

# 体育经济分析: 原理与应用

DiD

周正卿

29 October 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。

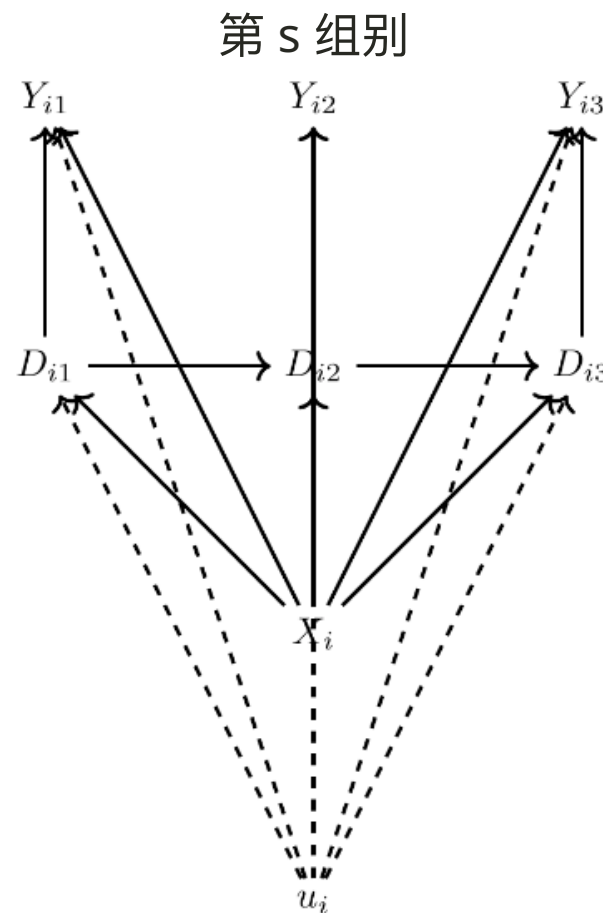
# 面板数据框架

## 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)

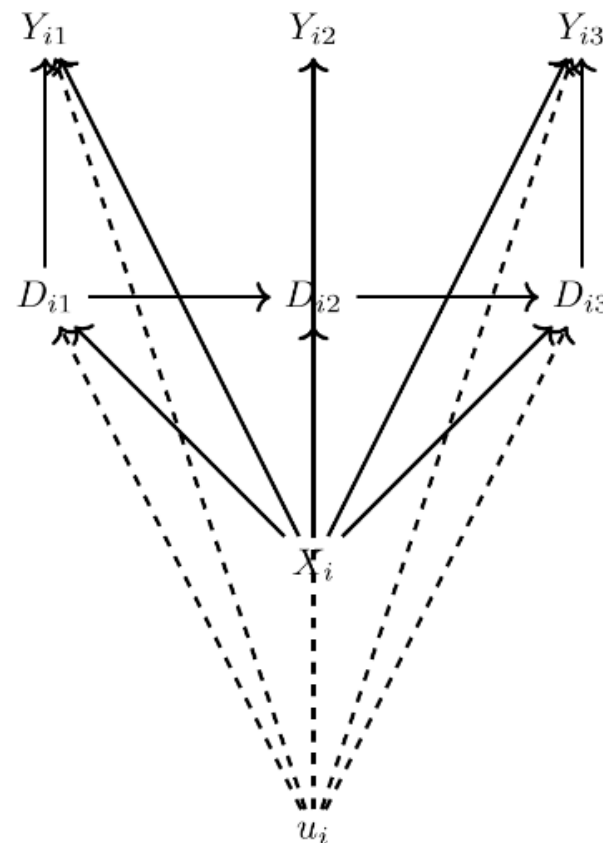
- 结果变量  $Y_i$ 
  - 包含 3 期, 分别是  $Y_{i1}, Y_{i2}, Y_{i3}$
- 协变量矩阵  $D_i$  随时间推移而变化  $D_{i1}, D_{i2}, D_{i3}$
- 随个体变化但不随时间变化, 可被观测的  $X_i$
- 随个体变化但不随时间变化, 但无法被观测到的  $u_i$ 
  - 无法观测
  - 随个体而变动
  - 给定个体  $i$ , 不随时间推移而变化





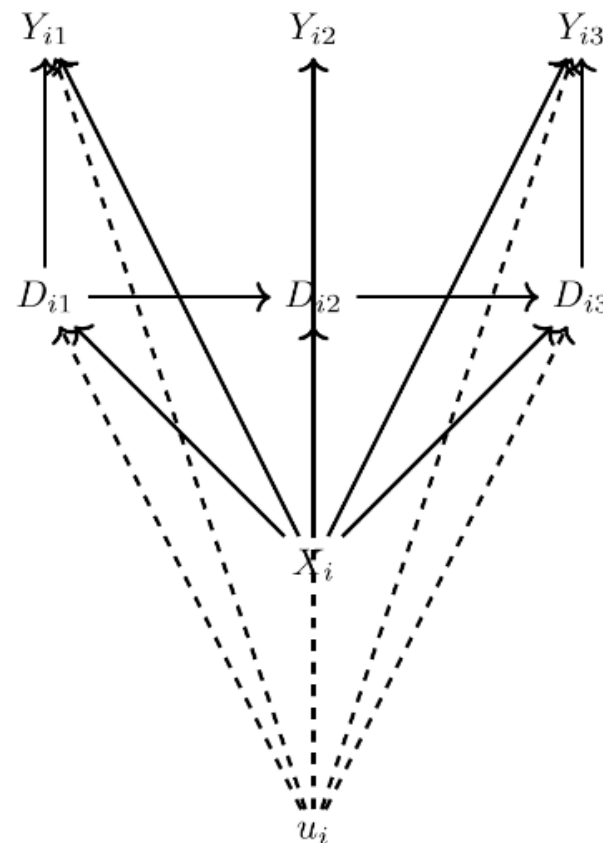
## 纵贯数据的有向无环图 (Imai 和 Kim, 2017)

- 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)
  - $\varepsilon_{it}$  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  $\delta$  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  $\eta_{it}$  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，要该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  $\varepsilon_{it}$  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  $\beta_2$  和  $\beta_3$ , 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  $\widehat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$  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  $\varepsilon_{it}$  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

