Y_{NJ, \text{Post}}^1 \right] - E \left[Y_{NJ, \text{Post}}^0 \right]$$

实际估计: 线性模型估计

- 除了满足**共同趋势假设**外，要获得 δ_{ATT} 的一致且有效估计量，还需要：
 - 控制**随时间推移而变化的内生协变量**，避免**遗漏变量偏差**
 - 通过控制适当的协变量，**减小残差方差，提高估计精度**
- 假设**州固定效应和时间固定效应为常数**，将 2×2 DiD 以线性回归方程来表达

$$Y_{its} = \alpha + \gamma NJ_s + \lambda D_t + \delta(NJ \times D)_{st} + \varepsilon_{its}$$

- NJ : 观测值来自 NJ 为 1
- D : 观测值来自 11 月 (Post) 为 1

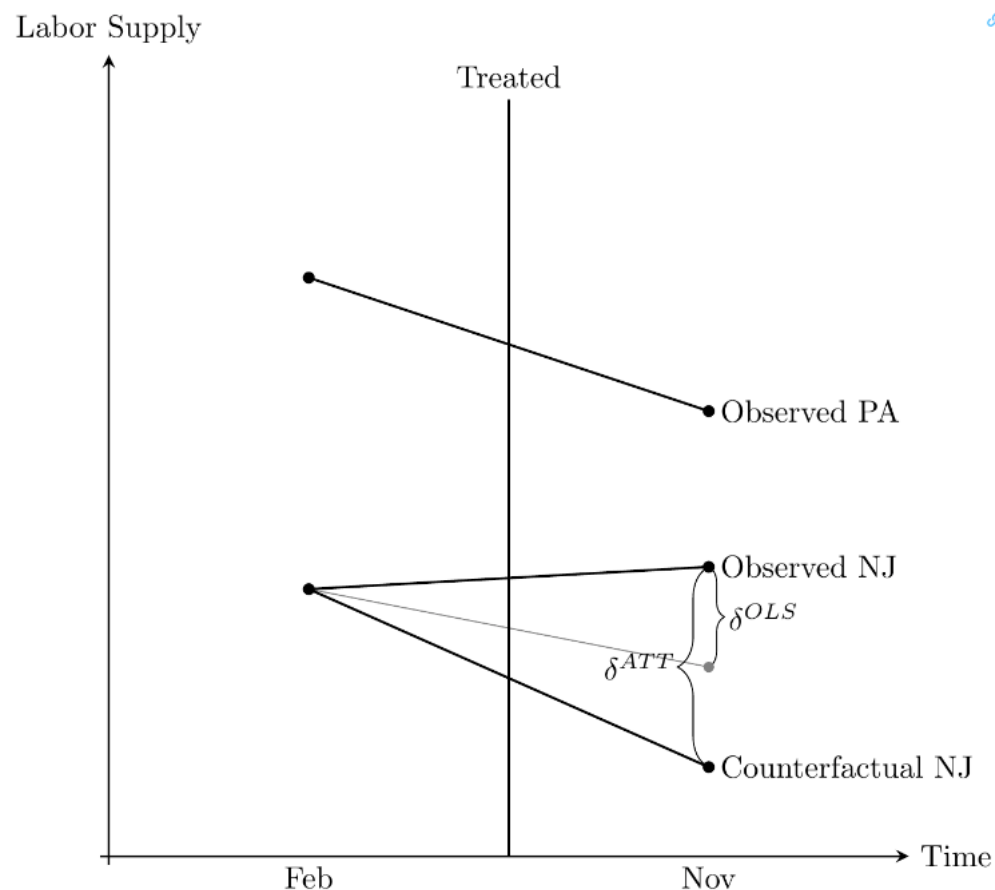
1. PA Pre: α

2. PA Post: $\alpha + \lambda$

3. NJ Pre: $\alpha + \gamma$

4. NJ Post: $\alpha + \gamma + \lambda + \delta$

共同趋势假设对于 $\hat{\delta}_{OLS}$ 去一致性估计 δ_{ATT} 非常重要



- OLS总是会估计出一个 δ 的值，关键是这个 $\hat{\delta}_{OLS}$ 是不是能够反映NJ 的反事实斜率

追求估计量的有效性（逻辑上没有通）

- CK1994 使用了 1 期处理前和 1 期处理后时期的数据
 - 对估计量有效性的干扰：每个州的就业存在序列相关
- Bertrand(2004)指出一般标准误通常低估估计量的标准差，估计量的标准误变得“过小”，从而导致会过度拒绝原假设，原假设本不应该拒绝实际中却拒绝了，因此建议使用以下方法：
 - 块状自助法标准误（block bootstrapping）：选择的“块”是州
 - 将数据聚合（aggregation）为 1 个前期和 1 个后期：完全忽略时间维度，使得数据只有前后两期和一个控制组
 - 若存在差分的时间序列，需要部分分离出省级和年份的固定效应，然后将分析转化为包含残差化的分析
 - 在组层面聚类（clustering）标准误：CK1994中群组数量只有2个

对共同趋势假设的讨论

诊断共同趋势假设

- 处理组个体的 Y_{it} 在没有接受处理的状态下拥有和控制组个体 Y_{it} 相同的时间变动趋势（包括事前事后两个阶段）
- 共同趋势假设在以下情况下会违背
- 当处理状态受**随时间推移而变化的因素**（time-varying factors）影响时，时间维度T 就成为了影响处理组和控制组的混杂因素，通常与处理组和控制组分别作用后，对结果造成**异质性的时间趋势**，造成估计的失效
- 即便在回归方程中控制时间维度，也不能得到一致的估计量
- **回忆**：DiD 只能处理**不随时间推移而变化的因素**（time-invariant factors）造成的影响
 - 例如人们参加工人培训计划是因为他们预期在进入该计划之前未来收入会减少

诊断共同趋势假设的逻辑Bug

$$\hat{\delta}_{kU}^{2 \times 2} = \underbrace{E[Y_k^1 | \text{Post}] - E[Y_k^0 | \text{Post}]}_{\text{ATT}} + \underbrace{[E[Y_k^0 | \text{Post}] - E[Y_k^0 | \text{Pre}]] - [E[Y_U^0 | \text{Post}] - E[Y_U^0 | \text{Pre}]]}_{\text{不满足共同趋势假设导致的偏差}}$$

- 第二行中 $E[Y_k^0 | \text{Post}]$ 是反事实项，是无法直接估计的，导致**共同趋势假设本质上是无法直接检验**
- 退而求其次，通过比较可观察的处理组和控制组**预处理前 DiD 系数**是否显著差异间接地检验平行趋势假设
- 如果处理组和控制组的事前趋势平行，那么有一定的信心认为事后趋势也是平行的
 - 若处理是内生的，显然会违背CT假设 → 两组本身有显著差异 → DiD 内生性问题必须要讨论的
 - 文章中不能出现：“如果处理发生前的趋势平行，则DID的共同趋势假设得到满足”

诊断 1: 检查处理组和控制组在处理前的平衡

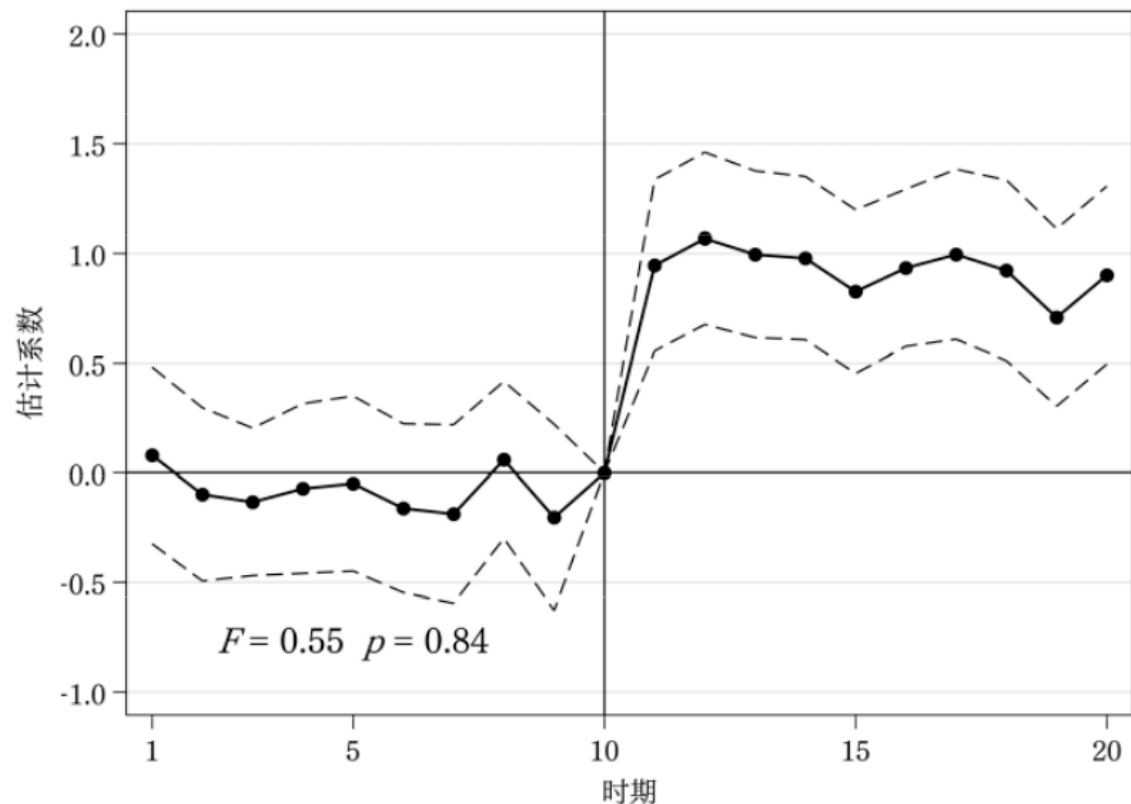
- 通常使用的是**事件研究**：在事前时期内，让处理效应在两组之间以及事件发生先后具有时变可比性
- **图示法非常重要**：目测个体接受处理的状态，是否同一时间？是否中间有退出？
- 如果接受处理的时间不统一，就需要额外注意（稍后讲解）
 - 该方法缺点是：1.处理组数量大，难实现;2.不美观;3.必须假设存在**控制组是从未被处理的组**
→ Goodman-Bacon (2019) **任何多期DiD的ATT 都可以看成是处理与未处理、早期处理和晚期处理、晚期处理和早期处理之间的比较，并加以综合**

事件分析的模型设定(稍后详解)

- 在 DiD 模型中包含**先期**和**滞后**，可以检查时点后处理的时变过程

$$Y_{its} = \gamma_s + \lambda_t + \sum_{\tau=-q}^{-1} \gamma_{\tau} D_{s\tau} + \sum_{\tau=0}^m \delta_{\tau} D_{s\tau} + X_{ist} + \varepsilon_{ist}$$

