

# 面板数据因果推断

## 第二讲：经典双重差分法：因果推断的视角

徐轶青

斯坦福大学

# 课程计划

- 第一讲：经典双重固定效应模型
- 第二讲：双重差分法：因果推断的视角
- 第三讲：对双重固定效应模型的重新认识
- 第四讲：对异质性稳健的新估计量
- 第五讲：合成控制法及其拓展
- 第六讲：潜因子法与双重稳健法

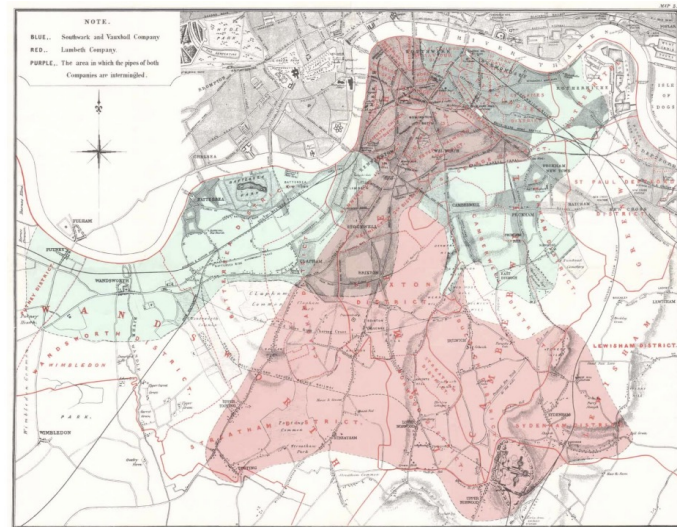
## 第二讲大纲

- DID设计的思想、历史和演进
- $2 \times 2$  DID设计的识别
  - 潜在结果框架
  - SUTVA和无预期效应
  - 选择性偏误与平行趋势假设
  - 对DID法的威胁
- $2 \times 2$  DID设计的估计
  - 点估计
  - 标准误估计
  - 大样本性质与统计推断\*

本讲内容部分借用了Pedro Sant'Anna教授讲义的内容。

# 约翰·斯诺医生与伦敦霍乱大流行

- 19世纪的伦敦霍乱大流行，而当时的人们并不知道霍乱的具体致病原因。一位名叫约翰·斯诺的医生着手调查此事。
- 当时，伦敦从两个自来水公司供水。
  - 其中一家在1852年，将水源从泰晤士河的靠下游处（在污水管下游）移到了上游。之后，使用这家自来水公司的住户，霍乱死亡率大幅降低。
- 斯诺医生因此认定，霍乱流行与被污染的水源有关。



Regions of South London Served by the Southwark & Vauxhall and the Lambeth Companies (Snow, 1955)

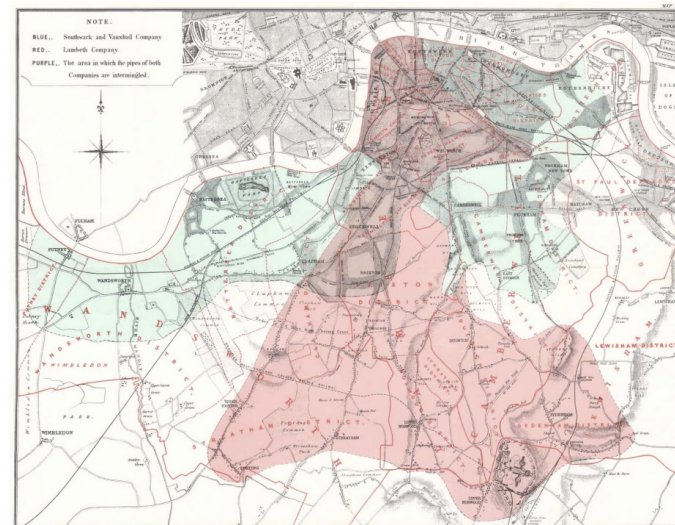
# 约翰·斯诺医生与伦敦霍乱大流行

Table 1:

John Snow's data on mortality from cholera in areas served by only one of the Southwark & Vauxhall or Lambeth Water Companies before and after the change in water source for the Lambeth Water Company. Rates for all time points were calculated based on 1851 population census

Water supply	Cholera deaths, 1849, rate per 100,000	Cholera deaths, 1854, rate per 100,000	Difference in rates comparing 1854 to 1849, rate per 100,000
Southwark & Vauxhall Company only	1349	1466	118
Lambeth Company Only	847	193	-653
Difference-in-difference, Lambeth versus Southwark & Vauxhall	502	1273	-771

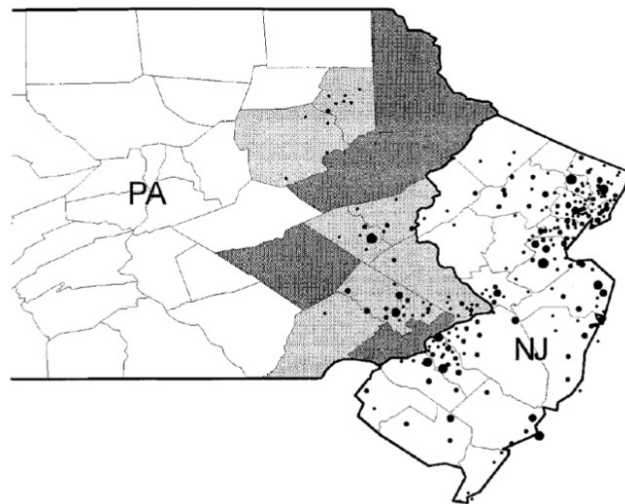
[Open in a separate window](#)



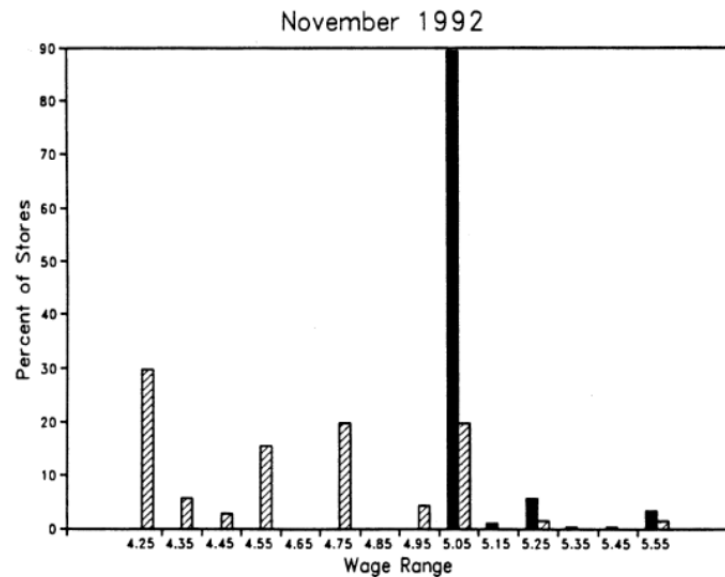
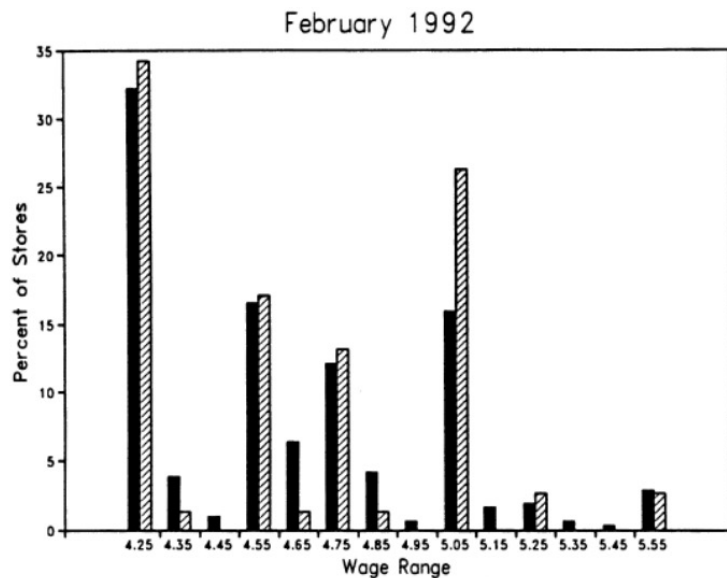
Regions of South London Served by the Southwark & Vauxhall and the Lambeth Companies (Snow, 1955)

# DID进入现代社会科学：最低工资和就业

- David Card和Alan Krueger将DID设计引入了经济学
- 传统经济理论认为，提高最低工资会抑制就业
- Card & Krueger (1994) 以1992年4月新泽西提高最低工资为契机，研究最低工资变化对快餐业就业的影响
- 样本：410家新泽西州和宾州东部的快餐店
- 两期问卷调查：工资、就业
  - 第一期：1992年3月
  - 第二期：1992年12月



# 工资分布



# 最低工资法的影响

TABLE 3—AVERAGE EMPLOYMENT PER STORE BEFORE AND AFTER THE RISE  
IN NEW JERSEY MINIMUM WAGE

Variable	Stores by state			Stores in New Jersey <sup>a</sup>			Differences within NJ <sup>b</sup>	
	PA (i)	NJ (ii)	Difference, NJ – PA (iii)	Wage = \$4.25 (iv)	Wage = \$4.26–\$4.99 (v)	Wage ≥ \$5.00 (vi)	Low– high (vii)	Midrange– high (viii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	–2.89 (1.44)	19.56 (0.77)	20.08 (0.84)	22.25 (1.14)	–2.69 (1.37)	–2.17 (1.41)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	–0.14 (1.07)	20.88 (1.01)	20.96 (0.76)	20.21 (1.03)	0.67 (1.44)	0.75 (1.27)
3. Change in mean FTE employment	–2.16 (1.25)	0.59 (0.54)	2.76 (1.36)	1.32 (0.95)	0.87 (0.84)	–2.04 (1.14)	3.36 (1.48)	2.91 (1.41)
4. Change in mean FTE employment, balanced sample of stores <sup>c</sup>	–2.28 (1.25)	0.47 (0.48)	2.75 (1.34)	1.21 (0.82)	0.71 (0.69)	–2.16 (1.01)	3.36 (1.30)	2.87 (1.22)
5. Change in mean FTE employment, setting FTE at temporarily closed stores to 0 <sup>d</sup>	–2.28 (1.25)	0.23 (0.49)	2.51 (1.35)	0.90 (0.87)	0.49 (0.69)	–2.39 (1.02)	3.29 (1.34)	2.88 (1.23)

Notes: Standard errors are shown in parentheses. The sample consists of all stores with available data on employment. FTE (full-time-equivalent) employment counts each part-time worker as half a full-time worker. Employment at six closed stores is set to zero. Employment at four temporarily closed stores is treated as missing.