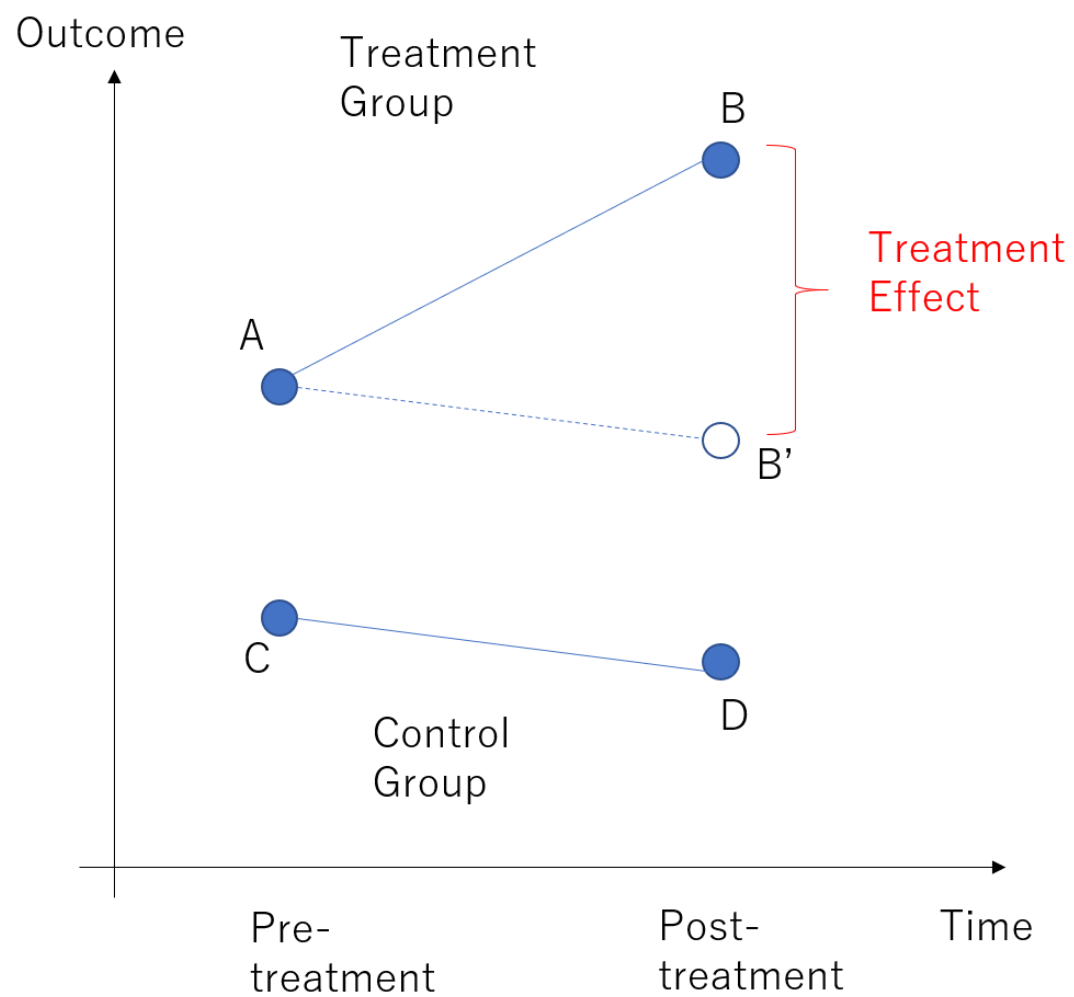
DiD



## 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

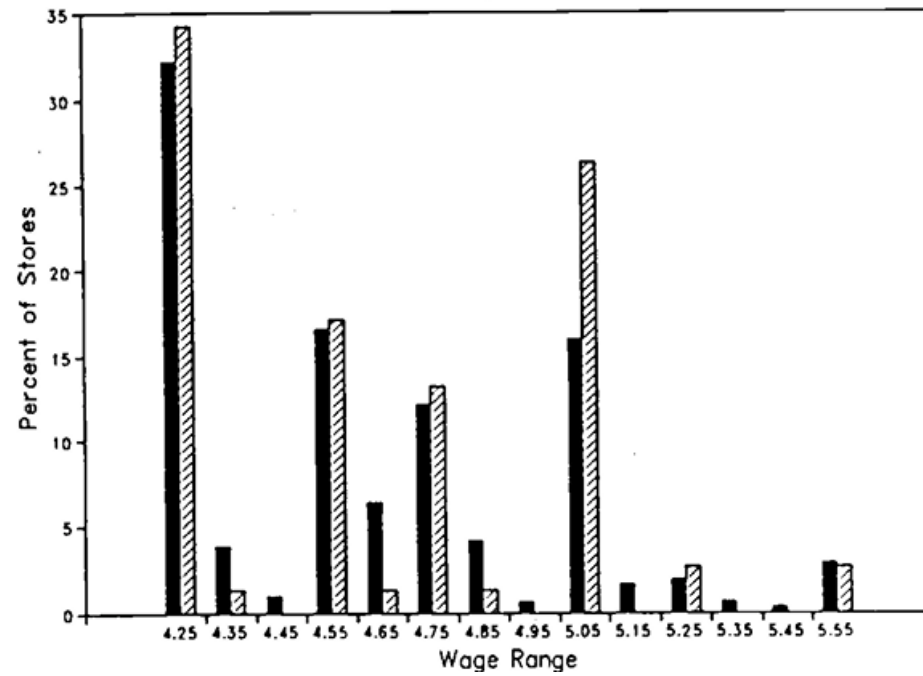
$$\begin{aligned}
 \hat{\delta}_{kU}^{2 \times 2} (1) &= (\bar{y}_k^{\text{post}(k)} - \bar{y}_k^{\text{pre}(k)}) - (\bar{y}_U^{\text{post}(k)} - \bar{y}_U^{\text{pre}(k)}) \\
 (2) &= (E[Y_k \mid \text{Post}] - E[Y_k \mid \text{Pre}]) - (E[Y_U \mid \text{Post}] - E[Y_U \mid \text{Pre}]) \\
 (3) &= \underbrace{(E[Y_k^1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

- 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）

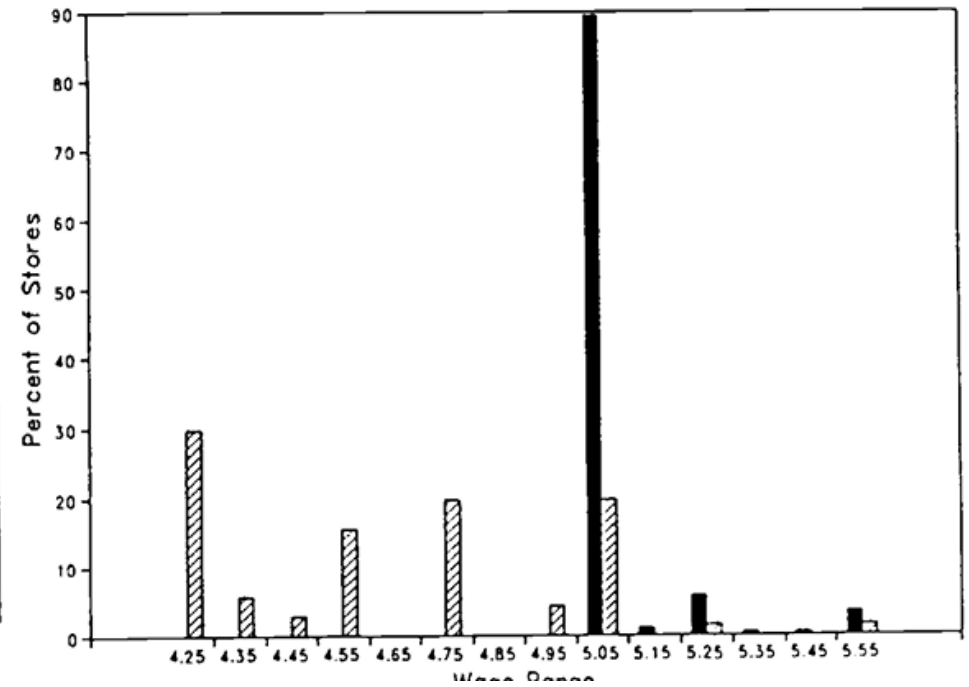
## 例子: Card and Kruger (1994, AER)

- 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?