- 处理发生在 **第 0 年**
- 处理之前的 q 期 (**先期效应**) ; m 期之后 (**处理后效应**)
- 需要强调的是, **事前平行趋势通过检验并不意味着共同趋势假设一定成立**



- 先期效应直观地理解为：在处理发生前，Y在两组之间的差异**相对于基期**的Y在两组之间的差异
 - 基期通常选择是 T-1 期
- 各时期的先期效应不显著 + 联合检验结果无法拒绝处理前系数都为0的原假设
- 事件分析法的优势是能观察到处理效应的时变变化 → 之后详细讨论

应用：《平价医疗法案》

- 背景：？？
- Mill等（2019）审视了《平价医疗法案》下医疗补助计划扩张对人口死亡率的影响
- 结论：《法案》应用于多个州，年死亡率下降了 0.13 个百分点，比样本均值下降了 9.3%；并且随着时间的推移，这种影响会越来越大
- 分析过程：构建了 4 个事件分析研究
 - 对关键估计值以及结果和安慰剂检验
 - 事件研究展现的图示非常具有说服力

应用：《平价医疗法案》

Fig 1 :法案对获得医疗补助资格的人数影响

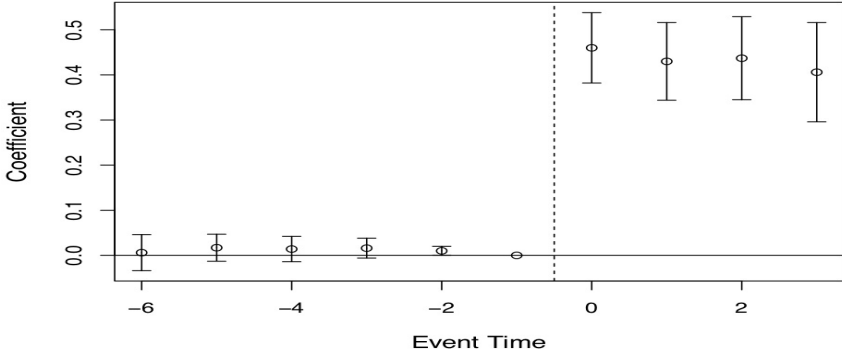


Fig 2 :法案对覆盖范围的影响

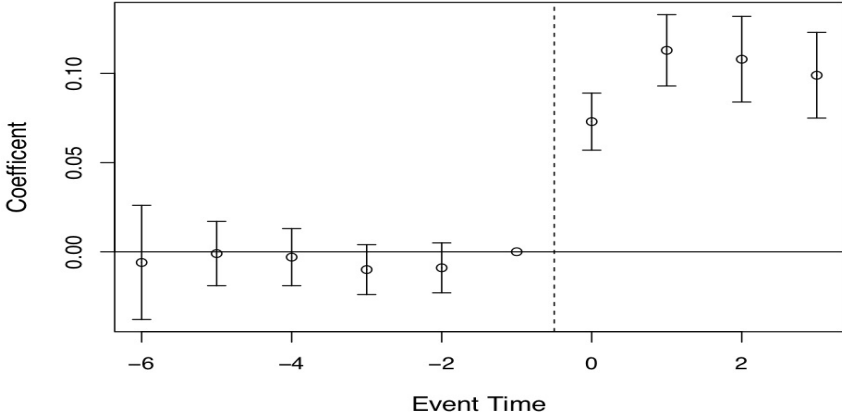


Fig 3 :法案对未参保率的影响

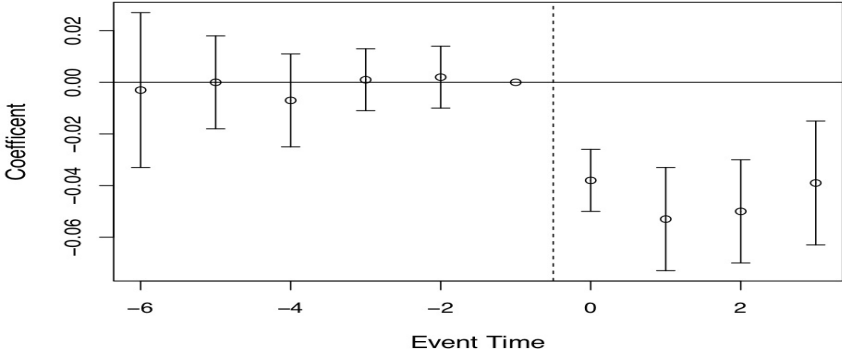
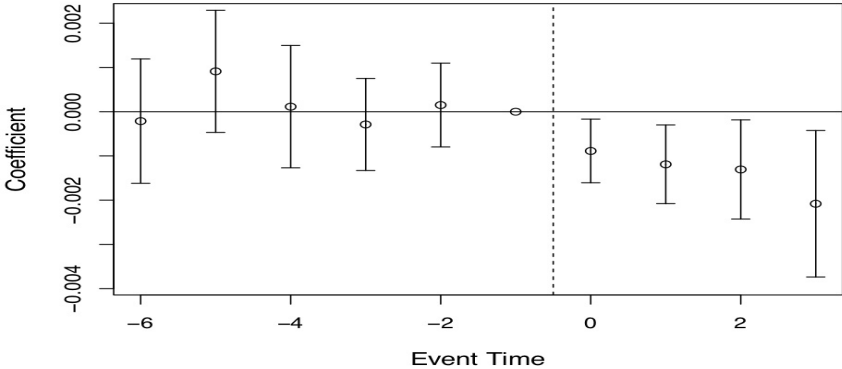


Fig 4 :法案对年死亡率的影响



图示先期效应

- **第一阶段：**图1-3 分别估计了医疗补助扩展对资格、覆盖范围和未参保率的影响
 - 处理前系数几乎都在零线上
 - 不仅点估计接近于零，而且标准误非常小
 - 意味着在扩张之前，不同州的处理组和控制组的个体之间的差异被精确地估计为零
 - 在处理后，某些人有资格获得医疗补助的概率立即上升到 0.4
 - 虽然不像处理签的系数那么精确，但作者可以排除低到 0.3-0.35 的影响
 - 有理由相信法案在隔周扩大了医疗补助计划
- **注意：**技术上来说，处理前的 0 系数并不意味着在处理后，反事实趋势和观测趋势之间的差异为 0
 - 但图示的确是令人瞩目的！
 - 因为图示给出了这种效果，但系数表格就不能够达到这样的效果

图示先期效应

- **第二阶段：**图4 显示医疗补助扩展对年死亡率的影响
 - 第一阶段的分析为第二阶段的结论增加了信服度
 - 第一阶段（处理对参保率的影响）和第二阶段（处理对感兴趣变量的影响）
- **机制分析阶段**
 - 文章还继续分析了为什么医疗补助计划扩展会降低死亡率
 - 分析了老年人群中接受针对危及生命的疾病的治疗的影响
 - 机制分析的代表作
- 关于事件分析与 DID 的关系与不同，在后面还会进一步来说

诊断 2：组别时间趋势的进一步分析 → 安慰剂三重差分设计

- DiD可靠性依赖于共同趋势假设：干预分配的过程可能使得假设不成立
- 制定政策往往根据组别的特点 → 奥运举办国
- 一个可能的选择是加入**组别线性趋势** $D_i \times Trend_t$ 以控制组间线性时间趋势的差异
- 当处理组和控制组存在明显的时间趋势差异时,直接使用DiD存在明显偏误, 但控制组间线性时间趋势后就能准确地估计处理效应
- 所以研究者的经常做法是：DiD基础上，额外控制组别线性趋势作为一种稳健性检验
 - 若平行趋势假设满足，那么是否加入组别线性时间趋势不会对估计结果产生明显影响
 - 若估计结果发生了明显改变，则预示着组别时间趋势可能存在差异，平行趋势假设可能并不满足

应用：CK1994中的 DDD

- 如果**第一阶段**和**第二阶段**的研究已经有预期结论，那么这个结论是否是肯定的呢？
- 设定一个**替代性原假设**（最低工资对就业率没有影响），那么试着去检验该备择假设
 - 通过DDD，若**不能拒绝**替代性原假设，那、就**有理由怀疑原始结论**的可信性
 - 通过DDD，若**拒绝**替代性原假设，那么就为**原始结论**提供了可信度
- CK1994 的安慰剂证伪分析中，作者考虑了另一类工人（不太可能被最低工资政策影响）
 - 理论上最低工资政策不会影响高工资工人的就业
 - 以高工资工人就业作为结果变量时，最低工资的系数为零
 - 以低工资工人就业作为结果变量时，最低工资的系数为负
 - 替代性原假设：是高收入工人和低收入工人之间的 DiD 估计量没有差异（H1 就是差异）

$$Y_{its}^{placbo} = \gamma_s + \lambda_t + \sum_{\tau=-q}^{-1} \gamma^{\tau} D_{s\tau} + \sum_{\tau=0}^m \delta_{\tau} D_{s\tau} + X_{ist} + \varepsilon_{ist}$$

- 预期： γ^{τ} 应该接近 0

应用：CK1994中的 DDD

- 引入了第三个维度“组别”（group），通过比较不同组别间的处理组和控制组在干预政策前后结果变量变化的差异来识别因果效应
- 应用场景
 1. 在共同趋势假设不满足时，引入第三个维度的差分来帮助消除处理组和控制组间的时间趋势差异（安慰剂证伪）
 2. 在共同趋势假设满足时，用于识别干预政策在不同群体间的异质性处理效应（丰富结论）

应用：CK1994中的 DDD

- 在先前分析的CK1994中一个假设是：
- 最低工资标准对 NJ 和 PA 存在**共同冲击T** → 仅仅是随机因素导致PA没接受到政策干预
- 事实上，PA 和 NJ 是不一样的，可能都有自己的发展路径
- 把这种路径叫做**特定于州的时间冲击**： NJ_t 和 PA_t
 - 不再满足共同趋势的假设 → 经典 DiD 无法准确估计 ATT
 - 存在**异质性组间时间趋势** 可以用 DDD 模型估计
- 在政策实施前，NJ的低薪和高薪就业水平由特定群组NJ固定效应表示（ NJ_h, NJ_l ）决定的，而 PA 是（ PA_h, PA_l ）
- 在政策实施后，发生如下变化，以 NJ 为例：
 1. 共同时间冲击导致就业变化的 T
 2. NJ的州特定时间冲击 NJ_t
 3. 低工资劳动者就业变化（政策仅对低收入者有效） l_t
 4. 最低工资政策带来的影响 D

应用：CK1994中的 DDD

州	群组	时期	结果	D1	D2	D3
NJ	底薪	后	$T + NJ_l + NJ_t + l_t + D$	$T + NJ_l + NJ_t + l_t + D$	$(l_t - h_t) + D$	D
		前	NJ_l			
NJ	高薪	后	$T + NJ_h + NJ_t + h_t$	$T + NJ_t + h_t$		
		前	NJ_h			
PA	底薪	后	$T + PA_l + PA_t + l_t$	$T + PA_t + l_t$	$(l_t - h_t)$	
		前	PA_l			
PA	高薪	后	$T + PA_h + PA_t + h_t$	$T + PA_t + h_t$		
		前	PA_h			