<sup>a</sup>Stores in New Jersey were classified by whether starting wage in wave 1 equals \$4.25 per hour ( $N = 101$ ), is between \$4.26 and \$4.99 per hour ( $N = 140$ ), or is \$5.00 per hour or higher ( $N = 73$ ).

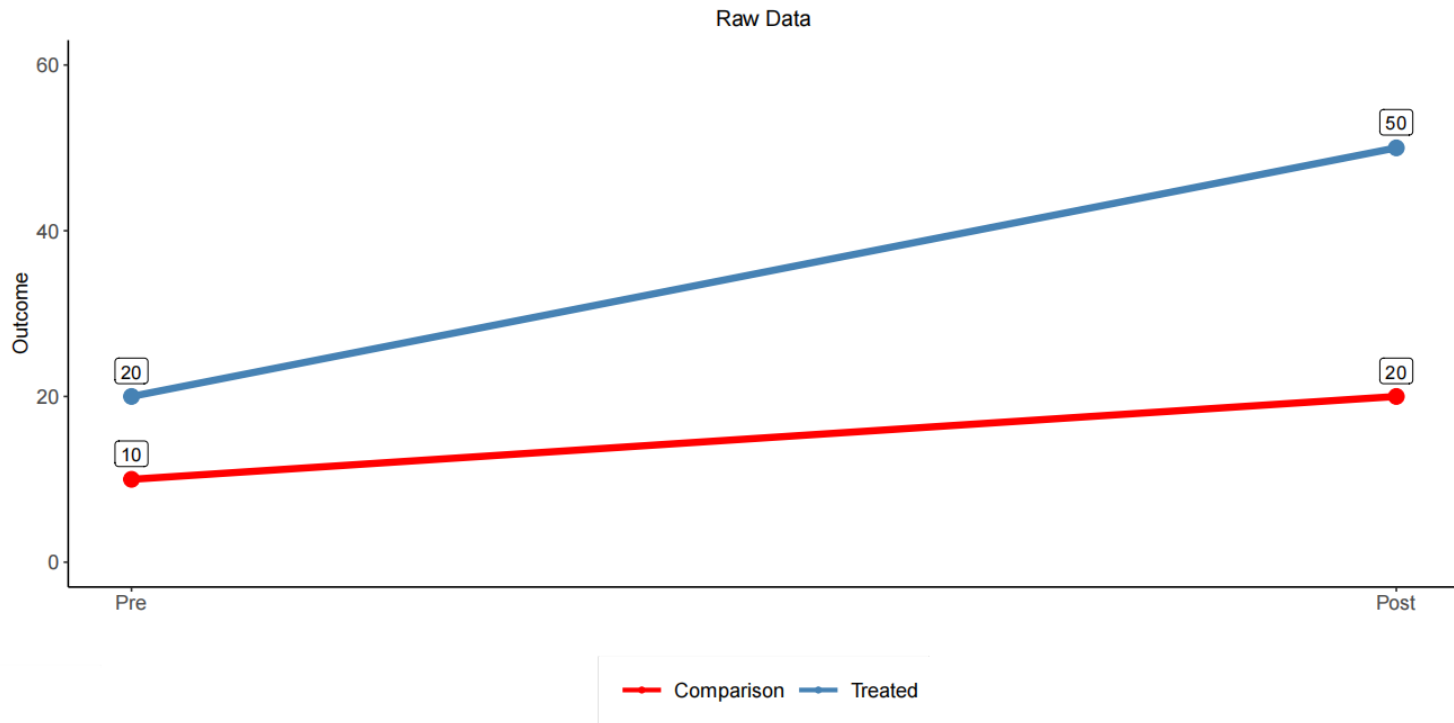
<sup>b</sup>Difference in employment between low-wage (\$4.25 per hour) and high-wage ( $\geq \$5.00$  per hour) stores; and difference in employment between midrange (\$4.26–\$4.99 per hour) and high-wage stores.

<sup>c</sup>Subset of stores with available employment data in wave 1 and wave 2.

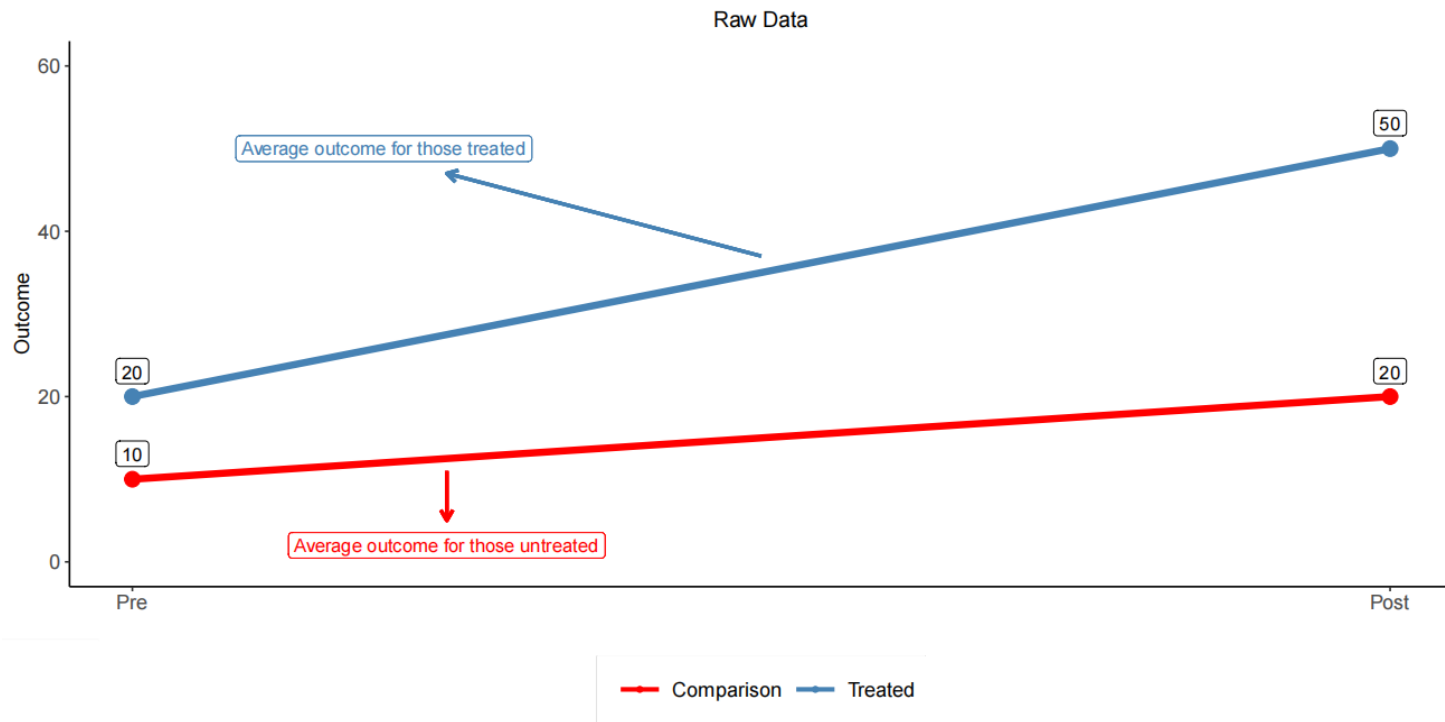
<sup>d</sup>In this row only, wave-2 employment at four temporarily closed stores is set to 0. Employment changes are based on the subset of stores with available employment data in wave 1 and wave 2.