## 估计方法 1: 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

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

## 估计方法 2 : 线性回归

- 除了满足**共同趋势假设**外，还希望：
  - 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 DiD 的估计精度
- 假设州固定效应和时间固定效应为常数，将 2×2 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

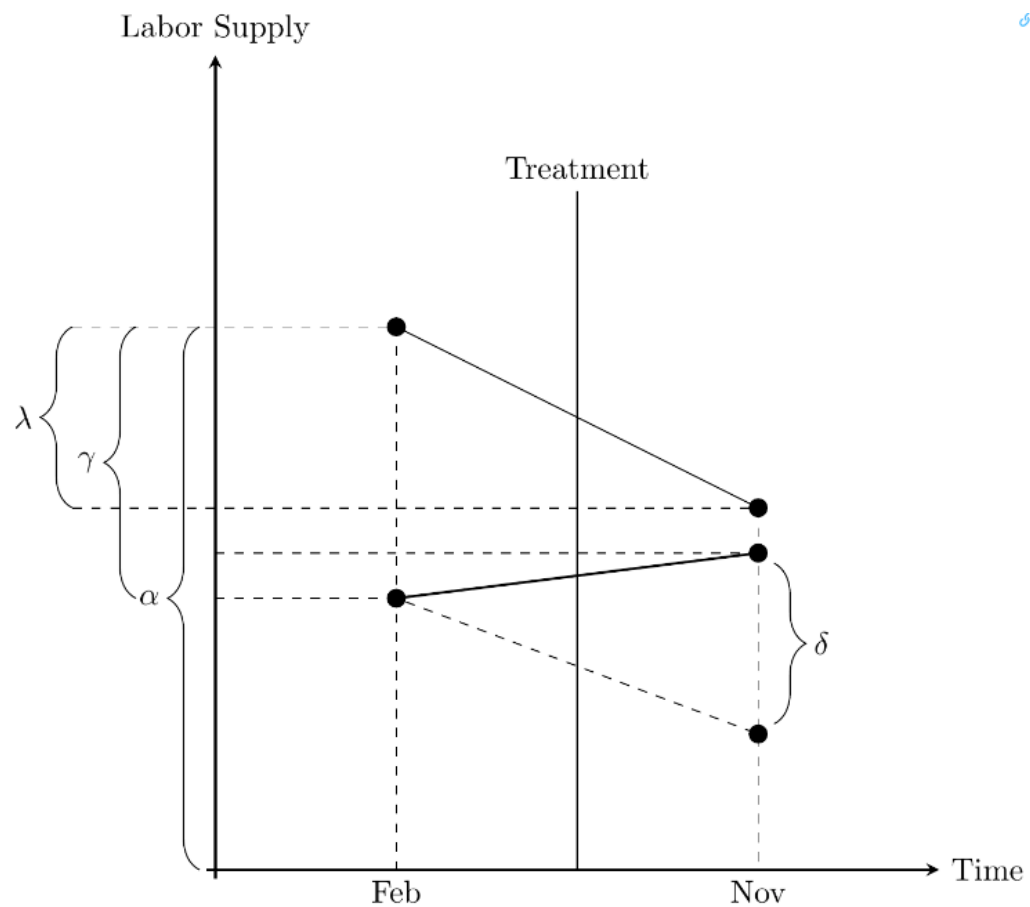
1. PA Pre:  $\alpha$

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

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

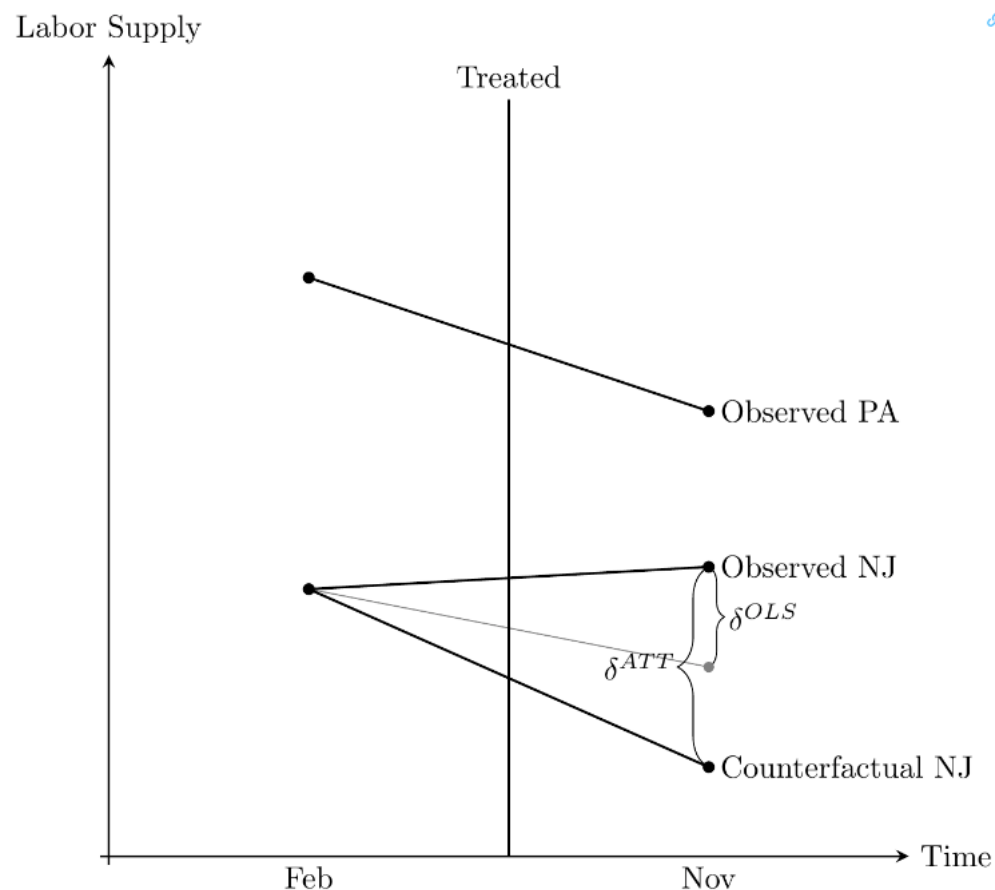
4. NJ Post:  $\alpha + \gamma + \lambda + \delta$

## DiD回归的理论图示



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

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



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

## 推断

CK1994 仅仅使用了 1 期的处理前和 1 期的处理后时期的数据

- CK1994 很可能每个州的就业不仅州内相关，而且是序列相关
- Bertrand (2004) 职场剧，传统标准误往往低估估计量的标准差，因此标准误向下偏误，变得“过小”，从而导致会过度拒绝原假设，因此建议：
  - 块状自助法标准误 (block bootstrapping)：选择的“块”是州或者省级，只需对州或者省进行替换
  - 将数据聚合 (aggregation) 为 1 个前期和 1 个后期：完全对时间维度进行忽略，只有前期、后期和一个控制组。
    - 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 中群组的数量只有 2 个 → 更为常见使用

# 对共同趋势假设的讨论

## 对共同趋势的诊断

- 共同趋势假设在不同情况下可能会被违反
- 如果处理状态取决于**随时间推移而变化的因素**（time-varying factors），处理组和控制组的结果可能具有**异质性的时间趋势**
- **注意**：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：基于随时间变化的因素自我选择进入治疗
  - 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

## 诊断共同趋势假设满足的逻辑难点

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

- 第二行中  $E[Y_k^0 | \text{Post}]$  是反事实项，无法直接估计
  - 实践中，通常比较**安慰剂预处理前的 DiD 系数**  
→ 潜台词是“如果它们之前相似，那么为什么处理后就不能是这样呢？”
- 但仅仅因为之前相似，在逻辑上不能得出在之后也是相似的
  - 如果假设未来和过去一样 → 就犯了“赌徒谬误”的错误
  - 但是，连假设都不设定，结果更**无法检验**
- 另外，如果处理本身是内生的，明显违背共同趋势假设 → 因为处理状态分配将直接取决于潜在结果，假设没有处理发生，潜在结果一定会改变



## 诊断 1: 检查处理组和控制组在处理前的平衡

- **事件研究**: 能够展现出处理组和控制组的个体在处理前的时期内, 具有动态可比性
- **展示原始数据**: 如果数据中有些组在同一时间接受处理, 就可以这样组 → 然后目测处理组在处理前的动态是否与控制组不同
- 如果没有统一的处理时间, 即不同的个体在不同的时刻接受处理
  - **逐年绘制原始数据, 简单肉眼大量**
    - 缺点是: 1.处理组数量大, 难实现; 2.不美观; 3.必须假设**唯一的控制组是从未被处理的组**
      - 针对第 3 点, Goodman-Bacon (2019) 认为该假设并不成立
      - **任何 DiD 都是处理与未处理、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**
      - 仅仅显示与从未接受过处理的个体比较, 实际上用了具有不同时间序列TWFE模型, 是误导性的表述
- **注意**: 对于共同趋势假设只能**诊断**, 而不能直接检验假设!
  - 记住: **无法检验平行趋势假设**
  - 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”