- $l_t - h_t$ 若等于 0，意味着高薪与底薪工人就业变化是相同的 → DiD 能识别，但这不现实
- 若 $l_t - h_t$ 在 NJ 和 PA 是相同，DDD 可以识别
→ 使用控制组（PA）内部的安慰剂替代处理组（NJ）内部的安慰剂

为什么 DDD 可以用来安慰剂证伪

- 对与时间趋势而言，有三个类别
- T 、 (l_t, h_t) 和 (PA_t, NJ_t)
- 将共同趋势假设更加明确为是群组差异 $l_t - h_t$ 在控制组和处理组之间是相同的 DiD基础上，额外控制组别线性趋势可以作为一种稳健性检验
- 若平行趋势假设满足，那么是否加入组别线性时间趋势不会对估计结果产生明显影响
- 若估计结果发生了明显改变，则预示着组别时间趋势可能存在差异，平行趋势假设可能并不满足
 - DiD的核心解释变量 $D_i \times Post_t$
 - 加入组别线性时间趋势 $D_i \times Trend_t$
- 二者存在比较明显的共线性，控制组别线性时间趋势会大大减少核心解释变量的变动程度从而降低估计效率、提高标准误
 - 估计系数分布明显更加分散，这表明估计量效率降低、标准误变得更大了
 - 如果处理效应是随着时间变化的，那么组别线性时间趋势会吸收一部分处理效应，导致DiD法会低估真实效应

为什么 DDD 可以用来安慰剂证伪

- 本质上看，组别时间趋势衡量了**不随时间变化的 组间差异因素**或者是**存在随时间变化的混淆因素**可能带来的内生性问题
 - 对于可观测的因素，通过添加控制变量的方法加以控制
 - 对于不可观测的因素很难直接处理，通过控制组间线性趋势差异可以部分缓解这一问题，**然而当组别时间趋势差异和时变处理效应同时存在**时目前还无法完全解决这一问题

应用：承办赛事对城市发展影响的DiD设计

- 亚运会相当于政府在杭州、嘉兴与湖州（简称杭嘉湖）三个地区的各乡镇实行了一项改革试验，任务是对赛事经济绩效进行评价
- 横向比较。承办赛事后杭嘉湖地区城市发展绩效高于其他没有承办赛事的浙江非杭嘉湖地区的绩效，那么能否认为亚运会改革取得成功的证据呢？
 - 答案是不确定的，其理由是：杭嘉湖地区与非杭嘉湖地区若在承办赛事之前就存在绩效差异，则承办赛事对两个地区的绩效差异很可能是因为继承了承办赛事前的绩效差异。因此，仅仅对承办赛事后的绩效进行横向比较，无法准确评估杭嘉湖地区赛事承办的净效应
- 纵向比较。如果杭嘉湖地区承办赛事前后的绩效出现非常明显的变化，那么我们 能否认为这就是改革取得成功的证据呢？
 - 若在杭嘉湖地区办赛前后，浙江非杭嘉湖地区同期也发生了相同幅度的变化，则答案很可能是否定的，理由是：既然在非杭嘉湖地区没有办赛，那么这些地区发生的同等变化就应该与办赛无关。我们可以认为，杭嘉湖地区与非杭嘉湖地区发生的同等变化是源于全国宏观经济形势与浙江整体经济形势的变化
 - 若在杭嘉湖地区改革前后，浙江非杭嘉湖地区同期发生了较小的变化，则对上述问题的回答就是肯定的

应用：承办赛事对城市发展影响的DiD设计

- 在进行纵向比较时，必须施加共同趋势假定，即：如果杭嘉湖地区未办赛的试验，那么也会发生与非杭嘉湖地区相同的较小变化
- 可以认为，浙江非杭嘉湖地区发生的较小变化是源于全国宏观经济形势与浙江整体经济形势的变化，而杭嘉湖地区之所以有较大变化，是因为还存在额外的赛事红利
 - 设立两个虚拟变量：

$$D_1 = \begin{cases} 1 & \text{浙江杭嘉湖地区} \\ 0 & \text{浙江非杭嘉湖地区} \end{cases} \quad D_2 = \begin{cases} 1 & \text{办赛后} \\ 0 & \text{办赛前} \end{cases}$$

- 建立 DID 模型

$$Score_i = \alpha_0 + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \alpha_3 D_{1i} \cdot D_{2i} + \varepsilon_i$$

i 代表各乡镇，Score代表经济绩效评价指标； α_0 代表办赛前所有乡镇共同的初始经济绩效均值； α_1 代表杭嘉湖地区与非杭嘉湖地区在办赛前的初始经济绩效差异； α_2 代表杭嘉湖地区与非杭嘉湖地区在办赛前后共同发生的绩效变化, 即共同趋势，具体是指全国宏观经济形势与浙江整体经济形势变化的影响； α_3 代表在控制了初始绩效差异与共同趋势之后，杭嘉湖地区所具有的额外经济绩效变化，此即赛事效应

应用：承办赛事对城市发展影响的DiD设计

在所有参数中, α_2 与 α_3 是最关键

1. 对于浙江非杭嘉湖地区 ($D_1 = 0$) :

- 办赛后 ($D_2 = 1$) 的期望绩效为 $\alpha_0 + \alpha_2$
- 办赛前 ($D_2 = 0$) 的期望绩效为 α_0
- 浙江非杭嘉湖地区改革前后的绩效差异为: $(\alpha_0 + \alpha_2) - \alpha_0 = \alpha_2$ 。其中 α_2 是共同趋势，即全国宏观经济形势变化影响与浙江整体经济形势变化影响之和

2. 对于浙江杭嘉湖地区 ($D_1 = 1$)

- 办赛后 ($D_2 = 1$) 的期望绩效为 $\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3$
- 办赛前 ($D_2 = 0$) 的期望绩效为: $\alpha_0 + \alpha_1$
- 杭嘉湖地区办赛前后的绩效差异为: $(\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3) - (\alpha_0 + \alpha_1) = \alpha_2 + \alpha_3$ 。杭嘉湖地区的绩效变化是全国宏观经济形势变化影响、浙江整体经济形势变化影响与赛事效应这三者之和

- DiD 后 α_3 就是杭嘉湖地区的办赛效应

应用：承办赛事对城市发展影响的DDD设计

- 现在考察共同趋势假设不成立时的情况
 - 假定在杭嘉湖地区办赛前后，上海发生了巨大变化，而杭嘉湖地区毗邻上海，故上海发生的巨大变化会对该地区产生外溢影响
 - 对于浙江非杭嘉湖地区，由于距离上海较远，我们假定这一外溢效应不存在 → 即使杭嘉湖地区没有进行办赛试验，上海的外溢影响也会使得杭嘉湖地区与浙江非杭嘉湖地区具有不同的绩效变化，从而导致共同趋势假设被违背
- 拓展的DiD思路是控制外溢效应所造成的趋势差异
 - 增加一个 x 设置为一虚拟变量，其中取值为1代表距离上海较近，取值为 0代表距离上海较远，然后将交互项 $x \cdot D2$ 作为解释变量模型引入模型
 - 但 x 与 $D1$ 完全共线，进而 $x \cdot D2$ 与 $D1 \cdot D2$ 完全共线，结果导致 DID 模型无法被识别

应用：承办赛事对城市发展的影响

- 共同趋势假设因上海外溢效应的存在而被违背
- 引入一个新的对照组，即与杭嘉湖地区临近的苏南地区。引入该对照组的根本目的在于识别出上海的外溢效应
- 假设苏南地区对照组具有如下四个性质
 1. 这些地区与浙江地区一样，受到全国宏观经济形势变化的影响
 2. 这些地区与浙江杭嘉湖地区一样，因毗邻上海而受其外溢影响
 3. 这些地区不属于浙江，不受浙江整体经济形势变化的影响
 4. 这些地区没有办赛，不受办赛的影响

应用：承办赛事对城市发展影响的DDD设计

- 基于上述性质可以推论
 - 在杭嘉湖地区办赛前后，苏南地区的同期绩效变化：全国宏观经济形势变化影响 + 上海外溢影响
 - 浙江非杭嘉湖地区的同期绩效变化：全国宏观经济形势变化影响 + 浙江整体经济形势变化影响
 - 浙江杭嘉湖地区的同期绩效变化：全国宏观经济形势变化影响 + 浙江整体经济形势变化影响 + 上海外溢影响 + 办赛绩效
 - 综合利用上述三类信息，可以同时识别出全国宏观经济形势变化影响、浙江整体经济形势变化影响、上海外溢影响与办赛绩效
- 通过对杭嘉湖地区办赛前后的绩效变化与同期浙江非杭嘉湖地区的绩效变化进行比较，可以识别出上海外溢影响与办赛绩效之和
- 若在此基础上，进一步剔除上海的外溢影响，则可获得办赛效应

应用：承办赛事对城市发展影响的DDD设计

- 设立三个虚拟变量：

$$D_1 = \begin{cases} 1 & \text{浙江地区} \\ 0 & \text{非浙江地区} \end{cases} \quad D_2 = \begin{cases} 1 & \text{办赛后} \\ 0 & \text{办赛前} \end{cases} \quad D_3 = \begin{cases} 1 & \text{临近上海地区} \\ 0 & \text{非临近上海地区} \end{cases}$$

- 建立 DDD 模型

$$\begin{aligned} Score_i = & \alpha_0 + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{1i} \cdot D_{2i} + \alpha_5 D_{1i} \cdot D_{3i} + \alpha_6 D_{2i} \cdot D_{3i} \\ & + \alpha_7 D_{1i} \cdot D_{2i} \cdot D_{3i} + \varepsilon_i \end{aligned}$$

α_0 代表办赛前所有乡镇共同的初始绩效均值； α_1 代表浙江地区与非浙江地区在办赛前的初始绩效差异； α_2 代表所有地区在办赛前后共同发生的绩效变化，即共同趋势，具体是指全国宏观经济形势变化的影响； α_3 代表临近上海地区与非临近上海地区在办赛前的初始绩效差异； α_4 代表浙江整体经济形势变化的影响； α_5 代表浙江杭嘉湖地区既属于浙江又临近上海这一独特区位特征造成的初始绩效差异； α_6 代表上海的外溢效应； α_7 代表在控制了由各种原因造成的初始绩效差异、全国宏观经济形势变化的影响、浙江整体经济形势变化的影响、上海的外溢效应之后，杭嘉湖地区所具有的额外绩效变化，即办赛绩效