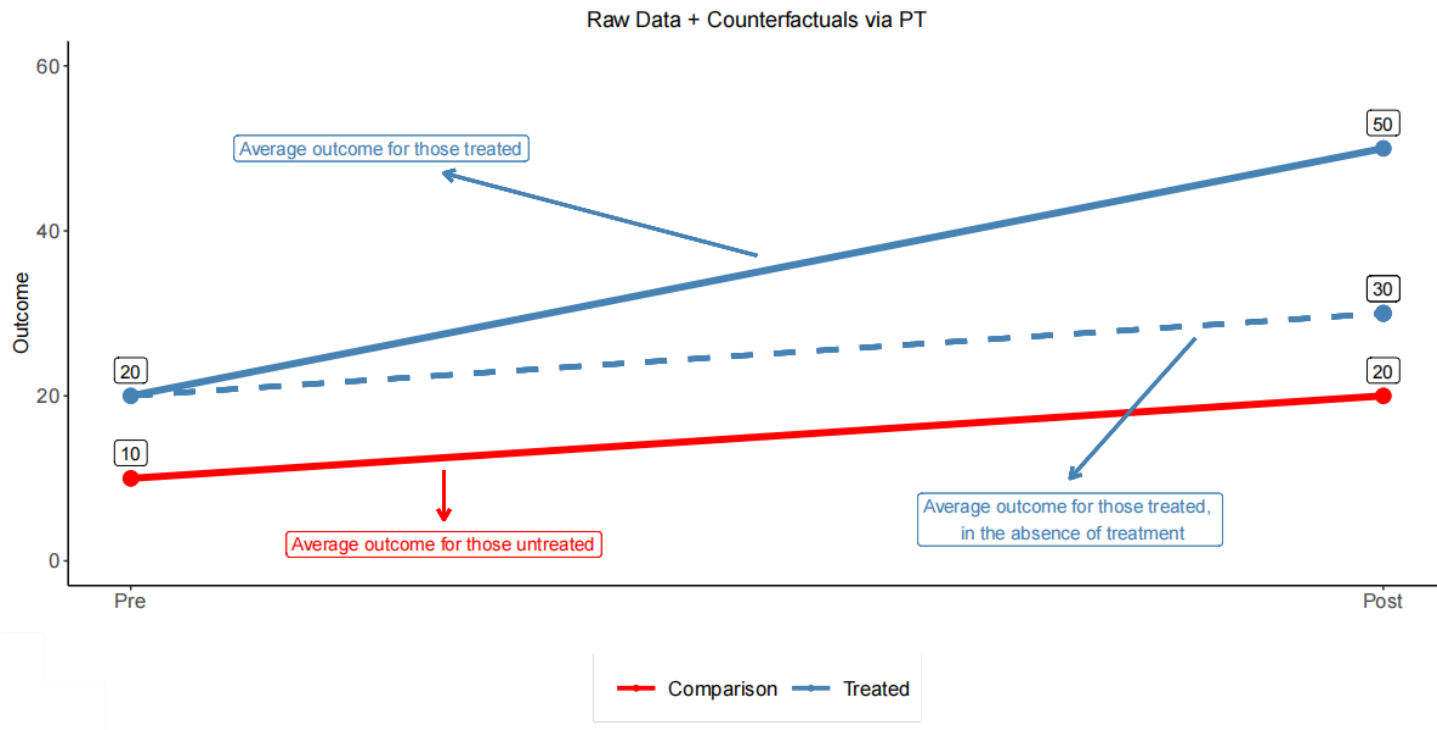
# 图示DID方法



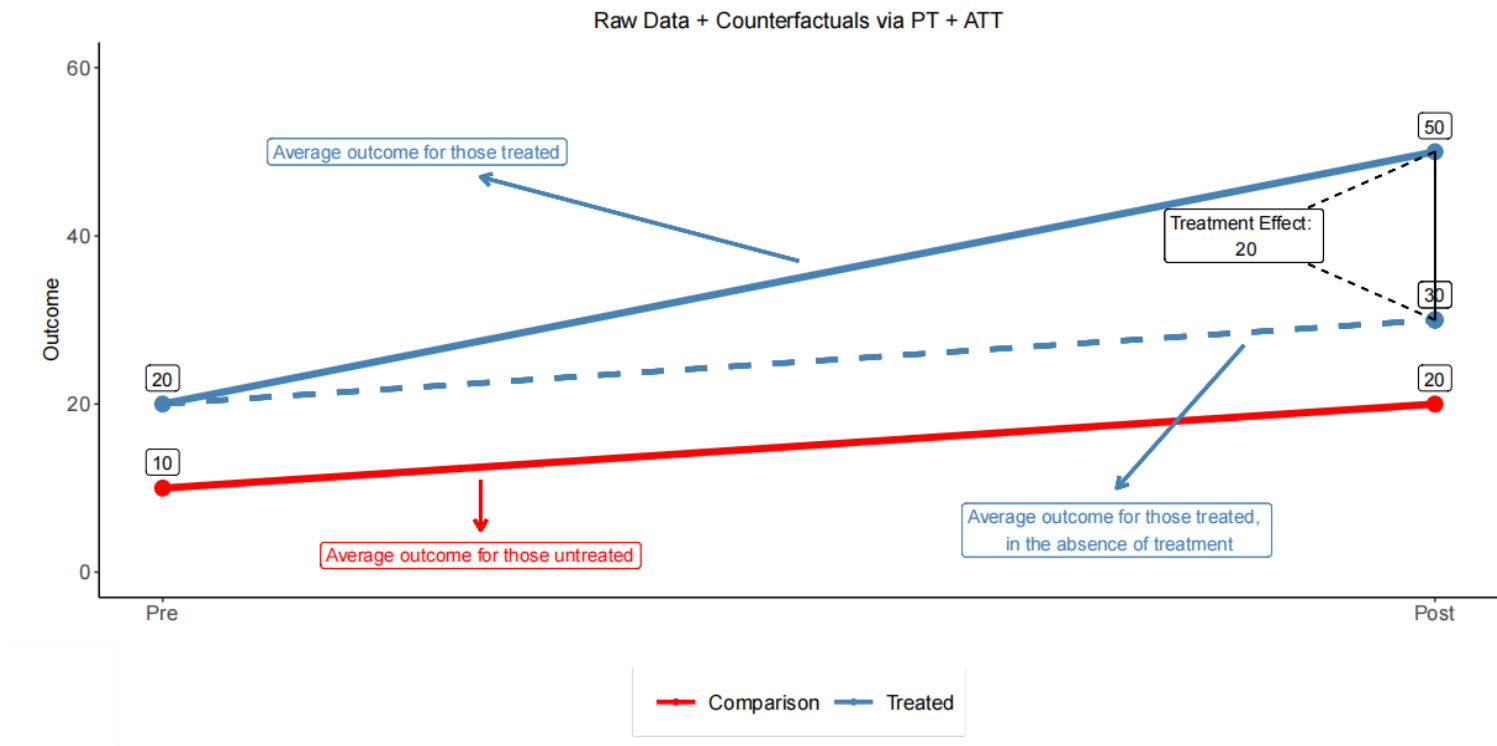
# 图示DID方法



# 图示DID方法



# 图示DID方法



# 经典的DID估计

	T = 1	T = 2
D = 0		
D = 1		

- 经典DID估计量的计算公式为：

$$\hat{\theta}_n^{\text{DiD}} = (\bar{Y}_{g=\text{treated}, t=\text{post}} - \bar{Y}_{d=\text{treated}, t=\text{pre}}) - (\bar{Y}_{d=\text{untreated}, t=\text{post}} - \bar{Y}_{g=\text{untreated}, t=\text{pre}})$$

其中， $\bar{Y}_{d=s, t=j}$  是在符合组别  $D = S$  且在时期  $j$  的样本的因变量  $Y$  的平均值，

$$\bar{Y}_{d=s, t=j} = \frac{1}{N_{d=s, t=j}} \sum_{i=1}^{N_{all}} Y_i \mathbb{1}\{D_i = s\} \mathbb{1}\{T_i = j\}$$

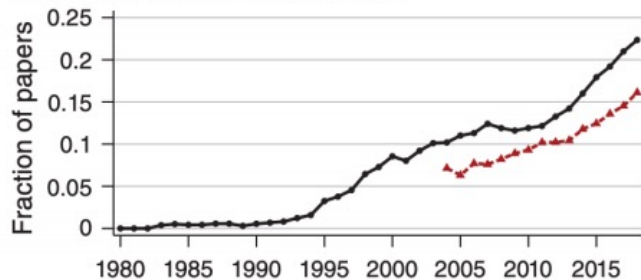
其中：

$$N_{d=s, t=j} = \sum_{i=1}^{N_{all}} \mathbb{1}\{D_i = s\} \mathbb{1}\{T_i = j\}$$

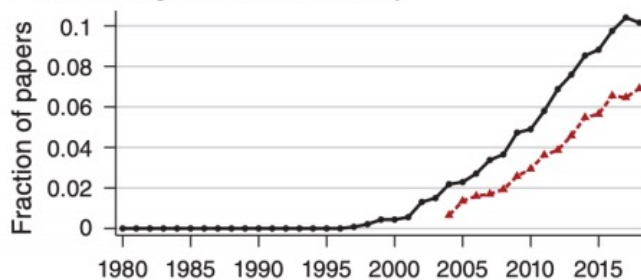
$D_i$  和  $T_i$  分别是代表组别和时期的虚拟变量， $Y_i$  表示数据中的因变量。

# DID方法近年来越来越流行

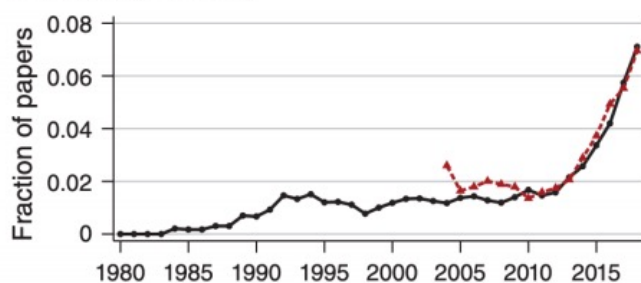
Panel A. Difference-in-differences



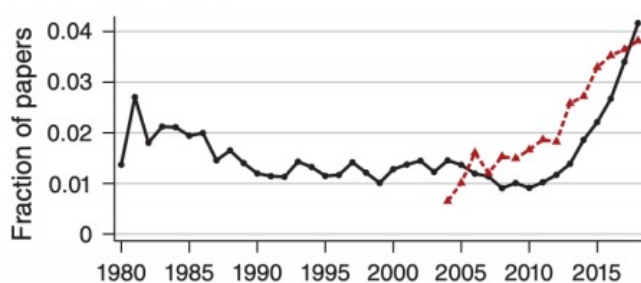
Panel B. Regression discontinuity



Panel C. Event study



Panel D. Bunching



—●— NBER working papers    -▲- Top-five journals

Currie, Kleven & Zwiars (2020)

FIGURE 4. QUASI-EXPERIMENTAL METHODS

# 为什么DID法如此流行？

- 使用观察性数据进行因果推断很难，因为总是存在遗漏变量
- DID利用了时间（处理前与处理后）和组间（处理组与未处理组）的差异，来估计我们关心的因果效应
- 优点：可以排除（1）基于不随时间变化的混淆变量造成的选择性偏误、（2）处理组和非处理组的共同趋势。
- 需要平行趋势假设，即在未处理的条件下，条件于可观测的协变量，处理组和非处理组的因变量前后变化相似
- 平行趋势假设排除了（1）随时间变化的混淆变量；（2）反馈效应（过去的Y影响今天的D）

# 非混淆性、事前-事后与 DID

## 事前—事后比较

- 比较：相同单位（个人/团体/组别）事前和事后的比较
- 方法：简单易行；只是跨时间段的平均值比较
- 缺点：未考虑结果变量的潜在趋势

当我们将研究窗口局限在非常短期的影响时，该方法更具可信度

## 处理组—非处理组比较

- 比较：与未接受处理（至少尚未接受处理）的单位进行比较
- 方法：通常使用匹配、回归或机器学习方法
- 缺点：无法排除基于不可观测变量的选择性

非混淆性（unconfoundedness）假设：不存在混淆变量，即那些既影响是否接受处理、又影响因变量的变量。



问题：  $\hat{\theta}_n^{DID}$  到底估计了什么？

核心元素： 1. 潜在结果框架； 2. 识别假设

## 第二讲大纲

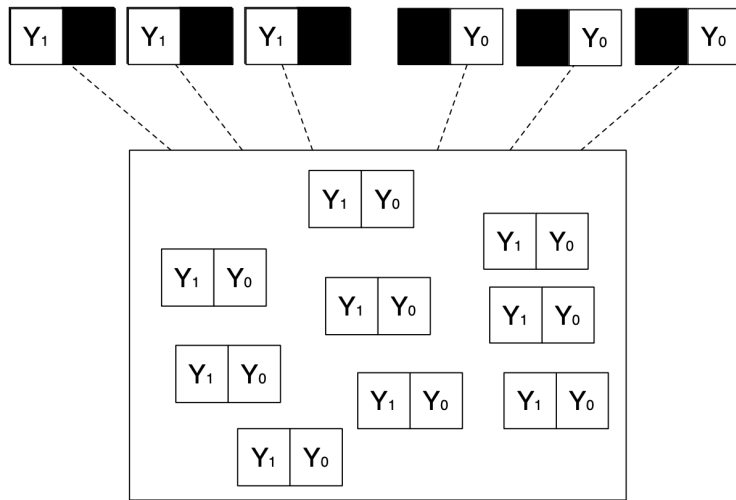
- 准备知识
- DID设计的思想、历史和演进
- $2 \times 2$  DID设计的识别
  - 潜在结果框架
  - SUTVA和无预期效应
  - 选择性偏误与平行趋势假设
  - 对DID法的威胁
- $2 \times 2$  DID设计的估计
  - 点估计
  - 标准误估计
  - 大样本性质与统计推断\*