## 事件分析的设定(会有专门的章节)

- 在 DiD 模型中包含**先期**和**滞后**，可以检查时点后处理的动弹过程

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

- 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**
  - 共同趋势假设本身不可检验, 而事前平行趋势只是整个共同趋势假设的一部分
  - 即使事前平行趋势通过检验也只是表明处理组和控制组在处理发生前保持相同时间趋势, 并不能确保事后趋势也一定有共同趋势
  - 文章中不应该出现“事前平行趋势检验通过, 平行趋势假设成立”的说法

Sun&Abraham (2018) 指出存在当处理效应存在异质性时, 共同趋势检验可能存在偏误; Liu (2021) 分享了一种应对异质性处理效应的稳健作图方法。关于异质性处理效应在 DiD 扩展部分中还会讨论

## 应用：《平价医疗法案》

- 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

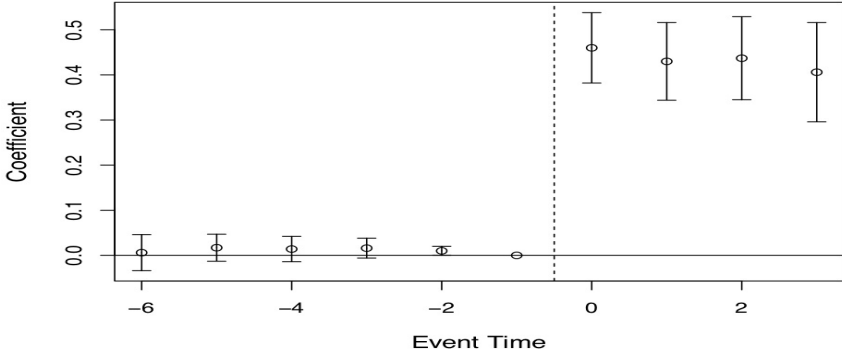


Fig 2 :法案对覆盖范围的影响

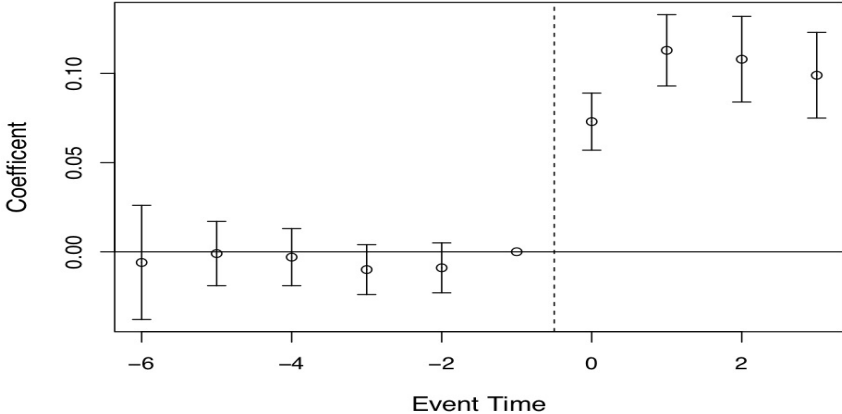


Fig 3 :法案对未参保率的影响

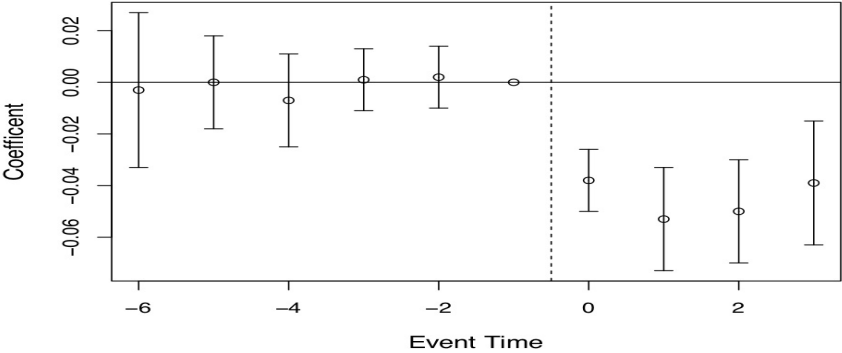
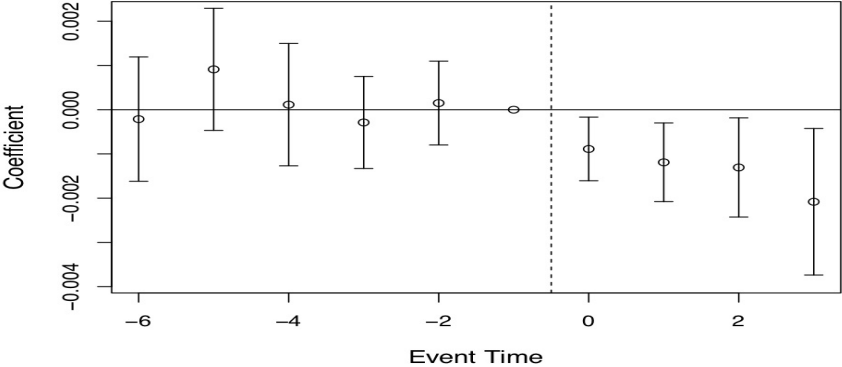


Fig 4 :法案对年死亡率的影响



## 事件分析的图示效果显著

- **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
  - 虽然不像处理前的系数那么精确，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

## 事件分析的图示效果显著 + 机制分析

- **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**
  - 文章还继续分离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

# 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
- DiD 是一个估计量，更是一种研究设计
- 作为估计量，估计的是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能正确识别因果效应
  - 共同假设经过变形就是DiD下的外生性假设  $E(\Delta \varepsilon | D) = 0$
  - 作为估计量的DiD没有解决内生性问题，而是“假设”不存在内生性问题
- 作为研究设计，DiD追溯到物理学家 John Snow 对伦敦霍乱成因的研究（Snow, 1855），CK1994 是这种思想的延续
  - 没有研究设计的“双重比对”的想法，是不会产生DiD估计量
    - 从研究方法史的角度看，是有了双重对比的研究设计，才有DiD
- **错误的说法**：DiD估计量，就自然有了对应的研究设计，就可以直接避开内生性问题
  - 这是不正确的。DiD解决内生性问题，本质上仍然依赖于干预或政策冲击本身的外生性



## 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

## 控制变量

在回归方程中加入控制变量起到两个作用。第一，保证条件独立假设

(conditional independence assumption, CIA) 成立。①条件独立假设成立意味着给定控制变量时处理变量 $iD$ 与误差项 $i\epsilon$ 不相关，从而保证了OLS估计量 $b$ 是我们所关心的因果效应 $\beta$ 的一致估计。这是观测性研究的因果推断中控制变量所发挥的最核心作用。第二，减小误差，提高估计精度。如果处理变量 $iD$ 与误差项 $i\epsilon$ 已经不相关，无论是否加入控制变量， $b$ 都是因果效应 $\beta$ 的一致估计。此时加入合理的控制变量可以降低误差从而提高估计精度。当然，作为经验法则而言这种判断标准有些过于严苛了，在一些极少见的特殊情况下控制事后变量可能有助于纠正选择性偏误（Cinelli, 2021），不过绝大多数情形下控制事后变量都需要非常谨慎。②作者的建议是对控制变量采取较为保守的策略，除非有很强的理由认为某变量会导致严重的选择性偏误而必须加以控制，否则一般不加入回归方程。即使是必须控制的变量，也要注意选择干预时点之前的前定变量。