应用：承办赛事对城市发展影响的DDD设计

在所有参数中, α_2 、 α_4 、 α_6 、 α_7 是最为关键的

- 对于浙江非杭嘉湖地区 ($D_1 = 1, D_3 = 0$)
 1. 办赛后 ($D_2 = 1$) 的期望绩效: $\alpha_0 + \alpha_1 + \alpha_2 + \alpha_4$
 2. 办赛前 ($D_2 = 0$) 的期望绩效: $\alpha_0 + \alpha_1$
 3. 浙江非杭嘉湖地区办赛前后的绩效差异为: $\alpha_2 + \alpha_4$ 。这一差异是全国宏观经济形势变化影响 + 浙江整体经济形势变化影响
- 对于苏南地区 ($D_1 = 0, D_3 = 1$)
 1. 办赛后 ($D_2 = 1$) 的期望绩效: $\alpha_0 + \alpha_2 + \alpha_3 + \alpha_6$
 2. 办赛前 ($D_2 = 0$) 的期望绩效: $\alpha_0 + \alpha_3$
 3. 与上海临近的苏南地区在办赛前后的绩效差异为: $\alpha_2 + \alpha_6$ 。这一差异是全国宏观经济形势变化影响 + 上海外溢影响之和

应用：承办赛事对城市发展影响的DDD设计

- 对于浙江杭嘉湖地区 ($D_1 = D_3 = 1$) :
 1. 办赛后 ($D_2 = 1$) 的期望绩效: $\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + \alpha_6 + \alpha_7$
 2. 办赛前 ($D_2 = 0$) 的期望绩效: $\alpha_0 + \alpha_1 + \alpha_3 + \alpha_5$
 3. 浙江杭嘉湖地区办赛前后的绩效差异: $\alpha_2 + \alpha_4 + \alpha_6 + \alpha_7$ 。这一差异是全国宏观经济形势变化影响 + 浙江整体经济形势变化影响 + 上海外溢影响 + 办赛绩效
- 综上所述，DDD模型中的 D_2 捕捉了全国宏观经济形势变化的影响 α_2 ；交互项 D_2D_1 捕捉了浙江整体经济形势变化的影响 α_4 ；交互项 D_2D_3 捕捉了上海的外溢影响 α_6 ；交互项 $D_1D_2D_3$ 捕捉了办赛绩效 α_7 。为了进一步验证此结论，考察浙江杭嘉湖地区办赛前后的绩效差异与同期浙江非杭嘉湖地区的绩效差异这两者的差：

$$(\alpha_2 + \alpha_4 + \alpha_6 + \alpha_7) - (\alpha_2 + \alpha_4) = \alpha_6 + \alpha_7$$

- 两者之差应该等于上海外溢影响与办赛绩效之和
- 在此基础上，若再进一步剔除上海外溢影响，则可获得办赛绩效。显然，当 α_6 与 α_7 分别对应上海外溢影响与改革绩效时，情况确实如此。

对经典 DiD 一般问题归纳

DiD 既是估计量，又是研究设计

- DiD 与传统计量模型解决内生性问题思路一样么？不一样
- 作为估计量，目标是估计ATT，但能否获得一致估计量取决于：共同趋势假设是否成立
 - 更严谨的说法是，在满足识别假设的前提下的DiD能正确识别ATT
 - 共同趋势假设经过变形其实就是DiD下的外生性假设 $E(\Delta \varepsilon | D) = 0$
 - DiD解决内生性问题的思路是“假设”不存在内生性问题
- 作为研究设计，DiD要追溯到1855 年物理学家约翰·斯诺对伦敦市的霍乱成因研究，CK1994正是延续了该思想
 - 没有研究设计中“双重比较”的想法，就不会产生所谓的DiD估计量
 - 从研究方法历史角度看，现有双重比较，才有双重差分
 - 印证了这类因果推断的方法起源是**比较**的起点
- **错误的说法是**：DiD可以避开内生性问题
 - DiD解决内生性问题仍然依赖于干预或政策冲击本身的外生性

DiD应用注意事项1/4：制度背景和政策实施真实情况

- 最多场景是评估政策效应 → 制度背景 + 真实实施情况一定要阐述清楚
 - 假如政策发布了，却没有很好地实施，或者“上有政策，下有对策”，进而得到误导性结论
 - 需要对政策的具体实施情况有深入、清晰的了解，包括政策何时开始实施、是否按要求准确执行，以及行为主体是否采取了应对措施等问题
- 通常认为地方政府高效率的财政补贴对企业研发创新能力产生积极作用。但是范子英和王倩（2019）的研究，发现在财政补贴实施过程中存在“列收列支”问题，即地方政府为了增加名义上的税收收入，会先向企业多征收一部分税款，再以财政补贴的名义返还回去。这使得名义上的财政补贴实际上是企业自有资金，而这部分“虚假”的财政补贴不会对企业经营行为产生实质性影响。因此，财政补贴的低效率可能是由于对政策实施真实情况了解不够导致的

DiD应用注意事项2/4：干预政策需要严格外生或随机分配吗？

- DiD核心假设之一是处理组和控制组的时间趋势在没有干预的情况下是平行的
- 但是，DiD法本身并不能解决内生性问题，而是假设干预政策是外生的，即干预政策的实施与模型中的扰动项（误差项）无关
- 贫困县政策的经济发展效应评估为例
 - **水平意义**上的外生性中，要求政策的分配过程是随机的或近似随机的，但在现实中，由于政策的目标和对象的特定性，很难有完全随机分配的政策
 - 满足**差分意义**上的外生性：关注的是处理组和控制组的经济发展趋势是否相似，而不是政策的分配过程是否随机
 - 为了满足差分意义上的外生性，常常使用PSM数据进行预处理，选取与贫困县禀赋条件类似的非贫困县作为控制组，以使平行趋势假设成立，然后再应用DiD法来估计因果效应

DiD应用注意事项3/4：有无溢出效应

- DiD另一核心识别假设——SUTVA（稳定单位处理值假设），即干预政策没有溢出效应
- 在现实政策分析中，SUTVA往往难以满足，因为政策干预往往会对非处理组产生间接的影响
- Lu等（2019）研究了中国经济开发区对当地经济发展的影响，并通过两种不同的策略来检验溢出效应
 - 首先，检验了经济开发区附近的同县其他村庄的经济发展情况
 - 其次，检验了村庄离经济开发区的距离与其经济发展的关系
 - 通过这两种策略，研究者发现经济开发区政策的溢出效应并不显著
- **注意：**如果研究本身就是探讨政策的溢出效应，那么DiD法则不适用
 - 例如地区产业政策对体育企业选址和集聚效应的影响

DiD应用注意事项4/4：一般均衡视角的成本收益分析

- 评估公共政策效果时，仅依赖双重差分法（DiD法）可能是不足够的，需要结合成本-收益分析从更宏观、更全面的角度来审视政策的效果和效率
 - Duflo（2001）的研究评估了印度尼西亚修建学校的投资效果，并通过成本-收益分析发现这项投资的内部回报率非常高，显示了教育投资的高收益
 - Lu等（2019）的研究评估了中国经济开发区政策的效果，通过成本-收益分析发现该政策带来了很高的净收益
- 虽然DiD能评估政策对特定结果变量的影响，但不能提供政策的机会成本和净收益信息
 - 成本-收益分析依赖关键假设和因素：折现率、时间范围、边际成本和边际收益、是否有外部性、风险和不确定性、受益人和付费、替代方案、非货币效应等
 - 从一般均衡的视角考虑政策效应对其他核心因素的影响：公共政策通常影响多个市场和多个方面的经济活动、长期效应和短期效应等

时变DiD

——具有时间异质性的双向固定效应模型

时变DiD的设定

- 在 DiD 模型中包含**先期**和**滞后**，可以检查时点后处理效应的时变变化
 - 直观地观察和检验政策发生前后个体行为的时变反应与组间差异
 - 时变DiD可以检验处理组和控制组在政策发生前是否具有相似的时间趋势，从而论证平行趋势假设是否成立
- 估计政策实施后各期处理效应
 - 标准DiD只能估计在样本期内的平均效应，但现实中一些政策效应存在滞后，需要施行一段时间才能发挥作用
 - 假如存在一般均衡效应，一项政策的短期和长期效果可能存在比较明显的差异
- 时变DiD的识别假设相较于DiD更为宽松
 - 对于政策交错发生的情形，若处理效应在群组或时间维度上存在异质性，DiD估计结果存在偏误
 - 目前时变DiD能很好估计**时间维度上异质性的处理效应**

使用标准DiD进行因果推断的前提

- **严格外生假设**（源于面板数据的基本假设）（DiD语境中的共同趋势假设）
 - 混淆因素中不能出现随时间推移而变化的 | 过去的结果变量不能当期结果变量产生影响 | 过去的结果变量或协变量不能对当期和未来的处理状态产生影响 | 当期的处理状态不能对未来的结果变量产生影响
 - 数学上严格外生假设**强于**共同趋势假设；实际中差别不大（Xu,2022）
- **无预期效应假设**
 - 个体在当期的结果变量不会受到个体在未来的接受政策处理状态的影响
 - 个体并不能预知其在未来是否会接受政策处理，从而根据这种预期改变其行
- **SUTVA**：不同个体是否受到政策冲击是相互独立的，受政策冲击的个体不影响任何其他个体的结果
- **同质性处理效应**
 - 组别间是同质，即同一政策对于处理组内所有组别的影响是相同的
 - 时间维度同质，即对于同一时间受到政策处理的所有个体，随着时间推移，处理效应的大小不变