# 潜在结果框架

---

# 因果推断的潜结果模型

- 个体  $i$
- $D_i$ : 是否接受了某种处理 (treatment)
- $Y_i$ : 观察到的结果 (outcome)
- $Y_i(0)$ 和 $Y_i(1)$ : 如果没有接受这种处理或接受这种处理的潜在结果 (potential outcome)
- 观测结果 $Y_i$ 可以表示为潜在结果和处理状态的函数:  $Y_i = Y_i(0) + [Y_i(1) - Y_i(0)] \times D_i$
- 个体处理效应  $= Y_i(1) - Y_i(0)$
- 因果推断的基本问题: 无法同时观测到两个潜在结果



# 因果推断中的潜在结果

根据鲁宾潜在结果模型：

- 令  $Y_{i,t}(1)$  代表个体  $i$ , 在时间  $t$  接收处理时的潜在结果
- 令  $Y_{i,t}(0)$  代表个体  $i$ , 在时间  $t$  不接受干预时的潜在结果
- 令  $D_i = 1$  代表该样个体属于处理组,  $D_i = 0$  代表该个体属于控制组
- 一共有  $T$  期:  $t = 1, 2, \dots, T$
- 个体  $i$  在时间  $t$  的观测到的结果为:  $Y_{i,t} = \sum_{d \in \{0,1\}} \mathbb{1}\{D_i = d\} Y_{i,t}(d)$
- 干预发生在一个固定时间点

# 在经典 $2 \times 2$ DID 情形中的使用潜在结果进行因果推断

- 下面我们来讨论典型的  $2 \times 2$  情形(Canonical  $2 \times 2$  setup)
- 一共有  $n$  个样本,  $i = 1, 2, \dots, n$
- 一共有两个时期,  $t = 1$  和  $t = 2$
- 在  $t = 1$  期, 没有样本进行了处理
- 在  $t = 2$  期, 一部分样本进行了处理 (处理组), 其余样本没有进行处理 (控制组)
- 对于处理组样本 (在  $t = 2$  期进行了处理的样本), 我们可以观测到  $Y_{i,t=2}(1)$  和  $Y_{i,t=1}(1)$
- 对于控制组样本, 我们可以观测到  $Y_{i,t=1}(0)$  和  $Y_{i,t=2}(0)$

	T = 0	T = 2
D = 0		
D = 1		

# 在经典 $2 \times 2$ DID 情形中的使用潜在结果进行因果推断

## 处置效应

- 样本  $i$  在时间  $t$  的处置效应（或因果效应）是其两个潜在结果之差：

$$Y_{i,t}(1) - Y_{i,t}(0)$$

## 观测结果

- 在时间  $t$  观测到的结果可以表示为：

$$Y_{i,t} = D_i \times Y_{i,t}(1) + (1 - D_i) \times Y_{i,t}(0)$$

## 因果推理的根本问题

- 在时间  $t$  我们无法同时观测到两个潜在结果  $Y_{i,t}(1)$  和  $Y_{i,t}(0)$

# 因果推断的根本问题：数据缺失

Unit	Data				$D_i$
	$Y_{i,t=1}(1)$	$Y_{i,t=2}(1)$	$Y_{i,t=1}(0)$	$Y_{i,t=2}(0)$	
1	?	?	✓	✓	0
2	✓	✓	?	?	1
3	?	?	✓	✓	0
4	✓	✓	?	?	1
⋮	⋮	⋮	⋮	⋮	⋮
n	✓	✓	?	?	1

✓: Observed data

?: Missing data (unobserved counterfactuals)



# 在经典 $2 \times 2$ DID情形中的使用潜在结果进行因果推断

问题：

- 由于数据缺失问题，因果推断很困难
- 在时间  $t$  我们如何得到  $Y_{i,t}(1) - Y_{i,t}(0)$  ？

若不考虑异质性：

- 假设不同样本的  $Y_{i,t}(1)$  和  $Y_{i,t}(0)$  是相同的
- 假设不同时期的  $Y_{i,t}(1)$  和  $Y_{i,t}(0)$  是相同的，并施加无预期假设 (no-anticipation assumption)

但异质性总是存在的

- 在异质性情形下，解决方案不能得到一致因果估计，仍需其他途径
- 我们需要找到更合适的假设 (之后我们将对此展开讨论)

## 2×2 DID 情形中的目标参数

- 一旦我们承认处置效应的异质性，获得特定个体的处置效就变得异常困难
- 因此，关注处理的平均因果效应
- 首先探讨 2×2 DID 情形
- 只关注处理后的效果（ATT）——为什么？

	T = 1	T = 2
D = 0		
D = 1		

## 2×2 DID 情形中的目标参数

- ATT

处理组在时期  $t=2$  时的平均处置效应为：

$$ATT = \mathbb{E} [Y_{i,t=2}(1) - Y_{i,t=2}(0) \mid D_i = 1]$$

- ATU

控制组在时期  $t=2$  时的平均处置效应为：

$$ATU = \mathbb{E} [Y_{i,t=2}(1) - Y_{i,t=2}(0) \mid D_i = 0]$$

- ATE

所有样本在时期  $t=2$  时的平均处置效应为：

$$ATE = \mathbb{E} [Y_{i,t=2}(1) - Y_{i,t=2}(0)]$$

## 2×2 DID 情形中的目标参数

这些参数回答了不同的问题：

- ATT

在时期  $t=2$  时实际接受处理的样本中，处理的平均效果如何？

- ATU

在时期  $t=2$  时没有接受处理的样本中，如果他们接收了处理，那么处理的平均效果如何？

- ATE

在时期  $t=2$  时的所有样本中，如果都接受了处理，那么处理的平均效果如何？

## 第二讲大纲

- DID设计的思想、历史和演进
- $2 \times 2$  DID设计的识别
  - 潜在结果框架
  - SUTVA和无预期效应
  - 选择性偏误与平行趋势假设
  - 对DID法的威胁
- $2 \times 2$  DID设计的估计
  - 点估计
  - 标准误估计
  - 大样本性质与统计推断\*

# 个体处理稳定性假设

---

## 个体处理稳定性假设 (SUTVA)

个体处理稳定性假设 (Stable Unit Treatment Value Assumption , SUTVA)

- 时间  $t$  的观测结果可以表示为:

$$Y_{i,t} = \sum_{d \in \{0,1\}} \mathbb{1} \{D_i = d\} Y_{i,t}(d)$$

- 在  $2 \times 2$  DID 情形中, 时间  $t$  的观测结果可以表示为:

$$Y_{i,t} = \mathbb{1} \{D_i = 1\} Y_{i,t}(1) + \mathbb{1} \{D_i = 0\} Y_{i,t}(0)$$

# 个体处理稳定性假设 (SUTVA)

个体处理稳定性假设 (Stable Unit Treatment Value Assumption , SUTVA)

- 时间  $t$  的观测结果可以表示为:

$$Y_{i,t} = \sum_{d \in \{0,1\}} \mathbb{1}\{D_i = d\} Y_{i,t}(d)$$

- 这个假设意味着, 样本  $i$  的潜在结果不因样本  $j$  是否接受处理而改变
  - 排除分析单位之间的干扰 (interference)
  - 排除溢出效应 (spillover effects)
  - 排除一般均衡效应 (general equilibrium)



# 个体处理稳定性假设 (SUTVA)

个体处理稳定性假设 (Stable Unit Treatment Value Assumption , SUTVA)

- 时间  $t$  的观测结果可以表示为:

$$Y_{i,t} = \sum_{d \in \{0,1\}} \mathbb{1} \{D_i = d\} Y_{i,t}(d)$$

- 这个假设在某些情境中可能不成立
- 我们应该选择合适的分析单位，以尽量减少样本间的干扰。

处理前是否存在“因果效应”？  
——无预期效应

# 无预期假设 (No-Anticipation)

无预期假设 (No-Anticipation) :

- 对于所有样本  $i$ ，无论其属于处理组还是控制组，处理前的所有时期都满足

$$Y_{i,t}(1) = Y_{i,t}(0)$$

- 这是久期分析 (duration analysis) 中的常见假设 (Abbring & van den Berg, 2003; Sianesi, 2004)
- 这一假设表明，在所有处理前阶段 (pre-treatment periods)，任意特定样本的处置效应为零。
- 允许处理后阶段 (post-treatment periods) 处置效应具有异质性。
- 这一假设在许多情境中是合理的，尤其是在没有提前公布处置方案的情况下。
- 但该假设并非总是成立的 (Malani & Reif, 2015)。

# 无预期假设 (No-Anticipation)

无预期假设 (No-Anticipation) :

- 对于所有样本  $i$ ，无论其属于处理组还是控制组，处理前的所有时期都满足

$$Y_{i,t}(1) = Y_{i,t}(0)$$

- 在该假设前提下，我们可以“简化”符号。
- 可以将所有的未处理（或尚未接受处理）的潜在结果用  $Y_{i,t}(0)$  表示

## 2×2 DID, 因果推断的基本问题

	t = 1	t = 2
D = 0	Y(0), ?	Y(0), ?
D = 1	?, Y(1)	?, Y(1)

Unit	Data				$D_i$
	$Y_{i,t=1}(1)$	$Y_{i,t=2}(1)$	$Y_{i,t=1}(0)$	$Y_{i,t=2}(0)$	
1	?	?	✓	✓	0
2	✓	✓	?	?	1
3	?	?	✓	✓	0
4	✓	✓	?	?	1
⋮	⋮	⋮	⋮	⋮	⋮
n	✓	✓	?	?	1

✓: Observed data

?: Missing data (unobserved counterfactuals)

# 数据缺失 + SUTVA + 无预期假设

	t = 1	t = 2
D = 0	Y(0) = Y(1)	Y(0), ?
D = 1	Y(0) = Y(1)	?, Y(1)

Unit	Data				$D_i$
	$Y_{i,t=1}(1)$	$Y_{i,t=2}(1)$	$Y_{i,t=1}(0)$	$Y_{i,t=2}(0)$	
1	✓	?	✓	✓	0
1	✓	✓	✓	?	1
3	✓	?	✓	✓	0
4	✓	✓	✓	?	1
⋮	⋮	⋮	⋮	⋮	⋮
n	✓	✓	✓	?	1

✓: Observed data

?: Missing data (unobserved counterfactuals)

# 数据缺失 + SUTVA + 无预期假设

	t = 1	t = 2
D = 0	Y(0)	Y(0), ?
D = 1	Y(0)	?, Y(1)