## SUTVA

单位处理变量值稳定假设（SUTVA）。单位处理变量值稳定假设（stable unit treatment values assumption, SUTVA）是指不同个体是否受到政策冲击是相互独立的，某一个体受政策冲击的情况（treatment status）不影响任何其他个体的结果。直观理解，不满足SUTVA意味着控制组个体也受到了干预政策的影响，因而不不再是事实上未受干预影响的“真实”控制组，也就无法使用控制组时间趋势来构建处理组时间趋势的反事实。在理想情况下，处理组和控制组被严格区分开来，彼此互不干涉，然而，在现实生活中，相当多的政策冲击具有一定的外部性，例如加强上游省份水污染企业的环境督查也会有利于改善下游省份水质。此外，个体的行为也往往具有一定的策略性和选择性，如处理组地区得到了较好的政策帮扶，那么原本控制组地区的个体可能会自发从控制组地区迁移至处理组地区，意味着宏观上非政策目标地区也受到了干预政策的影响，这就是通常所说的一般均衡效应（general equilibrium effect）或溢出效应（spillover effect）。一般均衡效应或溢出效应会使得SUTVA不再成立，进而导致DiD法无法正确识别因果效应

## 使用 DiD 的注意事项（1）：制度背景和政策实施真实情况

DiD法应用最多的场景是评估政策效应。对于制度背景的清晰梳理和政策真实实施情况的正确观察应该是政策评估类实证研究的基石。一项政策可能发布了却没有很好地实施，也可能受政策影响的个体采取了“上有政策，下有对策”的策略式行动影响了政策实施真实效果，如果研究者没有很好地厘清这些制度背景和政策实施的真实情况，就不可能准确地评估政策效应，甚至可能得到误导性的研究结论。这里举一个实例。相当多的研究发现地方政府的财政补贴相当低效，企业获得了大量的财政补贴却并没有激励企业的研发创新能力，甚至会引起企业寻租（王红建等，2014；张杰等，2015）。然而，范子英和王倩（2019）通过对地方政府税收征管实务的观察，发现财政补贴实施过程中存在相当明显的“列收列支”问题：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。所以，相当一部分名义上为财政补贴的资金实际上是企业自有资金，而这部分“虚假”的财政补贴自然不会对企业经营行为产生影响。因此，财政补贴的低效率很可能是由于对政策实施真实情况的把握不够深入导致的错误结果。总体而言，使用DiD法评估政策效应要求对政策的具体实施情况有深入、清晰的了解：政策什么时候开始真正实施？政策是否按照要求得到了准确执行？行为主体是否采取了一些应对措施？等等。这一系列问题与双重差分法是否合理、可行程度密切相关，也是进一步深入分析政策机制效应的良好开端。因此，政策评估类的实证研究有必要高度重视制度背景和政策实施情况。

## 使用 DiD 的注意事项（2）：干预政策需要严格外生或随机分配吗？

在第二部分DiD法的识别假设部分，我们强调了DiD法本身并没有解决内生性问题，而是“假设”干预政策是外生，内生性问题的解决仍然依赖于干预政策本身的外生性。然而，这里的外生性是什么意义上的外生性？换言之，DiD法下需要干预政策和谁之间是外生的？一种看法认为干预政策必须是完全随机（自然实验）或者近似随机分配（准自然实验），即干预政策和模型未考虑的所有因素（扰动项）之间不相关，只有在这种情况下才适用DiD法（陈林和伍海军，2015）。但是，现实中的任何一项政策几乎都有特定的政策目标和政策对象，完全随机分配的政策几乎并不存在，那么这类政策是否完全不适用DiD法呢？本文认为并非如此。第二部分对识别假设的讨论清楚地表明，DiD法所需要的外生性是干预政策和扰动项在差分意义上的外生性，这与水平意义上的外生性显然并非是等价的。①我们以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要求贫困县名额的分配过程要近似完全随机，无论是贫困地区还是富裕地区都有差不多的机会入选贫困县，显然这并不符合现实——贫困县的选取标准主要是人均GDP、人均财政收入等指标，被选为贫困县的地区都是经济发展十分落后的县域，因此贫困县政策并不满足水平意义上的外生性。但是，差分意义上的外生性是有可能满足的，即贫困县可能和非贫困县有相同的经济发展趋势。如果研究设计能够尽量满足这一识别假设，就可以使用DiD法。例如黄志平（2018）的做法是首先使用倾向得分匹配法（PSM）对数据预处理，在非贫困县中尽量选取与贫困县的各方面禀赋条件类似的控制组，从而尽可能地使得平行趋势假设成立（等价于差分意义上的

## 使用 DiD 的注意事项 (3)：溢出效应

DiD法的另一个核心识别假设是 SUTVA，即干预不存在一般均衡效应或溢出效应。然而，现实中的各项政策几乎或多或少都会存在一定的一般均衡效应，例如前文提到的上游省份加强水质环境规制会影响下游省份水质的例子。特别是在长期中，当处理组个体的决策发生变化时，控制组个体一定会随之调整自身的行为决策。因此，干预政策是否存在溢出效应是任何一个使用双重差分法的实证研究必须考虑的潜在威胁。不过，检验溢出效应是否存在并非一项简单的工作，研究者需要根据制度背景仔细识别可能受到溢出效应影响的控制组个体，而后检验溢出效应。Lu 等（2019）研究中国经济开发区对当地经济发展的影响，其对溢出效应的讨论和处理是一个较为成功的范例。他们采取了两种识别策略检验溢出效应，第一种是检验与经济开发区所属村庄邻近的同县其他村庄经济发展是否也得到了提高，第二种是检验经济开发区对经济发展的激励效应是否随着村庄离经济开发区越来越远而减弱。第一种方法的结果表明同县其他村庄的总产出、就业等仅有略微的提高且统计上不显著，第二种方法的结果表明距离经济开发区 2 千米之外的村庄基本上不受经济开发区的影响，两种方法都提供了证据表明经济开发区政策的溢出效应并不显著。还需要强调的一点是，如果研究重点本身就是政策的溢出效应的话，那么是不适用 DiD 法的。例如一些研究试图探讨地区产业政策对企业选择效应和集聚效应的影响：本地拥有更加优惠的产业政策

（如税收优惠）会吸引相邻地区的企业迁移到本地区，产生选择效应和集聚效应。这里的选择效应和集聚效应就是溢出效应的一个典型表现：本地区的政策对邻近地区的企业产生了

## 使用 DiD 的注意事项（3）：一般均衡视角下的成本收益分析

DiD法广泛应用于各类公共政策的评估，如果估计得到了政策效应符合预期，是否就意味着政策达到了初始目标或是政策本身就是有效的呢？不是。一般而言，DiD法只能评估干预政策对研究者感兴趣的结果变量的影响，但研究者并不清楚政策本身的机会成本有多大，也不清楚政策的净收益到底是多少。评估政策效应整体上是否符合预期或是政策是否有效率，并不能仅根据估计结果就判断政策是否有效，而是需要从更广泛的一般均衡角度，从整体上对政策进行成本收益分析。Duflo（2001）是在政策效用评估类文献中成功应用成本收益分析的早期经典代表，她研究了印度尼西亚修建学校对当地儿童的长期劳动力市场的影响。根据DiD法的基准结果，她估计了印度尼西亚政府投资学校建设的成本和对儿童未来的工资收益，发现投资学校建设的内部回报率为 8.8%-12%，远高于当地实际利率，因此投资教育是一个非常高收益的投资项目。①Lu 等（2019）对中国经济开发区的政策效应同样进行了成本收益分析，他们根据DiD法的估计结果计算得到 2006-2008 年间经济开发区为当地居民和企业提高的工资和利润总额约为 1 807 亿元，付出的税收成本则为 558 亿元，净收益高达 1 249 亿元。上述例子都体现了研究者在一一般均衡的视角下，从