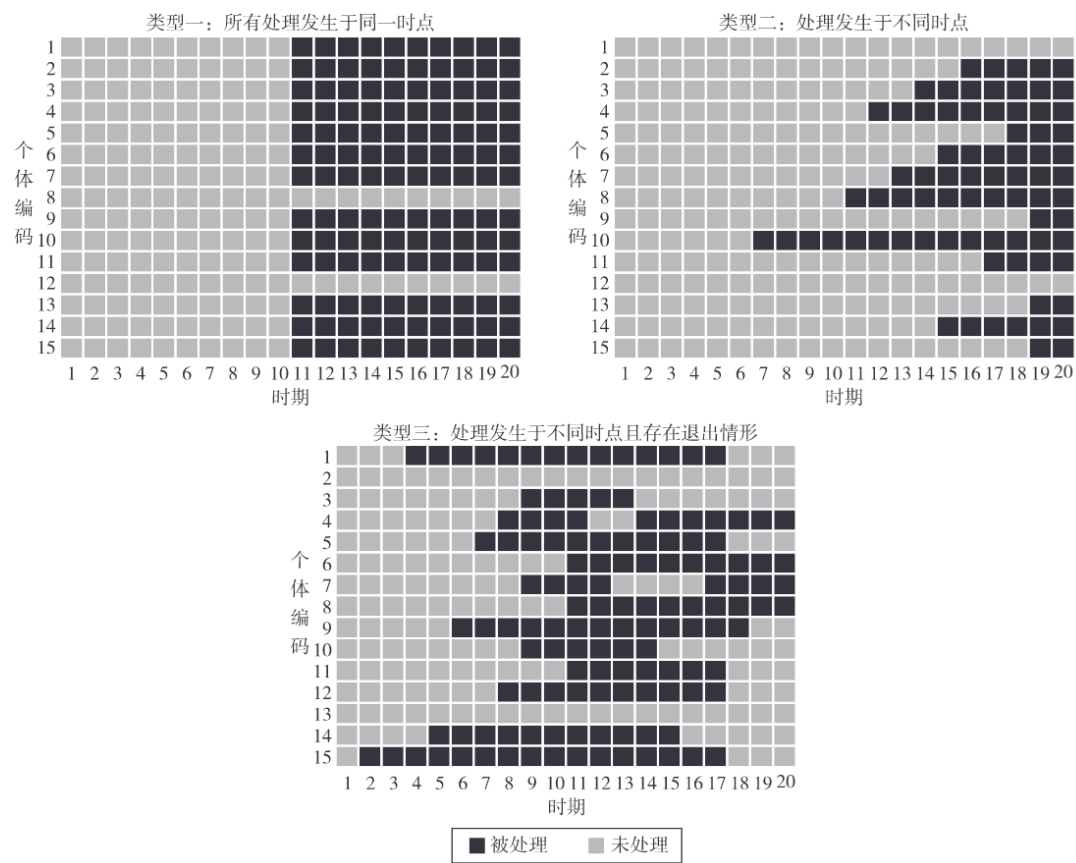
异质性处理效应会导致TWFE估计量出现偏误

- 在标准DiD中处理组在同一个时间点受到干预，然而现实中有相当多的政策并非是一次性全面实施，而是先在某些地区试点后再分批逐步推广，个体处理时点并不一致；即便在同一个时点所有个体接受同一个政策，但也可能比现在接受处理后政策出现效果的程度是不一样的
 - 示范区、体育产业基地、试点推广、联邦制国家
- 能够表达时变DiD的双重固定效应模型

$$Y_{it} = \alpha + \delta D_{it} + \theta X_{it} + \mu_i + \lambda_t + \varepsilon_{it}$$

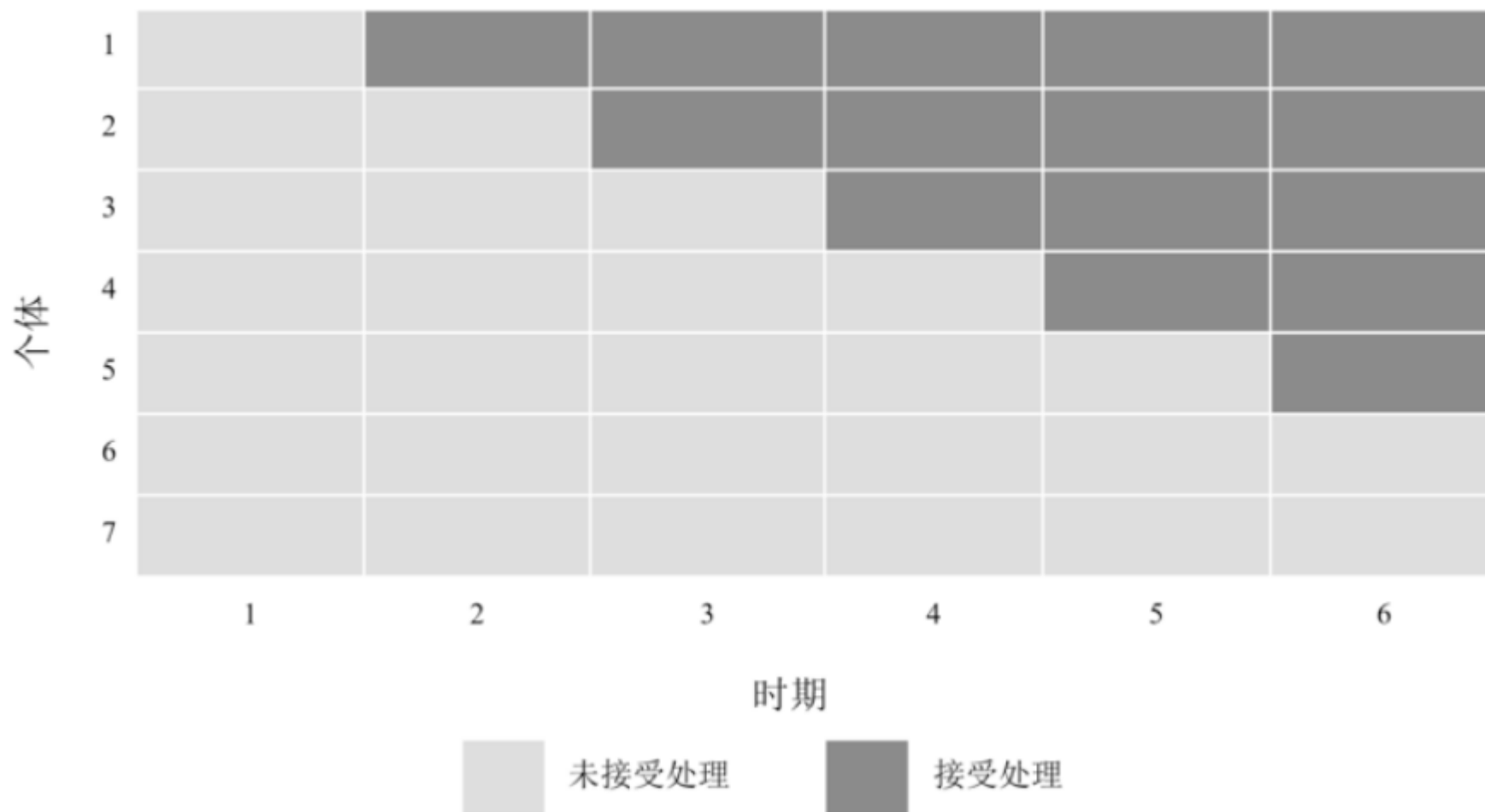
- D_{it} 表示个体 i 在 t 期的处理状态, 接受处理时取 1 , 未接受处理取 0
- 当处理组受到干预时点全部相同时, D_{it} 分解为 $D_{it} = D_i \times Post_t$, 还原成标准DiD
- 当处理组受到干预时点不同时, D_{it} 分解为 $D_{it} = D_i \times Post_{i,t}$
 - 注意：由于 δ 没有下标，时变DiD获得的是**整体的平均处理效应**

政策交错实施情形



- 单一时点
- 多时点，无退出
- 多时点，有退出

干预时点交错发生的 D_{it} 取值的典型示例



异质性处理效应会导致TWFE估计量出现偏误

- 时变 DiD 可能存在一些比较严重的问题（ Callaway 和 Sant'Anna, 2020 ； de Chaisemartin 和 d'Haultfœuille, 2020； Goodman-Bacon , 2021）
- 最主要的问题在于，当政策效应随着时间改变时，时变 DiD 估计结果 [即 δ] 并不是一个定义良好的平均处理效应，而是多个标准 2×2 DiD 的平均处理效应的加权平均，并且权重可能是负的（ de Chaisemartin 和 d'Haultfœuille, 2020）
 - 意味着即使干预本身对所有时点的处理组都是正效应，但时变DiD的估计系数仍然可能为负
 - **异质性**处理效应情形下时变DiD的单一系数估计出现偏误

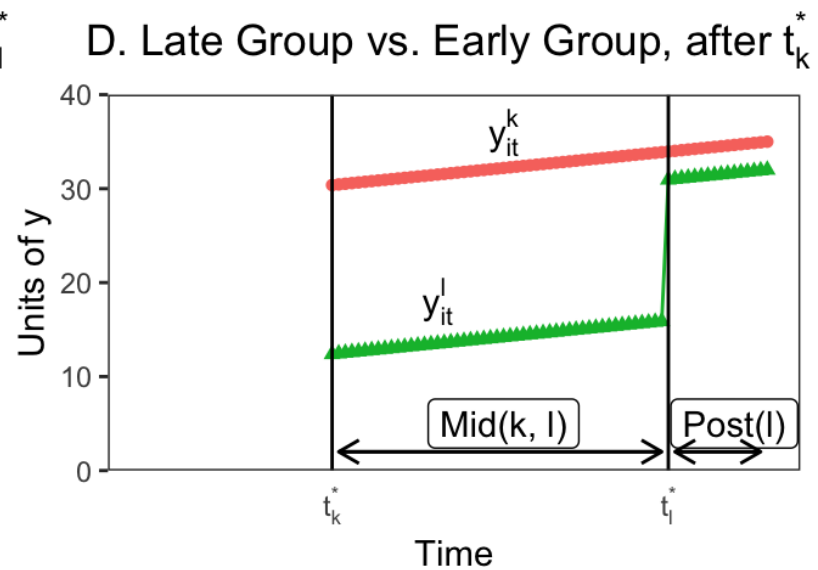
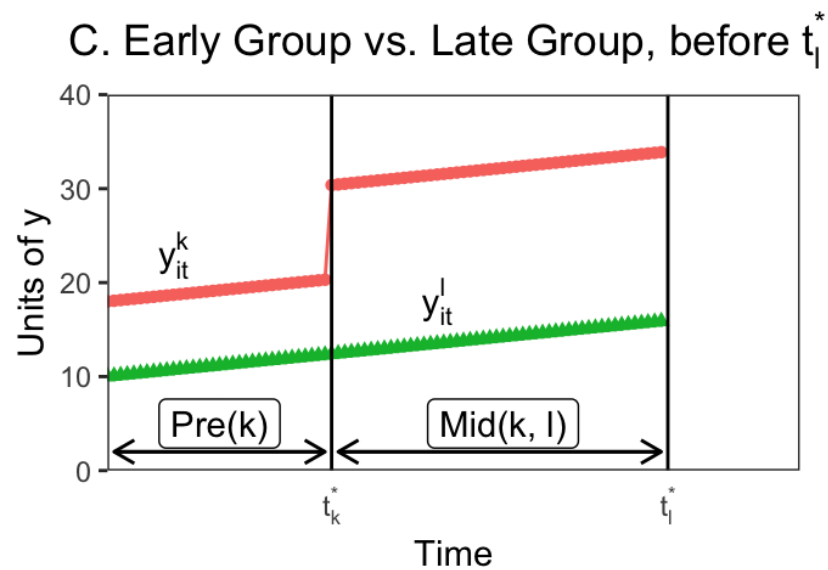
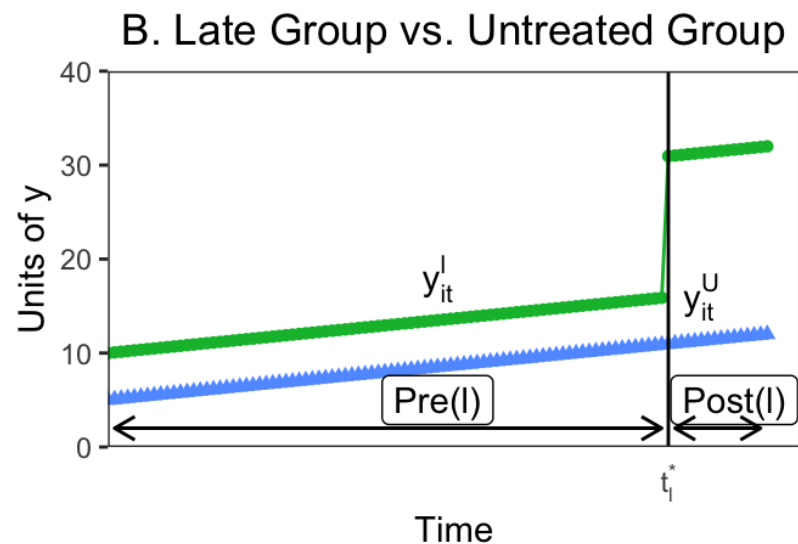
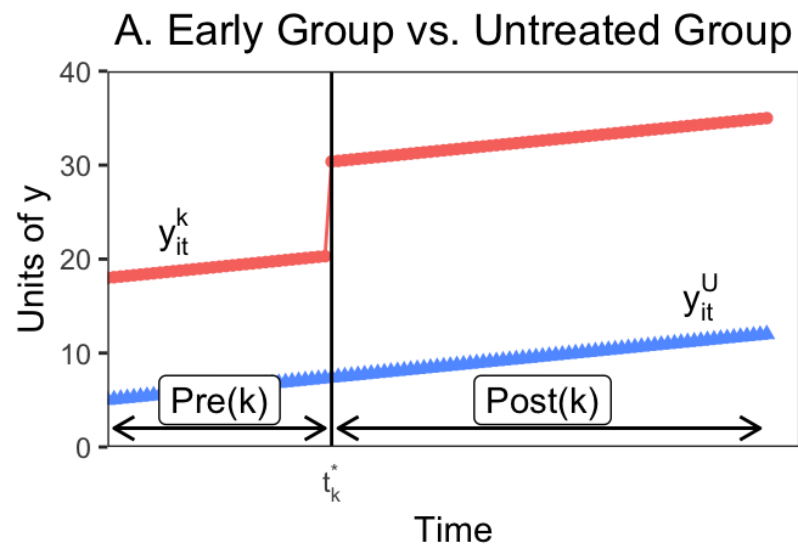
培根分解定理