Unit	Data			$D_i$
	$Y_{i,t=1}(0)$	$Y_{i,t=2}(1)$	$Y_{i,t=2}(0)$	
1	✓	?	✓	0
0	✓	✓	?	0
3	✓	?	✓	0
4	✓	✓	?	0
⋮	⋮	⋮	⋮	⋮
n	✓	✓	?	0

✓: Observed data

?: Missing data (unobserved counterfactuals)

# 第一讲大纲

- 准备知识
- DID设计的思想、历史和演进
- $2 \times 2$  DID设计的识别
  - 潜在结果框架
  - SUTVA和无预期效应
  - 选择性偏误与平行趋势假设
  - 对DID法的威胁
- $2 \times 2$  DID设计的估计
  - 点估计
  - 标准误估计
  - 大样本性质与统计推断\*



# 选择性偏差

---

# 选择性偏误

问题：直接的对比处理组和控制组在  $t = 2$  时的结果，通常不能得到正确的因果效应

$$\begin{aligned}\mathbb{E}[Y_{i,t=2} \mid D_i = 1] - \mathbb{E}[Y_{i,t=2} \mid D_i = 0] &= \mathbb{E}[Y_{i,t=2}(1) \mid D_i = 1] - \mathbb{E}[Y_{i,t=2}(0) \mid D_i = 0] \\ &= \mathbb{E}[Y_{i,t=2}(1) - Y_{i,t=2}(0) \mid D_i = 1] \\ &\quad + (\mathbb{E}[Y_{i,t=2}(0) \mid D_i = 1] - \mathbb{E}[Y_{i,t=2}(0) \mid D_i = 0]) \\ &= ATT + \text{Selection Bias}\end{aligned}$$

- 在大多数情境中，选择性偏误项不可能为零
- 是否接受处置往往与潜在结果相关

# 选择偏差

例子：弱势群体就业培训计划：

- 参与者是从就业困难的子群体中自行选择的
- 如果没有该计划，处理后时期参与者的收入将低于非参与者

$$\mathbb{E}[Y_{i,t=2}(0) \mid D_i = 1] - \mathbb{E}[Y_{i,t=2}(0) \mid D_i = 0] < 0$$

利用时间维度来克服“选择性偏误”

# 平行趋势假设

---

## 平行趋势假设 – $2 \times 2$ 情形

由于通过简单比较时期  $t = 2$  时处理组和控制组的结果变量平均值无法得到我们的目标参数（ATT），因此我们需要采用其他方法。

- 平行趋势假设（Parallel Trends Assumption）

$$\mathbb{E}[Y_{i,t=2}(0) \mid D_i = 1] - \mathbb{E}[Y_{i,t=1}(0) \mid D_i = 1] = \mathbb{E}[Y_{i,t=2}(0) \mid D_i = 0] - \mathbb{E}[Y_{i,t=1}(0) \mid D_i = 0]$$

平行趋势假设表明，在没有处理的情况下，处理组和控制组结果变量随时间的平均发展趋势相同。

# 平行趋势假设与ATT的估计

首先我们将分析的目标参数定为  $t = 2$  时的ATT

	t = 1	t = 2
D = 0	Y(0)	Y(0), ?
D = 1	Y(0)	<b>Y(0), Y(1)</b>

- 根据 SUTVA 假设 和 ATT的定义，我们可以得到：

$$ATT \equiv \mathbb{E}[Y_{i,t=2}(1) | D_i = 1] - \mathbb{E}[Y_{i,t=2}(0) | D_i = 1]$$

$$= \underbrace{\mathbb{E}[Y_{i,t=2} | D_i = 1]}_{\text{by SUTVA}} - \mathbb{E}[Y_{i,t=2}(0) | D_i = 1]$$

- 绿色部分可从数据中估计（根据SUTVA假设）。
- 红色部分仍然取决于潜在结果，因此我们的目标是找到“估计”它的方法。
- 这就需要用到平行趋势假设！

# 平行趋势假设和ATT

- 首先，回顾平行趋势假设：

$$\mathbb{E}[Y_{i,t=2}(0) \mid D_i = 1] - \mathbb{E}[Y_{i,t=1}(0) \mid D_i = 1] = \mathbb{E}[Y_{i,t=2}(0) \mid D_i = 0] - \mathbb{E}[Y_{i,t=1}(0) \mid D_i = 0]$$

- 通过简单的移项，我们可以将其写为：

$$\mathbb{E}[Y_{i,t=2}(0) \mid D_i = 1] = \mathbb{E}[Y_{i,t=1}(0) \mid D_i = 1] + (\mathbb{E}[Y_{i,t=2}(0) \mid D_i = 0] - \mathbb{E}[Y_{i,t=1}(0) \mid D_i = 0])$$

- 现在，利用无预期假设和SUTVA：

$$\mathbb{E}[Y_{i,t=2}(0) \mid D_i = 1] = \underbrace{\mathbb{E}[Y_{i,t=1}(2) \mid D_i = 1]}_{\text{by No-Anticipation}} + (\mathbb{E}[Y_{i,t=2}(0) \mid D_i = 0] - \mathbb{E}[Y_{i,t=1}(0) \mid D_i = 0])$$

$$\mathbb{E}[Y_{i,t=2}(0) \mid D_i = 1] = \underbrace{\mathbb{E}[Y_{i,t=1} \mid D_i = 1] + (\mathbb{E}[Y_{i,t=2} \mid D_i = 0] - \mathbb{E}[Y_{i,t=1} \mid D_i = 0])}_{\text{by SUTVA}}$$