DiD扩展

——具有时间异质性的双向固定效应



## 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

## DiD 扩展-交错双重差分法 (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

## DiD 扩展-广义双重差分法 (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

## DiD 扩展-队列双重差分法 (cohort DiD)

- 也被称为截面双重差分法，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；另一个维度为**出生队列间差异**，标识**个体**是否受到了干预政策的影响
- 队列双重差分法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，近年来使用这一方法的代表性文献有 Chen等 (2020)

## DiD 扩展-模糊双重差分法 (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
- 但是，有时冲击并未带来急剧 (sharp) 变化，所谓的“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，而所谓的“控制组”中也并非完全没有受到冲击，即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) 详细介绍了该方法，重新评估了印度尼西亚的教育回报率

# 从 DiD 到 Event Study

## ——具有时间异质性的双向固定效应

## 事件分析の設定

- 在 DiD 模型中包含**先期**和**滞后**，可以检查时点后处理的动弹过程

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

- 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**
  - 共同趋势假设本身不可检验, 而事前平行趋势只是整个共同趋势假设的一部分
  - 即使事前平行趋势通过检验也只是表明处理组和控制组在处理发生前保持相同时间趋势, 并不能确保事后趋势也一定有共同趋势
  - 文章中不应该出现“事前平行趋势检验通过, 平行趋势假设成立”的说法

- 共同趋势假设，即如果政策没有发生，处理组和控制组的结果变量具有相同的时间趋势
- 事件研究法能够直观地观察和检验政策发生前后个体行为的动态反应与组间差异，因此大量使用双重差分法的实证研究选择借助事件分析法来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。
- 由于使用事件分析法来检验事前平行趋势是 DiD 的标准步骤，但其作用远不止检验事前平行趋势
- 事件研究法可以得到更加丰富的政策效应信息。双重差分法只能估计政策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。并且，由于存在一般均衡效应，一项政策的短期和长期效果可能存在比较明显的差异。上述政策效应的动态特征是双重差分法无法说明的，而事件研究法则能以图形形式清晰地展示政策效应动态变化，这是事件研究法最吸引人的特征(Miller, 2023)。
- 第二，事件研究法的识别假设相较于双重差分法更为宽松。近年来理论计量的新进展表明，对于政策交错发生的情形，若处理效应在组群或时间维度上存在异质性，双重差分法的估计结果存在偏误。事件研究法能够很好地应对时间维度上的异质性处理效应，比双重差分法的适用情景更为广泛。
- 如何检验事前平行趋势、如何选择基期、对非平衡面板数据是否需要做归并或截断处理等



## 使用事件分析进行因果推断的前提

- **严格外生假设**（面板数据的基本假设，DiD+TWFE+ES）
  - 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
  - 即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，某一个体受政策冲击的情况不影响任何其他个体的结果
- **同质性处理效应**（异质性处理效应的修正是近期热点）
  - 组别间是同质，即同一政策对于不同处理组的影响是相同的
  - 时间维度上同质，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

## 诊断 2：事前处理的安慰剂效应证伪

- 想法是：以最低工资的研究为例，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？假如设定一个**备择假设**（最低工资对就业率没有影响），那么试着去检验该备择假设
  - 通过新的研究设计，**不能拒绝**该备择假设，那么就**有理由怀疑原始结论**的可信性
  - 通过新的研究设计，**拒绝**该备择假设，那么就为**原始结论**提供了可信度
- CK1994 的安慰剂证伪分析中，使用了另一类工人的数据（选择不太可能被最低工资政策影响的工人）
  - 不考虑一般均衡效应？什么意思，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

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

- $\gamma^{\tau}$  应该接近 0

## 作为安慰剂效应证伪的 DDD

- DDD基本思路引入了第三个维度“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
- 应用场景通常有2个
  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
  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

## 作为安慰剂效应证伪的 DDD

- 在 CK1994 的先前分析中，我们假设 NJ 在通过最低工资标准后发生的这件事情是一个 **共同冲击 T** (common shock) → PA 仅仅是因为随机因素导致没有接受到政策干预
- 但如果有一个特定于州的时间冲击， $NJ_t$  和  $PA_t$  呢？
  - 处理组和控制组就会有不同的时间发展路径
  - 不再满足共同趋势的假设
  - 经典 DiD 无法准确估计 ATT
- 对于这种特定于州的时间冲击导致的共同趋势假设不满足，可以利用 DDD 模型估计
- 在最低工资增加之前，NJ 的底薪和高薪就业是由特定群组的 NJ 固定效应 ( $NJ_h$ ) 决定的，而 PA 是 ( $PA_h$ )
- 最低工资上调后，NJ 发生 4 种变化；除了最低工资的影响，PA 也有属于自己的特定州效应
  1. 国家趋势导致就业变化的 T
  2. NJ 的州特定时间冲击  $NJ_t$  ( $PA_t$ )
  3. 低工资劳动者就业变化  $l_t$
  4. 最低工资产生的未知影响  $D$

# 作为安慰剂效应证伪的 DDD：基于CK1994 的设计

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

例如研究贫困县政策对经济发展的影响时，由于贫困县依据人均GDP等经济指标来认定，被划为贫困县的地区经济发展速度很可能原本就比非贫困县更慢，处理组（贫困县）和控制组（非贫困县）之间的经济发展状况很难满足平行趋势。一个可能的选择是加入组间线性趋势  $D_i \times Trend_t$  以控制组间线性时间趋势的差异，从而缓解这一问题。图4a给出了数值模拟的证据，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法估计出的处理效应存在明显偏误，但控制组间线性时间趋势后就能准确地估计处理效应。事实上，根据上述的分析，在DiD法中额外地控制住组间线性趋势可以作为一种稳健性检验：若平行趋势假设满足，那么是否加入组间线性时间趋势不会对估计结果产生明显影响；反之，若估计结果发生了明显改变，则预示着组间时间趋势可能存在差异，平行趋势假设可能并不满足