- **静态情形**下，Goodman-Bacon（2019）指出即使共同趋势假设满足，TWFE也会存在“坏的控制组”问题
- **动态情形**下，Sun和Abraham（2021）对培根分解定理进行了延伸：时变情形下，不仅存在“坏的控制组”问题，而且每期的估计系数还受到跨期交叉污染而变得难以解释，进而出现平行性趋势检验失效的风险

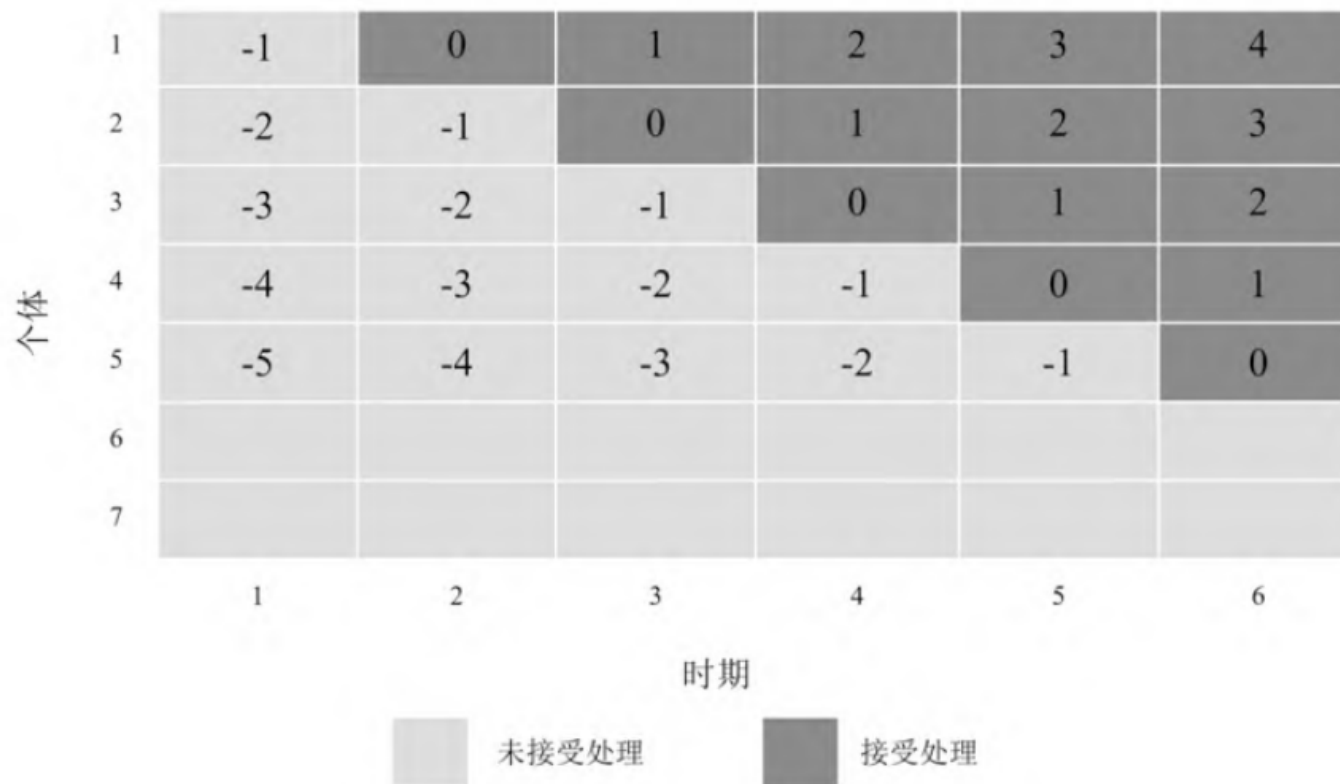
培根分解定理

- 假设存在早期处理组(k), 后期处理组(l)和从不接受处理的控制组(u)。k组个体在第34期接受政策 $t_k^* = \frac{34}{100}T$, l组个体在第85期接受 $t_l^* = \frac{85}{100}T$, 共有100期
 - “好的控制组”: 从未接受处理组 + 相对较晚接受处理的组
 - “坏的控制组”: 较早接受处理组 → 事前趋势已经发生了变化
 - 标准DiD中仅存在k组或l组中的一个 → 使用TWFE估计没有偏误

培根分解定理



练习：谁是控制组？



- 时变 DiD 中第s期的控制组：从未接受处理的控制组个体外 + 当期未受处理但在未来会受到处理的处理组个体

培根分解定理

- 时变DiD估计量是所有可能的 2×2 标准 DiD 估计量加权平均值，其中的权重基于群组的大小和处理的方差

$$p \lim_{N \rightarrow \infty} \hat{\beta}_{DiD} = VWATT + VWCT - \Delta ATT$$

- $VWATT$ 方差加权的平均处理效应” (Variance-Weighted ATT) 是**正权重加权的ATT**
- $VWCT$ 方差加权共同趋势” (Variance-Weighted Common Trend)是由所有可能的 2×2 标准DiD的处理组和控制组之间的平行趋势加权平均得到的
- ΔATT 加权的处理效应变化，仅在“坏”控制组出现时不为0

培根分解定理

- 时变DiD估计量是所有可能的 2×2 标准 DiD 估计量加权平均值，其中的权重基于群组的大小和处理的方差

$$p \lim_{N \rightarrow \infty} \hat{\beta}_{DiD} = VWATT + VWCT - \Delta ATT$$

- 处理效应**不随**时间变化($\Delta ATT = 0$) + 方差加权共同趋势满足 ($VWCT = 0$)
→ VWATT (方差加权ATT) 与ATT (样本加权) 完全匹配
- 处理效应**随着**时间变化 + 共同趋势假设满足
→ 时变DiD估计量仍可能有偏且不一致的

动态情形下的 DiD 推断

- 时变DiD，通过检验处理组和控制组在干预前和干预后的组间均值差异变化来识别政策的时变效应
- 主要问题是：在干预时点交错发生导致**无法定义一个绝对的时间参照点**作为处理前和处理后的分界线，不能以绝对时间作为参照系，而是以干预发生时点作为相对时间参照系

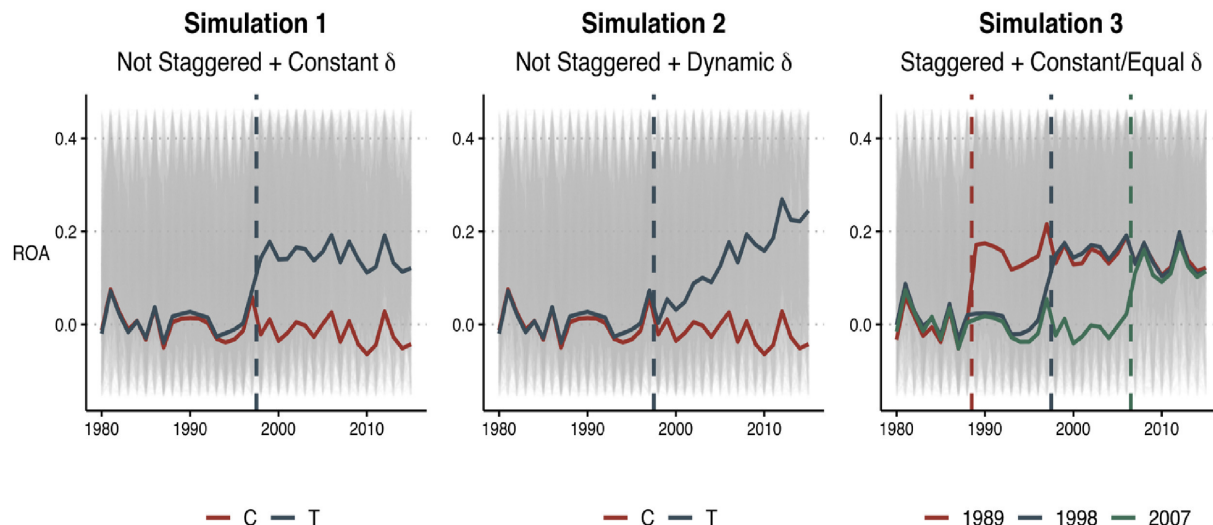
$$Y_{it} = \alpha + \beta_s^{\text{precut}} [D_i \times \mathbf{1}(t - T_D < \underline{EW})] + \sum_{s=\underline{EW}}^{-2} \beta_s^{\text{pre}} [D_i \times \mathbf{1}(t - T_D = s)] \\ + \sum_{s=0}^{\overline{EW}} \beta_s^{\text{post}} [D_i \times \mathbf{1}(t - T_D = s)] + \beta_s^{\text{postcut}} \left[D_i \times \mathbf{1}\left(t - T_D > \overline{EW}\right) \right] + \mu_i + \lambda_t + \varepsilon_{it}$$