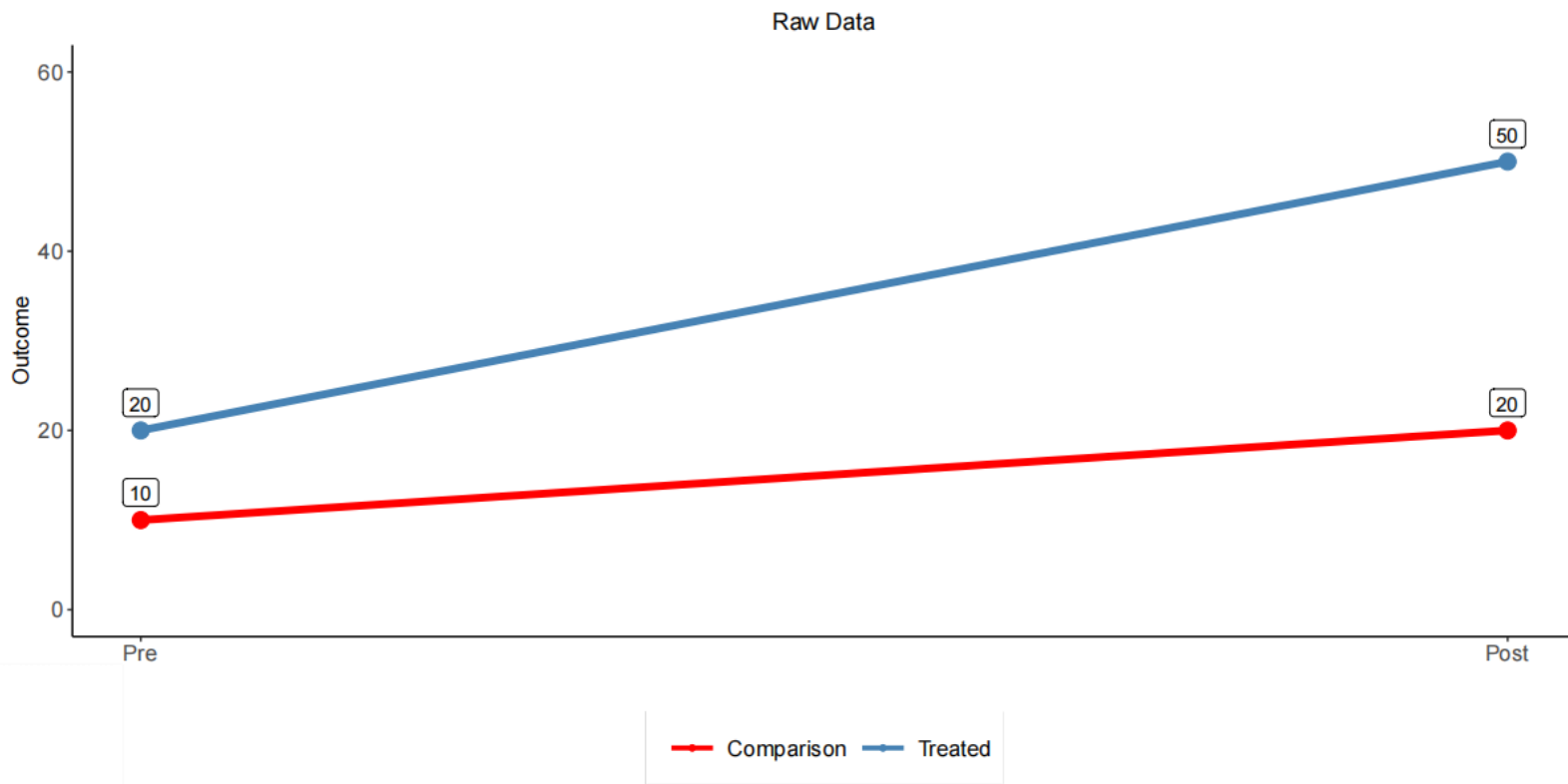
## 平行趋势假设和ATT

- 将这些结果结合在一起，在SUTVA+无预期+平行趋势假设下，可以得出：

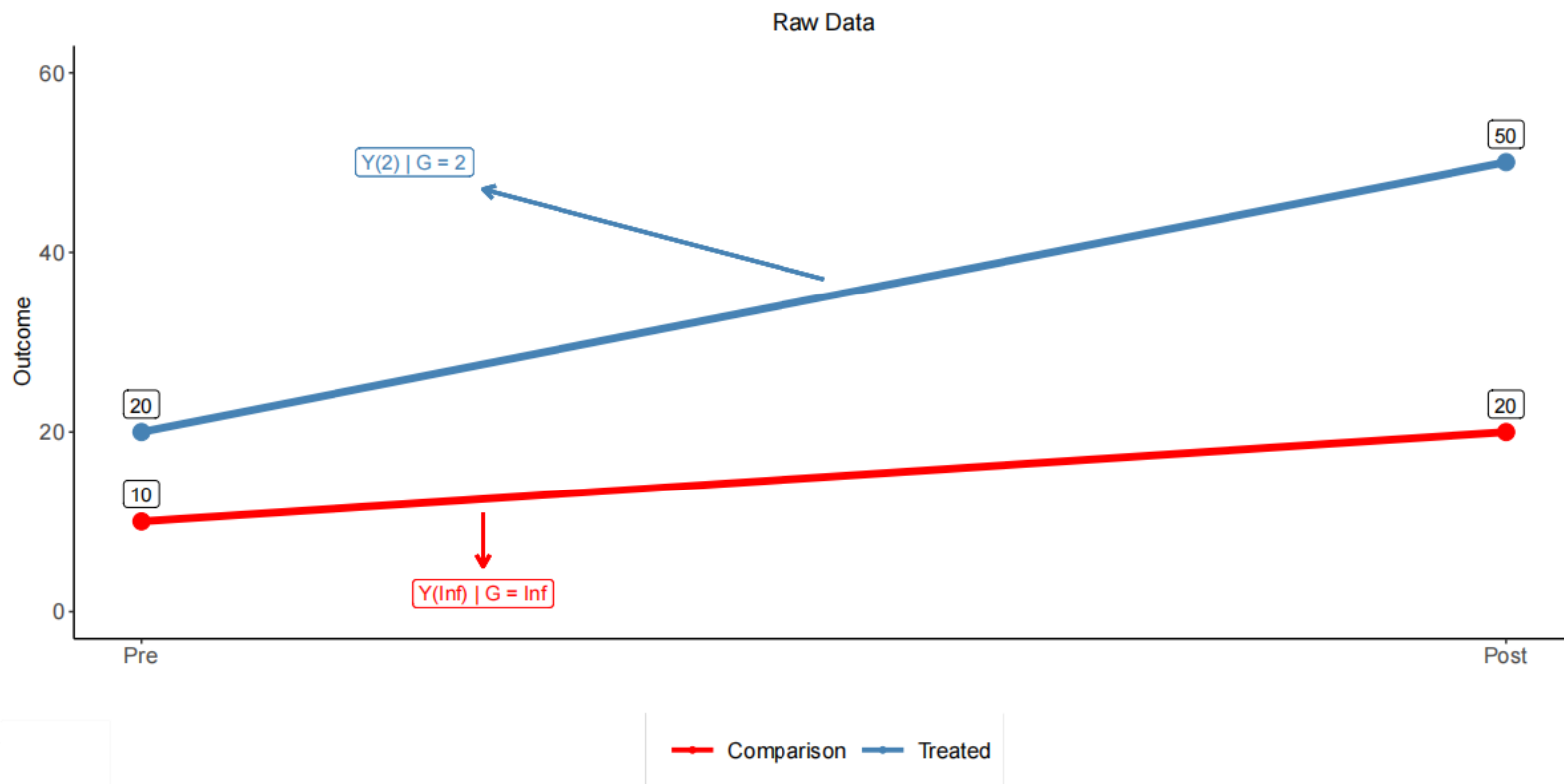
$$\begin{aligned} \text{ATT} &= \mathbb{E}[Y_{i,t=2} \mid D_i = 1] - (\mathbb{E}[Y_{i,t=1} \mid D_i = 1] + (\mathbb{E}[Y_{i,t=2} \mid D_i = 0] - \mathbb{E}[Y_{i,t=1} \mid D_i = 0])) \\ &= (\mathbb{E}[Y_{i,t=2} \mid D_i = 1] - \mathbb{E}[Y_{i,t=1} \mid D_i = 1]) - (\mathbb{E}[Y_{i,t=2} \mid D_i = 0] - \mathbb{E}[Y_{i,t=1} \mid D_i = 0]) \end{aligned}$$