## 为什么 DDD 可以用来安慰剂证伪

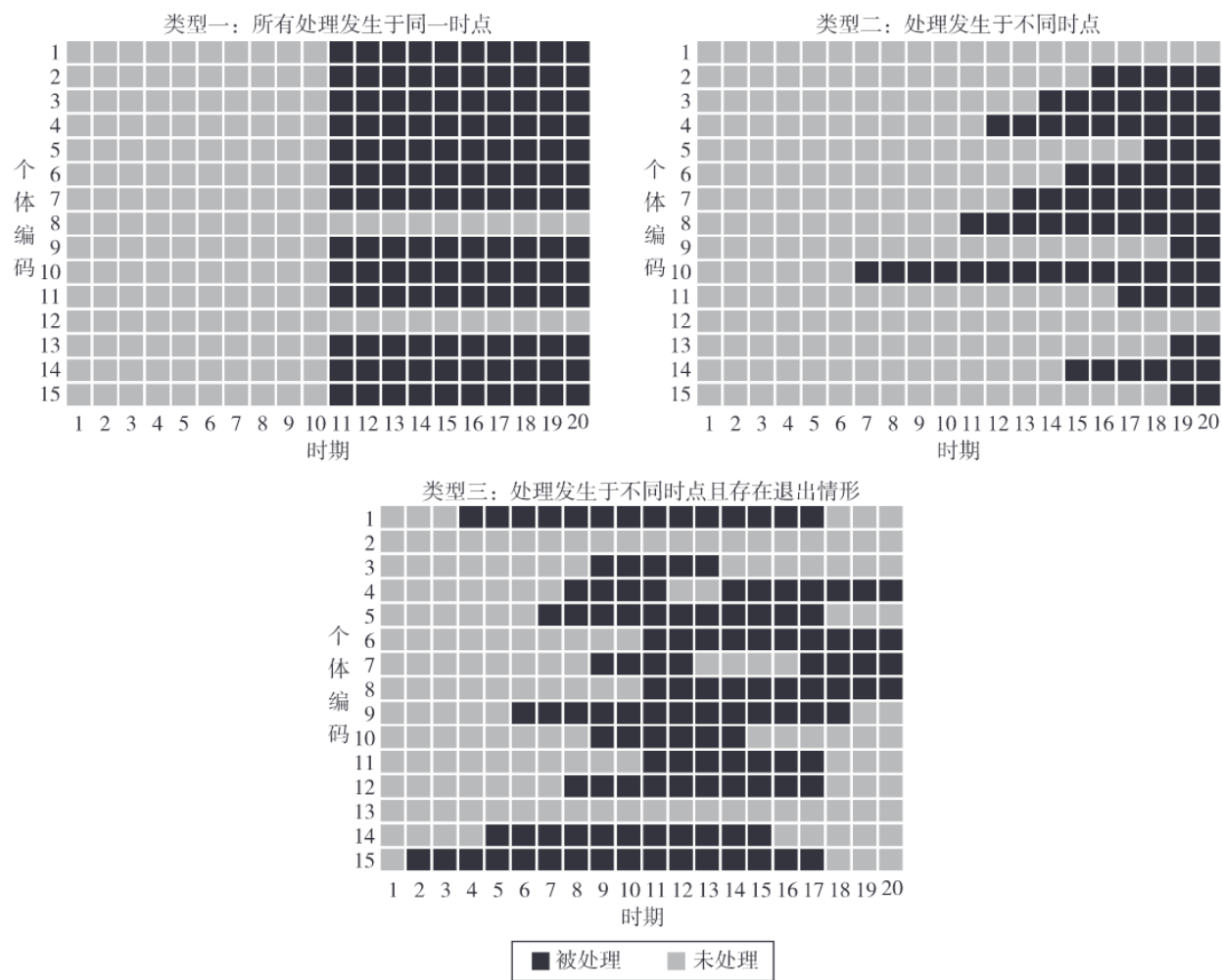
然而，控制组间时间趋势也是一把双刃剑，可能会产生一些不合意的后果。第一，组间线性时间趋势  $D_i \times Trend_t$  和DiD的核心解释变量  $D_i \times Post_t$  的构造方式相似，因此二者存在比较明显的共线性，控制组间线性时间趋势会大大减少核心解释变量的变动程度从而降低估计效率、提高标准误。从图 4a 中可以发现加入线性时间趋势后的估计系数分布明显更加分散，这表明估计量效率降低、标准误变得更大了。第二，如果处理效应不是一次性的，而是随着时间推移逐步显现出来，那么组间线性时间趋势会吸收一部分处理效应，导致DiD法会低估真实效应。图 4b 的模拟结果说明了这一点：在处理效应存在动态变化时，加入组间线性时间趋势会大大低估真实的处理效应。因此，是否控制组间时间趋势需要研究者结合具体的研究情景仔细斟酌。

## 为什么 DDD 可以用来安慰剂证伪

从本质上看，组间时间趋势存在差异的根本原因是存在某些可观测或不可观测的前定变量在处理组和控制组之间存在差异或者是存在随时间变化的混淆因素。比如前面提到的贫困县的例子，贫困县和非贫困县的经济发展趋势差异是由当地的初始经济发展水平、地理条件、文化等一系列因素综合造成的。对于可观测的因素，可以通过添加控制变量的方法加以控制，但对于不可观测的因素则一般很难直接处理，通过控制组间线性趋势差异可以部分缓解这一问题，然而当组间时间趋势差异和动态处理效应同时存在时也无法完全解决这一问题。针对这种复杂情况，目前主要有两种处理思路。一种思路是在DiD的框架下，通过使用未受处理的样本来更为干净地估计和剔除掉时间趋势。①另一种思路可能需要超越DiD法，寻找工具变量或使用空间断点回归设计等方法，不过这些问题超出了本文的范围，这里不再加以讨论。



# DiD 根据政策时点进行划分



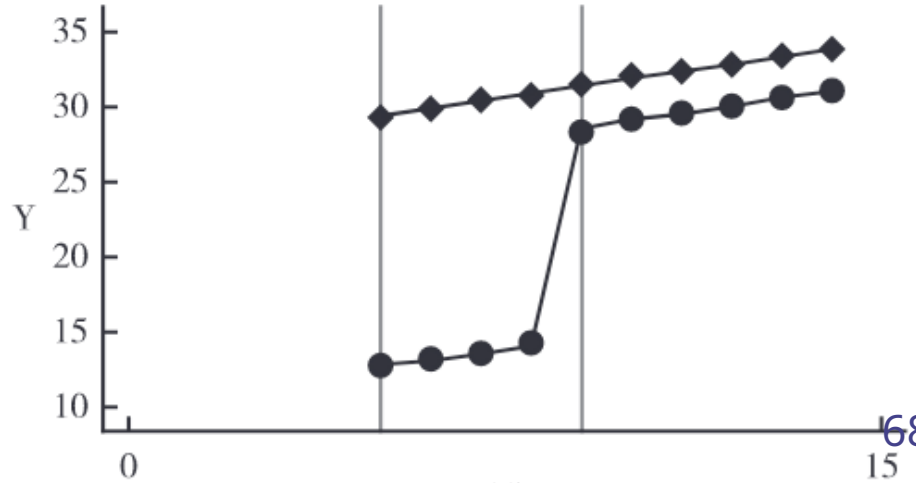
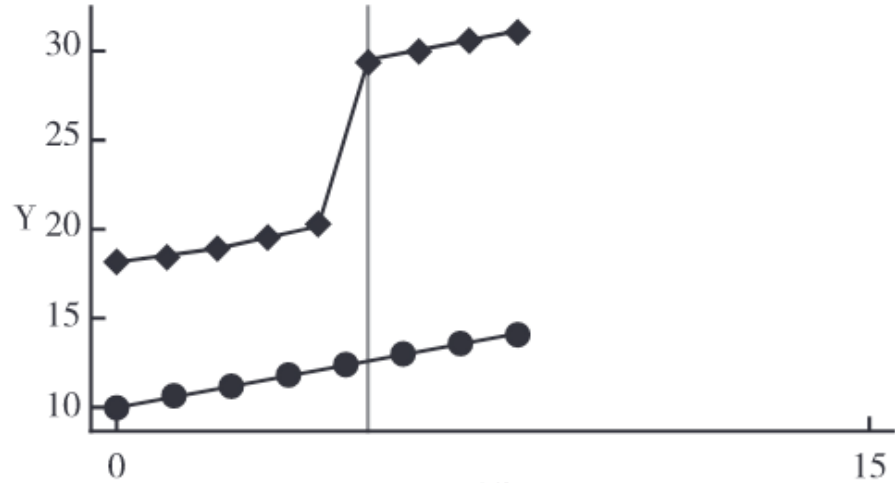
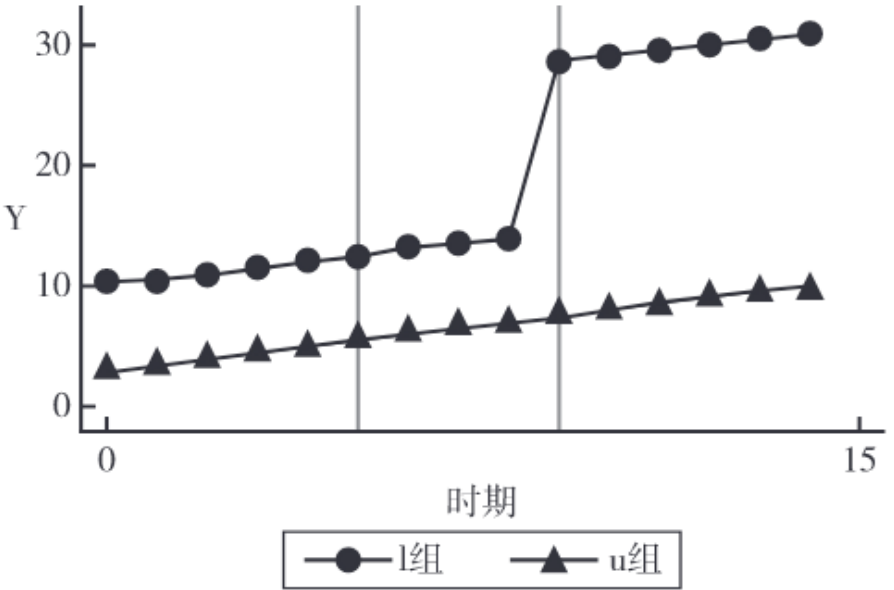
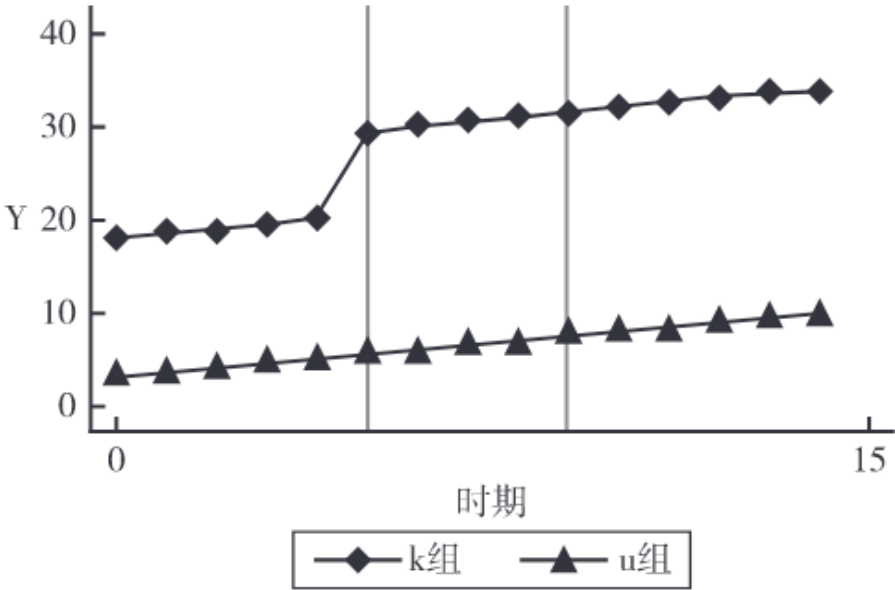
## 存在异质性处理效应会导致 TWFE 估计量出现偏误