- $\mathbf{1}(\cdot)$ 是示性函数， T_D 是政策发生当期， \underline{EW} 和 \overline{EW} 是事件窗口的开始期和结束期
 - 若以距离干预发生时点的**相对时间为参照系** ($t - T_D = s$)
 - 若以**绝对时间为参照系** (T_t^s) \rightarrow 意味着所有样本将于同一个时间点接受处理，时变DiD就成为“一刀切”DiD

应用：Simulation (Baker et al.,2021)

- 企业财务和会计应用中使用的数据集
- y_{it} 在平衡面板中， $T=36$ ， 1980/2015， 并且有1000家公司 i
- 时间不变个体效应和时间变化的年份效应 $\sim N(0, 0.5^2)$
- 公司总部位于50个州中随机选择一个， 这些州组成了处理组

(i) Trends in Outcome Path

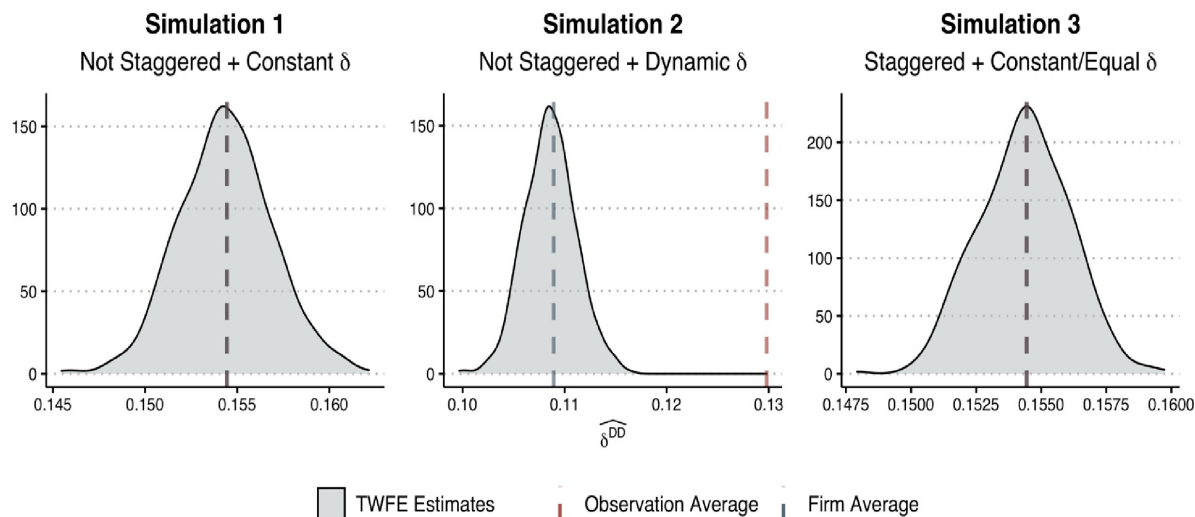


1.“一刀切” + 静态处理效应
→ TWFE一致(个体层面 + 群组层面)

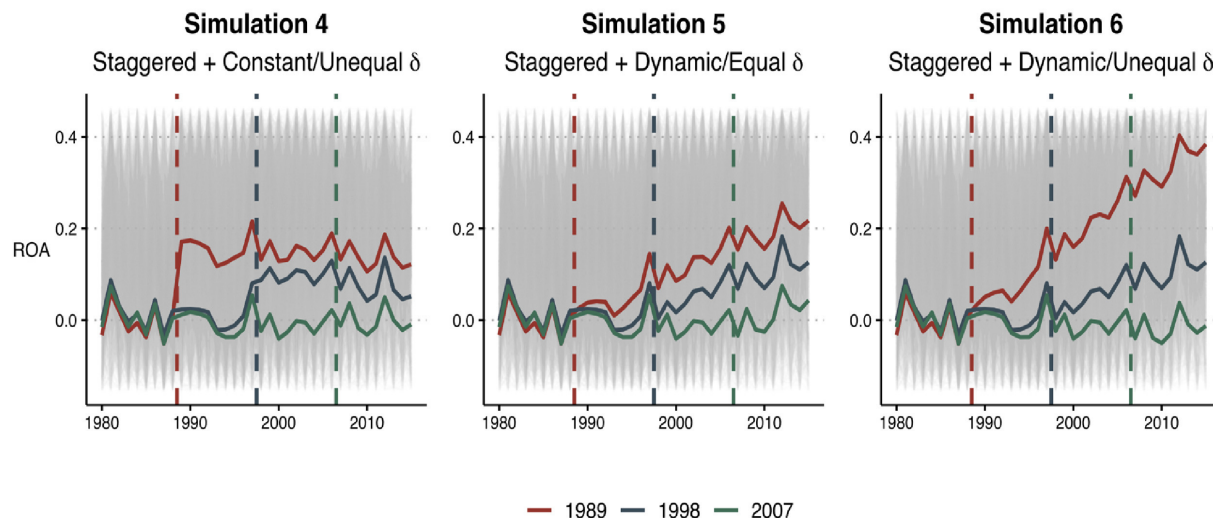
2.“一刀切” + 动态处理效应
→ TWFE一致(仅个体层面)

3.“多时点” + 静态处理效应
+ 无组间异质性
→ TWFE一致(个体层面 + 群组层面)

(ii) TWFE DiD Estimates on Simulated Data

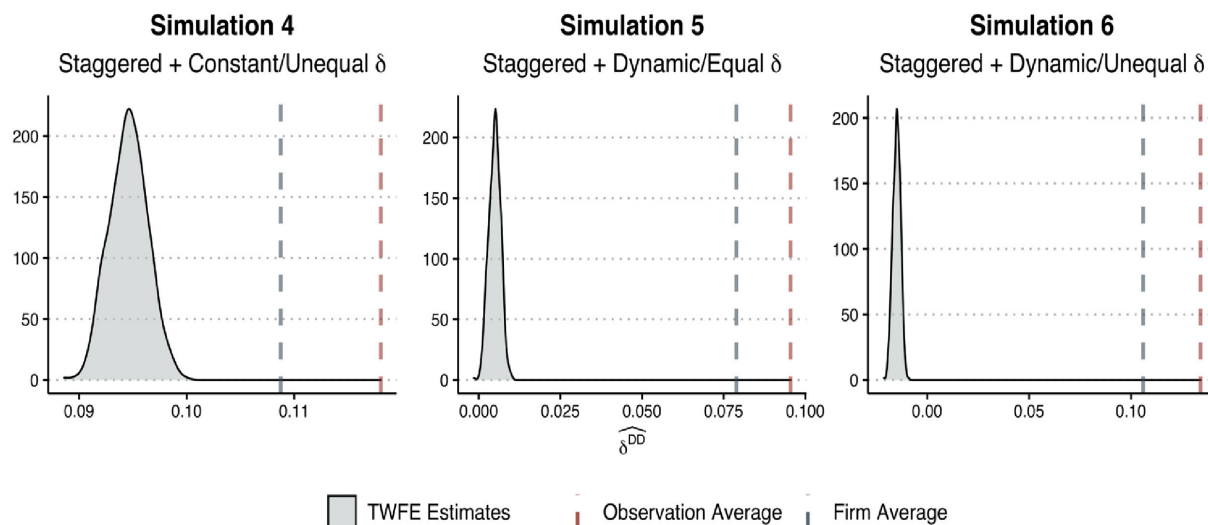


(i) Trends in Outcome Path



4.“多时点” + 静态处理效应
+ 有组间异质性 \rightarrow TWFE方向性偏误(个体层面可预估)
 $\rightarrow \Delta ATT = 0$ 且 $VWCT=0$
得到的是 **方差加权ATT** 而非
样本量加权ATT

(ii) TWFE DiD Estimates on Simulated Data



5.“多时点” + 动态处理效应
+ 无组间异质性
 \rightarrow TWFE严重偏误
 \rightarrow 使用**事件分析方法**OK

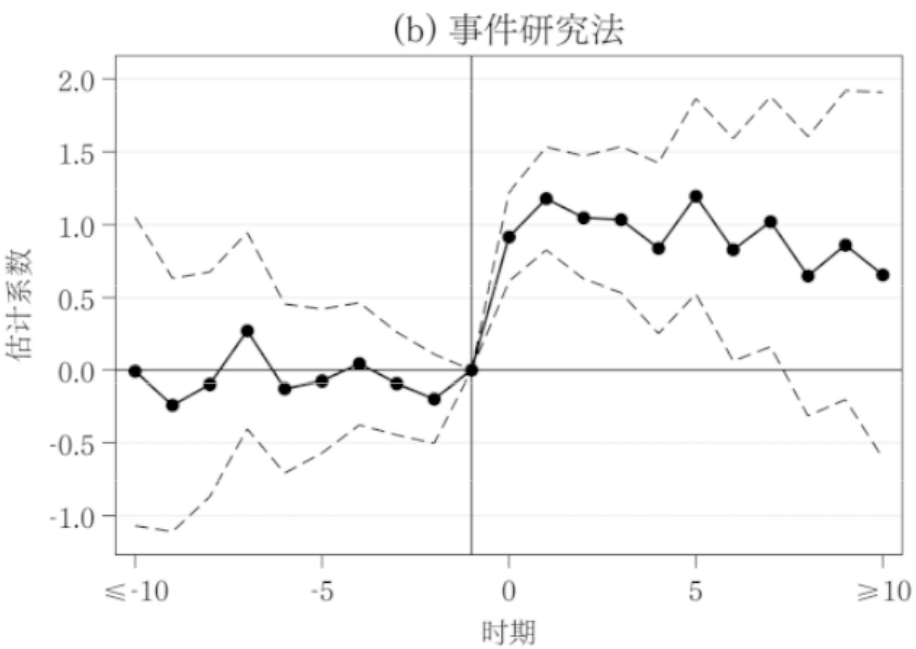
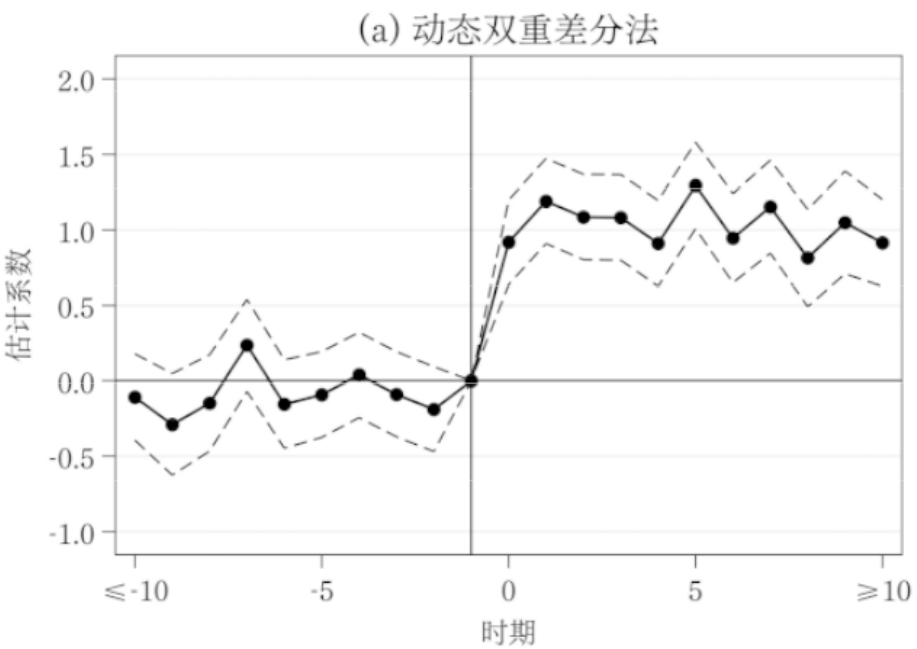
6.“多时点” + 动态处理效应
+ 有组间异质性
 \rightarrow TWFE**负权重**