- DID方法就这样“诞生”了！

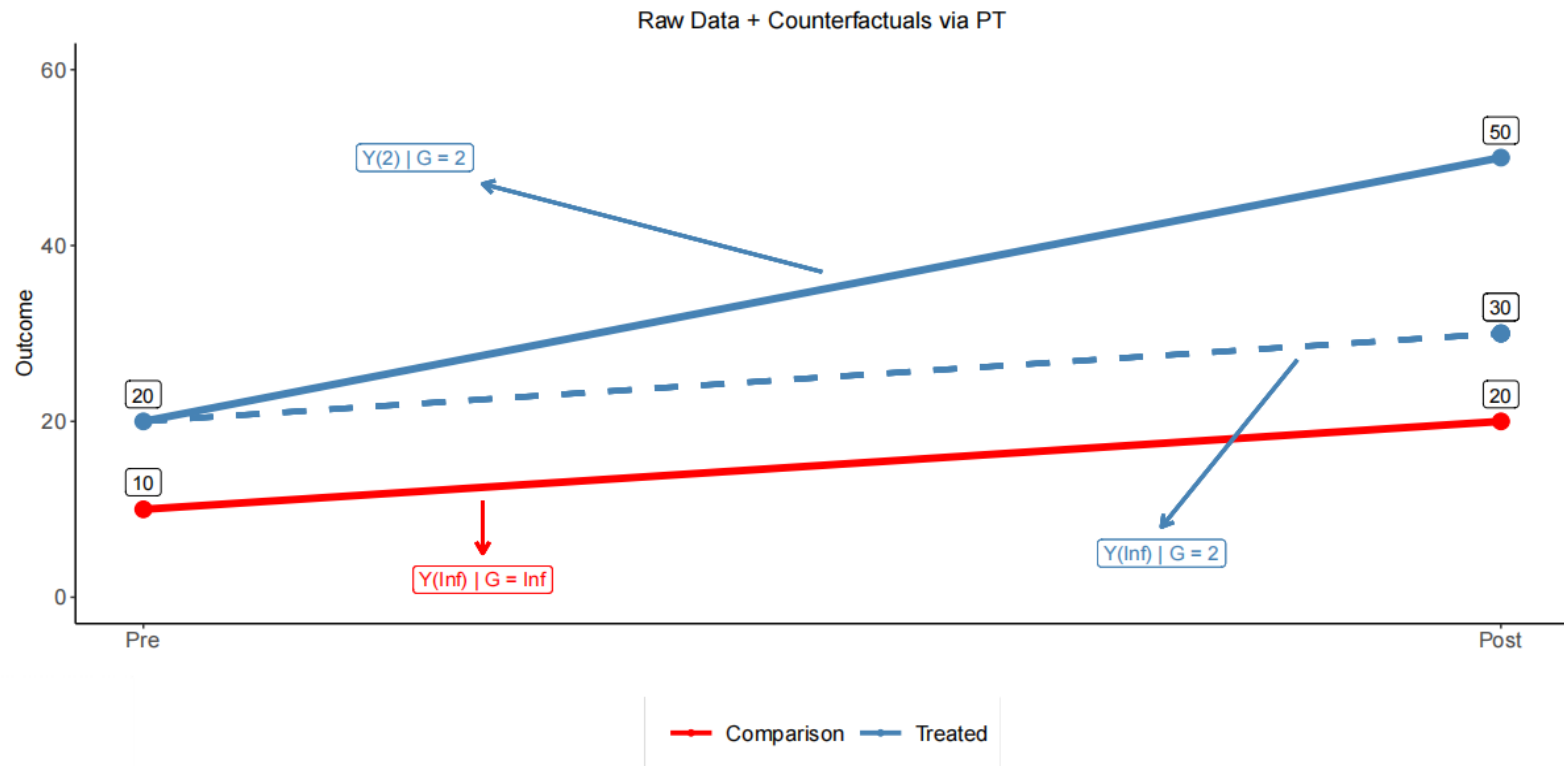
# 图示平行趋势假设



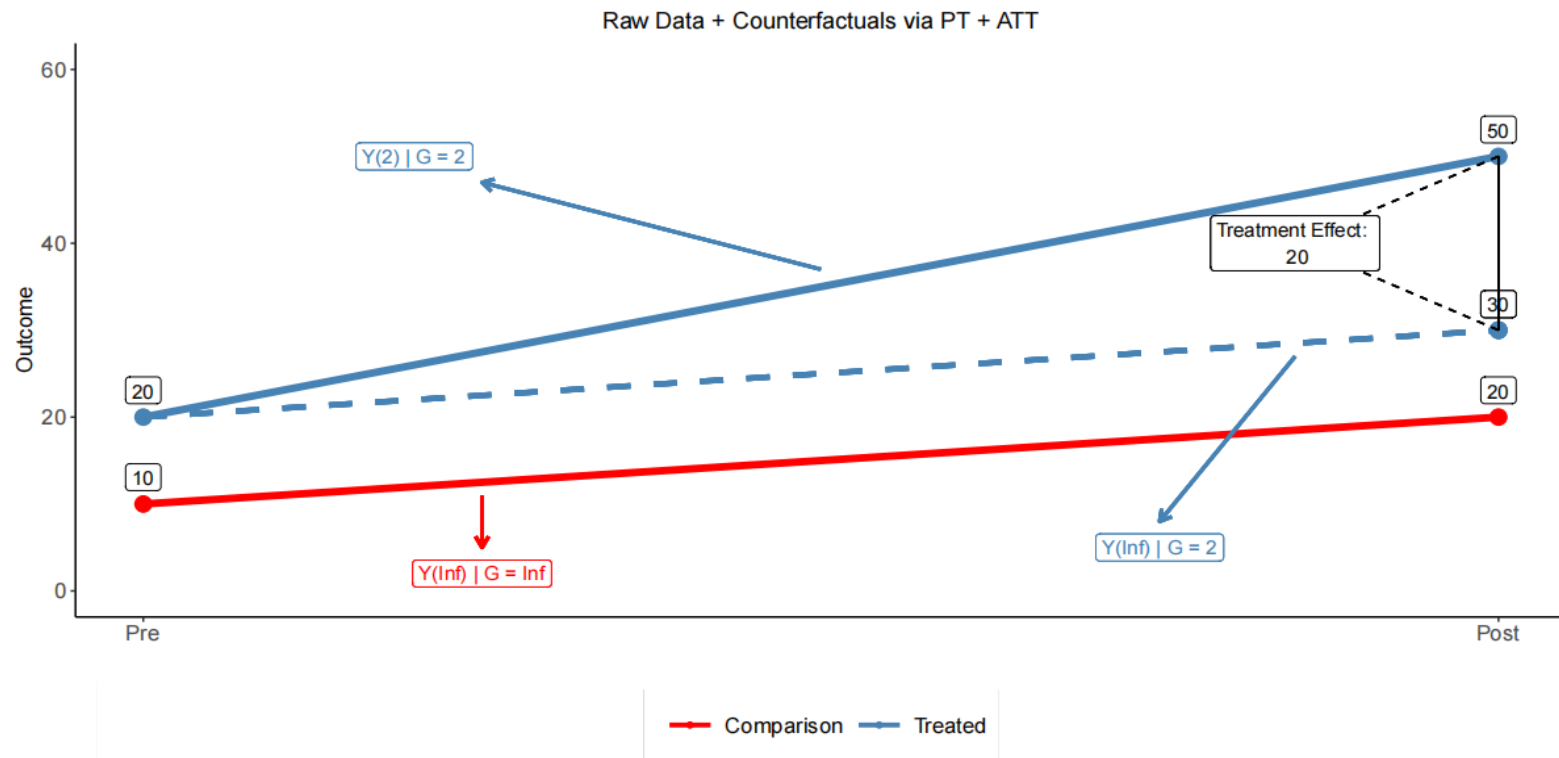
# 图示平行趋势假设



# 图示平行趋势假设



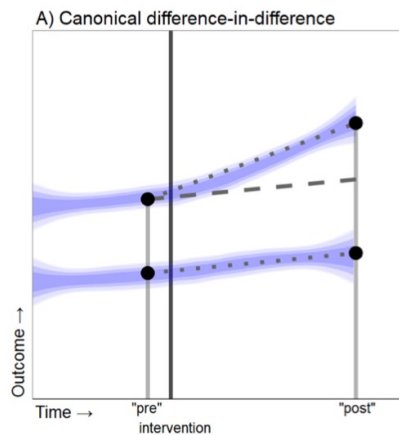
# 图示平行趋势假设



## 第二讲大纲

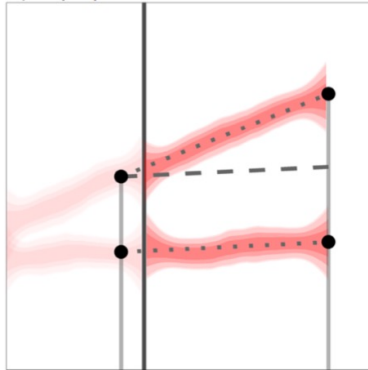
- DID设计的思想、历史和演进
- $2 \times 2$  DID设计的识别
  - 潜在结果框架
  - SUTVA和无预期效应
  - 选择性偏误与平行趋势假设
  - 对DID法的威胁
- $2 \times 2$  DID设计的估计
  - 点估计
  - 标准误估计
  - 大样本性质与统计推断\*

# 对DID法的威胁

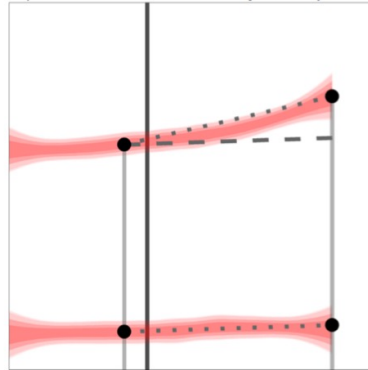


(Haber et al 2021)

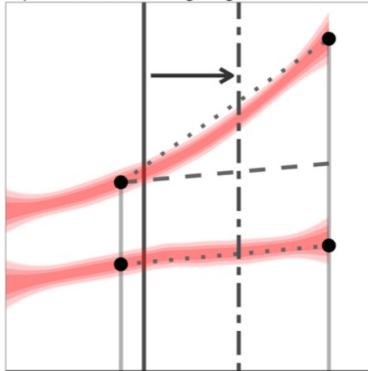
B) No pre/parallel-trend established



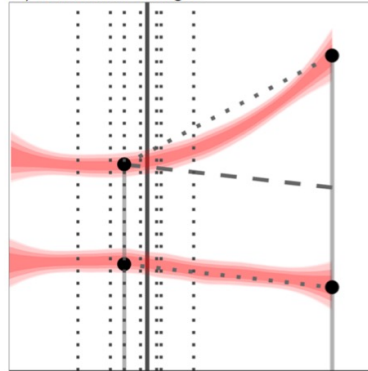
C) Scale differences / linearity assumptions



D) Misattributed timing / lags



E) Concurrent changes



## 第二讲大纲

- DID设计的思想、历史和演进
- $2 \times 2$  DID设计的识别
  - 潜在结果框架
  - SUTVA和无预期效应
  - 选择性偏误与平行趋势假设
  - 对DID法的威胁
- $2 \times 2$  DID设计的估计
  - 点估计
  - 标准误估计
  - 大样本性质与统计推断\*



## 2×2 DID法的估计

- 目前为止，我们只证实了在满足SUTVA + 无预期效应 + 平行趋势假设下，处理组的平均处理效应（ATT）可以被识别
- 但我们的估计涉及到总体期望，而且在实践中，我们并不真正知道真正的数据生成过程，故而无法计算总体期望
- 我们可以用“类比原则”来进行估计：用样本均值（sample analogues）来代替总体期望（population expectations）

## 2×2 DID估计量：“简单粗暴法”(Brute-force)

- 采用类比（或代入）原则，我们可以得到经典双重差分法估计的处理组平均处理效应 ATT

$$\hat{\theta}_n^{\text{DiD}} = (\bar{Y}_{d=1,t=2} - \bar{Y}_{d=1,t=1}) - (\bar{Y}_{d=0,t=2} - \bar{Y}_{d=0,t=1})$$

- 其中  $\bar{Y}_{d=s,t=j}$  是  $s$  组内单位在  $j$  时间段内结果变量  $Y$  的样本均值

$$\bar{Y}_{d=s,t=j} = \frac{1}{N_{d=s,t=j}} \sum_{i=1}^{N \cdot T} Y_i 1\{D_i = s\} 1\{T_i = j\},$$
$$N_{d=s,t=j} = \sum_{i=1}^{N \cdot T} 1\{D_i = s\} 1\{T_i = j\},$$

- $D_i$  和  $T_i$  分别是组别和时间的虚拟变量， $Y_i$  是堆叠（pooled）数据的结果变量

## 2 × 2 DID估计量：双向固定效应 (TWFE)

- 我通常将  $\hat{\theta}_n^{\text{DID}}$  称作“简单粗暴法”(Brute-force) 或 手动双重差分估计量 (DID by hand)
- 在明确了出发点是估计总体的参数之后，我们要做的就是遵循代入 (plug-in) 原则
- 然而，在实践中，许多研究者采用双向固定效应模型来估计双重差分法中的平均处理效应 ATT:

$$Y_{i,t} = \alpha_0 + \gamma_0 D_i + \lambda_0 1\{T_i = 2\} + \beta_0^{\text{twfe}} (D_i \cdot 1\{T_i = 2\}) + \varepsilon_{i,t}$$

- 误差项几乎必然收敛于0:  $\mathbb{E}[\varepsilon_{i,t} \mid D_i, T_i] = 0$
- 我们可以证明  $\beta_0^{\text{twfe}}$  与经典2 × 2 双重差分法估计结果相等

## 2×2 DID估计量：双向固定效应 (TWFE)

- 双向固定效应模型的设定如下：

$$Y_{i,t} = \alpha_0 + \gamma_0 D_i + \lambda_0 1\{T_i = 2\} + \beta_0^{twfe} (D_i \cdot 1\{T_i = 2\}) + \varepsilon_{i,t}$$

误差项几乎必然收敛于0:  $\mathbb{E}[\varepsilon_{i,t} \mid D_i, T_i] = 0$

- 那么，可以推导出：

$$\mathbb{E}[Y_{i,t} \mid D_i = 0, T_i = 1] = \alpha_0$$

$$\mathbb{E}[Y_{i,t} \mid D_i = 0, T_i = 2] = \alpha_0 + \lambda_0$$

$$\mathbb{E}[Y_{i,t} \mid D_i = 1, T_i = 1] = \alpha_0 + \gamma_0$$

$$\mathbb{E}[Y_{i,t} \mid D_i = 1, T_i = 2] = \alpha_0 + \gamma_0 + \lambda_0 + \beta_0^{twfe}$$

## 2 × 2 DID估计量：双向固定效应 (TWFE)

- 一组矩约束：

$$\mathbb{E}[Y_{i,t} \mid D_i = 0, T_i = 1] = \alpha_0$$

$$\mathbb{E}[Y_{i,t} \mid D_i = 0, T_i = 2] = \alpha_0 + \lambda_0$$

$$\mathbb{E}[Y_{i,t} \mid D_i = 1, T_i = 1] = \alpha_0 + \gamma_0$$

$$\mathbb{E}[Y_{i,t} \mid D_i = 1, T_i = 2] = \alpha_0 + \gamma_0 + \lambda_0 + \beta_0^{\text{twfe}}$$

- 这意味着：

$$\mathbb{E}[Y_{i,t} \mid D_i = 1, T_i = 2] - \mathbb{E}[Y_{i,t} \mid D_i = 1, T_i = 1] = \lambda_0 + \beta_0^{\text{twfe}}$$

$$\mathbb{E}[Y_{i,t} \mid D_i = 0, T_i = 2] - \mathbb{E}[Y_{i,t} \mid D_i = 0, T_i = 1] = \lambda_0$$

- 于是，我们可以证明  $\beta_0^{\text{twfe}}$  与经典 2 × 2 双重差分法估计结果相等

## 2×2 DID估计量：双向固定效应 (TWFE)

- 然后，我们可以采用最小二乘法来估计 $\beta_0^{\text{twfe}}$  (或平均处理效应ATT)

$$\left(\hat{\alpha}, \hat{\gamma}, \hat{\lambda}, \hat{\beta}^{\text{twfe}}\right) = \underset{\alpha, \gamma, \lambda, \beta^{\text{twfe}}}{\operatorname{argmin}} \sum_{j=1}^{N_{\text{all}}} \left(Y_j - \alpha - \gamma D_i - \lambda 1\{T_j = 2\} - \beta^{\text{twfe}} (D_i \cdot 1\{T_j = 2\})\right)^2$$

- 强调：我们用这个回归方程是为了得到平均处理效应 ATT
- 回归只是一种估计工具——并不能修正目标参数！
- 我们之所以喜欢用回归是因为我们对回归的理解比较充分！
- 我们可以利用它来进行渐近式的合理推断

## 第二讲大纲

- DID设计的思想、历史和演进
- $2 \times 2$  DID设计的识别
  - 潜在结果框架
  - SUTVA和无预期效应
  - 选择性偏误与平行趋势假设
  - 对DID法的威胁
- $2 \times 2$  DID设计的估计
  - 点估计
  - 标准误估计
  - 大样本性质与统计推断\*

# 聚类的重要性

- 我们至少应该聚类到横截面的单位层面，这样聚类是为了允许：
  - ✓ 同一个体，不同时期的结果变量之间任意的自相关性
- 标准的统计推断程序（无附加严格的分布假设）只有在聚类数较多时才可信
- 在没有正态分布假设的情况下，我们需要用中心极限定理来证明推断的合理性。
- 中心极限定理的可靠性取决于有效的样本量，即聚类的数量，要大