- 同一处理对于不同个体产生的效果存在差异，这种差异可能表现在接受处理后的时长或者不同时间接受处理的组别两个维度
- 在**静态情形**下，GB(2021) 指出即使“共同趋势”假设满足，TWFE也会存在“坏的控制组”问题
- 在**动态情形**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，甚至还会面临平行性趋势检验失效的风险（SA2021）

## TWFE在估计静态模型时的潜在问题

- GB(2021)理论上指出，当处理效应存在异质性时，同一处理对于不同个体产生的效果存在差异，将会导致TWFE产生潜在的估计偏误
- 假设存在k, l, u三组个体，k组个体最早接受政策处理（第5期），l组个体相对较晚（第9期），u组个体是从未接受处理组
  - “坏的控制组”：较早接受处理组 → 事前趋势已经发生了变化
  - “好的控制组”：从未接受处理组 + 相对较晚接受处理的组
- 传统单时点DiD中仅存在k组或l组中的一个，TWFE没有偏误

# TWFE在估计静态模型时的潜在问题



## 静态模型与 Goodman-Bacon 分解

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

- VWATT“方差加权平均处理效应” (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  $2 \times 2$  DiD 的处理组和控制组之间的平行趋势加权平均得到的
- $\Delta ATT$  “加权的处理效应变化”, 仅使用“坏”控制组（较早接受处理组）时出现
- 共同趋势假设满足  $\rightarrow VWCT = 0$  , 静态TWFE仍可能有偏且不一致的
  - 若**处理效应个体间异质性**, 那么  $\Delta ATT = 0$  , 但此时 VWATT (方差加权ATT) 与ATT (样本加权) 不一致
  - 若**处理效应在时间维度上是异质性**, 那么  $\Delta ATT \neq 0$  尽管此时VWATT与ATT (样本加权) 相同, 但  $\Delta ATT$  的存在使得TWFE 仍然是有偏且不一致的
  - 若**处理效应在个体和时间上均存在异质性**, 那么VWATT 不等于ATT (样本加权) 和  $\Delta ATT \neq 0$  都可能导致 TWFE 有偏且不一致

## TWFE在估计动态模型时的潜在问题

- 使用动态双向固定效应模型 (Dynamic TWFE) 进行事件研究做法是, 在样本期  $[\underline{T}, \overline{T}]$  内加入一系列表示相对于接受处理时点时长的虚拟变量, 采用如下设定回归:

$$y_{it} = \alpha_i + \lambda_t + \sum_{\substack{r \neq -1 \\ -\underline{T} \leq r \leq \overline{T}}} 1[R_{it} = r] \beta_r + \epsilon_{it}$$

- $1[R_{it} = r]$  是指示变量, 指示个体  $i$  观测时点  $t$  距离接受处理时点是否为  $r$  期,
- 实际操作中, 研究者一般会根据样本总时期长短对  $r$  进行部分归并 (Bin) 处理, 并且为了避免多重共线性问题常会将 “-1” 期排除在外作为基期
- $\beta_r$  则是我们感兴趣的估计系数, 即动态平均处理效应 (Dynamic Treatment Effect)

## TWFE在估计动态模型时的潜在问题

- Sun 和 Abraham (2021)指出, 当存在异质性处理效应时,该模型估计系数可能会存在偏误并难解释
  - 类似静态估计,  $\beta_r$  在估计的过程中使用了“坏的控制组”而产生偏误
  - 系数  $\beta_r$  可能会赋予一些不属于  $r$  期的样本以正权重, 从而使得每一个  $\beta_r$  可能受到来自其他时期的系数  $\beta_s (s \neq r)$  的“交叉污染” (Cross-Lag Contamination)
    - 异质性处理效应会导致在使用动态 TWFE 时的共同趋势检验方法失效

## 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)



## 异质性处理效应的严峻性

假设所有接受处理的个体给定一个反事实结果  $Y(0)$ , 即如果该个体未受处理的潜在结果, 而个体观测到的真实结果记为  $Y$ ;  $Y$  和  $Y(0)$  的差值就是每个个体的处理效应  $\Delta$ 。我们假定有两种情形:  $\Delta^1$  假设处理效应是同质的, 即处理效应在各个体之间以及在不同时期是一个相同的常数;  $\Delta^2$  则假设处理效应是异质的, 即个体 A、B 和 C 的处理效应大小不同并且个体 B 和 C 各自的处理效应在不同时期也是不同的。在实证

## 经典事件研究法的研究步骤通常为

- 首先定义事件发生前后可能对结果变量产生影响的时间段（事件窗口）
- 然后计算事件窗口内的股票超额回报率
- 最后比较事件发生前后的股票超额回报率变动来识别事件本身对股票价格的影响
- 早期的事件研究法没有引入未受处理的样本作为对照，仅比较个体在事件发生前后的自身差异来识别因果效应。从研究设计的角度看，这种做法非常类似于以时间作为驱动变量的断点回归设计

# 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

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

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