事件分析法是时变DiD的特例

$$Y_{it} = \alpha + \beta_s^{\text{precut}} \mathbf{1}(t - T_D < \underline{EW}) + \sum_{s=\underline{EW}}^{-2} \beta_s^{\text{pre}} \mathbf{1}(t - T_D = s) \\ + \sum_{s=0}^{\overline{EW}} \beta_s^{\text{post}} \mathbf{1}(t - T_D = s) + \beta_s^{\text{postcut}} \mathbf{1}(t - T_D > \overline{EW}) + \mu_i + \lambda_t + \varepsilon_{it}$$

- T_D 是政策发生当期
- 如果所有个体都会受到处理，仅仅是处理时点不同，且没有从未受到处理的控制组，那么样本中全部观测值的 $D_i = 1$ ，时变DiD 就变换为事件分析法
 - 事件研究法本质上是不存在不受处理的控制组的时变 DiD
- 对应情形 5
 - 动态处理效应
 - 无组间差异性
- 这就回应了时变 DiD（事件分析法）为什么可以解决**时间维度上异质性处理效应**

事件研究法剔除了控制组样本使得样本量偏小、估计系数的标准误更大



时变 DiD 或事件研究方法的注意事项

- 事件窗口宽度的选择
 - 干预交错发生的数据结构涉及到的事件窗口 宽度要更长一些： $2T-1$
 - 比如若数据集包含 10 期的观测值，其中既有第 1 期就接受干预的个体，也有到第 10 期才接受干预的个体，那么该样本涉及到的窗口宽度为干预前 9 期、干预发生当期以及干预发生后 9 期，共 19 期
 - 由于窗口宽度大于样本时间跨度，观测值在干预前后各期的分布是不平衡的，一般而言距离干预时点越远的样本越少
- 不平衡样本会带来样本选择偏误（selection bias）和样本消耗（attrition）问题的困扰
 - 选择的事件窗口越宽，样本不平衡现象越严重，会愈发加剧上述担忧
- 事件窗口越长，越有可能受到同时期发生的其他事件和混杂因素的干扰
 - 一般来说窗宽选择越宽则样本规模越大、估计越有效（efficient），但可能会有更大的偏误（bias）
- 由于事件研究法的估计结果对事件窗口的选择较为敏感，在实际研究中通常需要更换事件窗口宽度来做一些稳健性检验

在异质性处理效应中获得稳健估计量的三种思路

1. 计算“组别－时期”ATT (Cohort-Specific ATT, CATT) 并进行加权平均。先计算特定组别－特定时期的平均处理效应，再在组别、时期两个维度进行合理的加权加总。避免使用较早接受处理组作为控制组，从而避免估计偏误
 - 如de Chaisemartin 和 D'Haultfoeuille (2020a, 2022a)、Callaway 和 Sant'Anna (2021) 以及Sun和Abraham (2021)
2. 插补(Imputation Estimator)出合理的反事实结果。首先利用从未接受处理的样本或尚未接受处理的样本估计出每个处理组个体每个时期的反事实结果。然后，计算处理个体的处理效应。最后，加总个体层面的处理效应，得到平均处理效应思路：通过构造合理的反事实，避免坏的控制组问题。
 - Borusyak等(2021)、Liu等(2022) 和Gardner(2021)
3. 利用堆叠回归 (Stacked Regression Estimator)的方式。为每一个处理组的观测匹配了从未接受处理或尚未接受处理的观测，进而形成一个处理组。随后将这些数据堆叠在一起，通过进一步加入组别一个体、组别一时间固定效应进行线性回归。思路：避免使用较早接受处理组作为控制组。
 - Cengiz等(2019)和Gardner(2021)

总结

·单一系数的偏误来源就是组间或者时间上处理效应的异质性 ·只有所有cell的处理效应都一致，且满足平行趋势，此时单一系数才能捕捉ATT ·当只有组间的处理效应异质性时，此时单一系数是方差加权的ATT，可以类比OLS和Matching ·更加灵活的ESA方法当政策效果的动态性在组间一致时，是可以用的

应用: did2s

<https://pkuzzq-image.oss-cn-beijing.aliyuncs.com/uPic/yQL1vV.png>

<https://pkuzzq-image.oss-cn-beijing.aliyuncs.com/uPic/2UaVA1.png>

<https://pkuzzq-image.oss-cn-beijing.aliyuncs.com/uPic/v06rTQ.png>

<https://pkuzzq-image.oss-cn-beijing.aliyuncs.com/uPic/dnRmSy.png>

<https://pkuzzq-image.oss-cn-beijing.aliyuncs.com/uPic/7YPt46.png>

<https://pkuzzq-image.oss-cn-beijing.aliyuncs.com/uPic/G7bFlS.png>

<https://pkuzzq-image.oss-cn-beijing.aliyuncs.com/uPic/tdlxy4.png>

从 DiD 到 Event Study

——具有时间异质性的双向固定效应

选择性偏差

- **选择性偏差** (selection bias)：样本选择过程中的非随机而导致的结论偏差，包括**自选****择偏差** (self-selection bias) 和**样本选择偏差** (sample-selection bias)
 - 自选偏差：实验组的非随机分配导致的 → 是否进入实验组是个体选择的结果或是与个体相关的因素（包括可观测和不可观测两类）导致的
 - 例子
 - 样本选择偏差：非随机选择的样本不能够反映总体的特征，从而使结果估计量产生偏差
 - 例子
- 评估政策处理效应的方法就是尽可能的降低选择性偏差

选择性偏差的表达

- 直接面对数据时，通常使用处理样本结果变量的均值与未处理样本结果变量的均值之差来估计 ATE
 - differen-in-mean, 简称 DIM
 - simple-difference-in-outcome, 简称 SDO
 - Naive comprison

$$\begin{aligned} DIM &= E[Y^1 | D = 1] - E[Y^0 | D = 0] \\ &= \frac{1}{N_T} \sum_{i=1}^n (y_i | d_i = 1) - \frac{1}{N_C} \sum_{i=1}^n (y_i | d_i = 0) \end{aligned}$$

- 当满足**独立性假设** (independence assumption, 简称 IA) 时: $(Y_1; Y_0) \perp D$
 - $DIM = ATE = ATT = ATU$

DiD扩展

DiD 的逻辑本质

- 其基本思路是寻找观测样本在两个维度上的差异
 1. 控制不可观测的时间趋势
 2. 测度政策效应的变化
- 也就是说，涉及到任何这类两个维度的差异之差异，都可以从DiD的角度去理解
- 也就是说，几乎所有的交互项模型都可以理解为一种DiD法

类 DiD

Mayzlin等（2014）的研究，他们研究了造假成本对在线旅店预定网站的消费者评论的影响

- Expedia：只有实际完成订单的消费者可以评价服务质量
- Trip Advisor：任何人都可以评价服务质量
 - 导致造假成本是不同的
- 当一家旅店周围没有其他旅店存在时，该旅店在TripAdvisor上的好评率显著高于在Expedia上的好评率 → 理论上：该旅店试图操纵评论
- 当一家旅店旁边有同类竞争对手时候，该旅店Trip Advisor上的差评率显著高于Expedia上的差评率 → 试图打压竞争对手，为对手恶意评低分
 - 转化为 DiD 的识别策略
- Expedia和Trip Advisor的造假成本构成了一个维度的差异（类似政策维度变化）
- 旅店邻近范围内是否存在直接竞争者构成了另一维度的差异（类似不可观测的时间趋势）
 - 通过二者之差就能够识别出造假成本对网站消费者评论操纵的影响

交错DiD (staggered DiD)

- 标准双重差分法模型和双向固定效应双重差分法模型涉及的政策实施时点或冲击发生时点为**同一时期**。然而，现实中政策大多先有试点再逐步推广，是渐进的过程。
- 技术上将政策分组虚拟变量 D_i 变为 D_{it} ，可用来表示个体 i 在时间 t 处是否受到政策冲击，无需再生成交互项。
- 在实际应用中，staggered DiD 会遇到**难以找到控制组、部分样本始终为处理组、异质性处理效应**等问题

广义DiD (generalized DiD)

- 当**所有**研究对象**均或多或少**同时受到了政策干预，即仅有处理组而无控制组时，仍然能够考虑应用双重差分法。
- 根据研究对象受到的具体冲击情况来构建处理强度 (treatment intensity) 指标来进行分析
- 此时个体维度并不是从 0 到 1 的改变，而是连续的变化 → 将个体维度的政策分组虚拟变量替换为用以表示不同个体受政策影响程度的连续型变量
- Nunn 和 Qian (2011) 研究土豆种植扩散对欧洲人口增长的影响
 - 欧洲几乎所有地区都种植了土豆，不存在未种植土豆的地区，因此没有标准意义上的控制组
 - 他们的选择是将地区间土豆种植适宜度作为处理强度，以 1700 年前后为处理时点，使用广义双重差分法估计了引入土豆对人口增长的影响

队列DiD (cohort DiD)

- 也称截面DiD，即使用横截面数据来评估某一历史事件对个体的长期影响
- 同样是比较两个维度上的差异大小
 - 一个维度为**地区间差异**，标识该**地区**是否受干预政策影响或干预强度
 - 另一维度为**出生队列间差异**，标识**个体**是否受到了干预政策的影响
- 本质上是使用未受政策干预的出生队列作为受到政策干预的出生队列的反事实结果
 - Duflo (2001) 是经典研究

模糊DiD (fuzzy DiD)

- 在标准 DiD 应用情境中，处理组和控制组之间通常泾渭分明，因此通过分组差分得到较为“干净”的处理效应
- 但有时冲击并未带来急剧 (sharp) 变化
 - 处理组中虽然受冲击率高于其他组别，但并没有完全被干预或受政策冲击
 - 控制组中也并非完全没有受到冲击
 - 处理组和控制组之间没有明确的分野，不存在“干净”的处理组与控制组
 - de Chaisemartin 和 d'Haultfoeuille (2018)