## 聚类的重要性：双向固定效应模型 和 手动双重差分法

我们已经讨论了：

- 手动计算得到的双重差分法估计量  $\theta_n^{DID}$ ，在数值上等于  $2 \times 2$  双向固定效应估计量  $\hat{\beta}^{twfe}$
- 如果我们要推导出两个估计量在相同的假设条件下的**大样本性质**，它们应该是完全一样。

但是如果我们并不聚类的话，上述推论并不成立：

- 手动法得到的双重差分估计量明确指出有效样本量为：横截面单位个数或聚类的数量
- 双向固定效应模型的有效样本则是堆叠而成  $2 \times$  横截面单位个数 (平衡面板)

当单位的数目较少时，如  $< 40$ ，需要使用带聚类的自助抽样法 (clustered bootstrap)，来获得比较准确的不确定性估计。

# 不确定性估计：抽样的视角

- 在  $2 \times 2$  DID 设定下，主要有两种抽样设计：

## 假设前提（面板数据的抽样设计）

- 数据  $\{Y_{i,t=1}, Y_{i,t=2}, D_i\}_{i=i}^n$  是我们所关心的总体的一个随机样本
- 对于同一个体，我们可以观测到其  $t=1$  期和  $t=2$  期的数据
- 所以我们要将标准误聚类在个体层面！

## 第二讲大纲

- DID设计的思想、历史和演进
- $2 \times 2$  DID设计的识别
  - 潜在结果框架
  - SUTVA和无预期效应
  - 选择性偏误与平行趋势假设
  - 对DID法的威胁
- $2 \times 2$  DID设计的估计
  - 点估计
  - 标准误估计
  - 大样本性质与统计推断\*

# 面板数据下的2 × 2 双重差分法

- 回忆双重差分估计量:

$$\hat{\theta}_n^{\text{DiD}} = (\bar{Y}_{g=2,t=2} - \bar{Y}_{g=2,t=1}) - (\bar{Y}_{g=0,t=2} - \bar{Y}_{g=0,t=1})$$

- 对于面板数据, 可以进一步简化为:

$$\hat{\theta}_n^{\text{DiD}} = \overline{\Delta Y}_{g=2} - \overline{\Delta Y}_{g=0}$$

- 其中  $\overline{\Delta Y}_{g=d}$  是  $\Delta Y_i \equiv Y_{i,t=2} - Y_{i,t=1}$  组 d 中的个体的抽样均值

$$\overline{\Delta Y}_{g=d} = \frac{\sum_{i:D_i=d} \Delta Y_i}{n_{G=d}} = \frac{n^{-1} \sum_{i=1}^n \Delta Y_i 1\{D_i = d\}}{n^{-1} \sum_{i=1}^n 1\{D_i = d\}} = \frac{\mathbb{E}_n[\Delta Y \cdot 1\{G = d\}]}{\mathbb{E}_n[1\{G = d\}]}$$

- 其中  $n_{G=d} = \sum_{i=1}^n 1\{G = d\}$  是组 G=d 的抽样大小

- 因此对于一个通用变量A来说:

$$\mathbb{E}_n[A] \equiv \frac{\sum_{i=1}^n A_i}{n}$$

# 面板数据下的 $2 \times 2$ 双重差分法

- 然后我们可以得到：

$$\hat{\theta}_n^{\text{DiD}} = \overline{\Delta Y}_{g=2} - \overline{\Delta Y}_{g=0} = \frac{\mathbb{E}_n[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}_n[1\{D = 1\}]} - \frac{\mathbb{E}_n[\Delta Y \cdot 1\{D = 0\}]}{\mathbb{E}_n[1\{G = 0\}]}$$

我们想知道这个估计量是否可靠 (reliable):

- 在我们的假设下，随着单位个数的增加，它是否在概率上收敛于真正的 ATT?
- 我们怎样才能在不违反分布假设的情况下对ATT进行可靠的推断?
- 我们将利用大样本的近似结果
- 你在统计课上学到/教的知识派上用场了！
- 我们将用到大数法则 (LLN) + 连续映射定理 (CMT) + 中心极限定理 (CLT)

# 面板数据下的 $2 \times 2$ 双重差分法估计量的一致性

- 因为该估计量的一致性直接由大数定律和连续映射定理推导而来：

$$\hat{\theta}_n^{\text{DiD}} = \frac{\mathbb{E}_n[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}_n[1\{D = 1\}]} - \frac{\mathbb{E}_n[\Delta Y \cdot 1\{D = 0\}]}{\mathbb{E}_n[1\{D = 0\}]}$$

- 大数定律 (law of large numbers)**：在独立同分布 (iid)+有界矩 (bounded moments) 的情况下（我们的隐含假设），样本均值依概率收敛于总体均值
- 连续映射定理 (Continuous mapping theorem)**：连续函数保持极限 (continuous functionals preserve limits)
- 因此当  $n \rightarrow \infty$  时

$$\hat{\theta}_n^{\text{DiD}} \xrightarrow{p} \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}[1\{D = 1\}]} - \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 0\}]}{\mathbb{E}[1\{D = 0\}]} = \mathbb{E}[\Delta Y \mid D = 1] - \mathbb{E}[\Delta Y \mid D = 0] \equiv \theta^{\text{DiD}}$$

$$\theta^{\text{DiD}} = ATT$$

# 面板数据下的 $2 \times 2$ 双重差分法估计量的渐近正态性

- 现在，我们要推导出以下的渐近分布：

$$\begin{aligned} \sqrt{n} \left( \hat{\theta}_n^{\text{DiD}} - \theta^{\text{DiD}} \right) = & \sqrt{n} \left( \frac{\mathbb{E}_n[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}_n[1\{D = 1\}]} - \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}[1\{D = 1\}]} \right) \\ & - \sqrt{n} \left( \frac{\mathbb{E}_n[\Delta Y \cdot 1\{D = 0\}]}{\mathbb{E}_n[1\{G = 0\}]} - \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 0\}]}{\mathbb{E}[1\{G = 0\}]} \right) \end{aligned}$$

- 为此，我们可以使用中心极限定理和 Delta 方法（独立同分布+有限渐近方差+分母不等于零）
- 我们可以稍微变通一下，因为我们想得到影响函数（influence function）
- 将  $\sqrt{n} \left( \hat{\theta}_n^{\text{DiD}} - \theta^{\text{DiD}} \right)$  表示成：
  - › 一组独立同分布项（iid terms）的均值+可忽略项（negligible terms）

# 面板数据下的 $2 \times 2$ 双重差分法估计量的渐近正态性

- 首先来分析：

$$\sqrt{n} \left( \frac{\mathbb{E}_n[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}_n[1\{D = 1\}]} - \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}[1\{D = 1\}]} \right)$$

- 变形后可得：

$$\begin{aligned} & \frac{1}{\mathbb{E}[1\{D = 1\}]} \sqrt{n} (\mathbb{E}_n[\Delta Y \cdot 1\{D = 1\}] - \mathbb{E}[\Delta Y \cdot 1\{D = 1\}]) \\ & - \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}[1\{D = 1\}]^2} \sqrt{n} (\mathbb{E}_n[1\{D = 1\}] - \mathbb{E}[1\{D = 1\}]) \\ & + \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 1\}] \cdot (\mathbb{E}_n[1\{D = 1\}] - \mathbb{E}[1\{D = 1\}])}{\mathbb{E}[1\{D = 1\}]^2 \cdot \mathbb{E}_n[1\{D = 1\}]} \sqrt{n} (\mathbb{E}_n[1\{D = 1\}] - \mathbb{E}[1\{D = 1\}]) \\ & - \frac{(\mathbb{E}_n[1\{D = 1\}] - \mathbb{E}[1\{D = 1\}])}{\mathbb{E}[1\{D = 1\}] \cdot \mathbb{E}_n[1\{D = 1\}]} \sqrt{n} (\mathbb{E}_n[\Delta Y \cdot 1\{D = 1\}] - \mathbb{E}[\Delta Y \cdot 1\{D = 1\}]). \end{aligned}$$



# 面板数据下的 $2 \times 2$ 双重差分法估计量的渐近正态性

- 大数定律下，红色部分在概率上收敛为零
- 中心极限定理下，蓝色部分的分布收敛到有限方差的正态分布 (Normal with finite variance)
- 然后，根据斯卢茨基定理 (Slutsky's Theorem) 可得：

$$\begin{aligned} & \sqrt{n} \left( \frac{\mathbb{E}_n[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}_n[1\{D = 1\}]} - \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}[1\{D = 1\}]} \right) \\ &= \frac{1}{\mathbb{E}[1\{D = 1\}]} \sqrt{n} (\mathbb{E}_n[\Delta Y \cdot 1\{D = 1\}] - \mathbb{E}[\Delta Y \cdot 1\{D = 1\}]) \\ & \quad - \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}[1\{D = 1\}]^2} \sqrt{n} (\mathbb{E}_n[1\{D = 1\}] - \mathbb{E}[1\{D = 1\}]) \\ & \quad + o_p(1). \end{aligned}$$

## 面板数据下的 $2 \times 2$ 双重差分法估计量的渐近正态性

- 调整一下顺序可得：

$$\begin{aligned} & \sqrt{n} \left( \frac{\mathbb{E}_n[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}_n[1\{D = 1\}]} - \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}[1\{D = 1\}]} \right) \\ &= \sqrt{n} \mathbb{E}_n \left[ \frac{\Delta Y \cdot 1\{D = 1\}}{\mathbb{E}[1\{D = 1\}]} \right] - \sqrt{n} \mathbb{E} \left[ \frac{\Delta Y \cdot 1\{D = 1\}}{\mathbb{E}[1\{D = 1\}]} \right] \\ & \quad - \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}[1\{D = 1\}]} \sqrt{n} \left( \mathbb{E}_n \left[ \frac{1\{D = 1\}}{\mathbb{E}[1\{D = 1\}]} \right] - 1 \right) + o_p(1) \\ &= \sqrt{n} \mathbb{E}_n \left[ \frac{\Delta Y \cdot 1\{D = 1\}}{\mathbb{E}[1\{D = 1\}]} \right] - \sqrt{n} \mathbb{E} \left[ \frac{\Delta Y \cdot 1\{D = 1\}}{\mathbb{E}[1\{D = 1\}]} \right] \\ & \quad - \sqrt{n} \mathbb{E}_n \left[ \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}[1\{D = 1\}]} \frac{1\{D = 1\}}{\mathbb{E}[1\{D = 1\}]} \right] + \sqrt{n} \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}[1\{D = 1\}]} + o_p(1). \end{aligned}$$

## 面板数据下的 $2 \times 2$ 双重差分法估计量的渐近正态性

- 继续变形:

$$\begin{aligned} & \sqrt{n} \left( \frac{\mathbb{E}_n[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}_n[1\{D = 1\}]} - \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}[1\{D = 1\}]} \right) \\ &= \sqrt{n} \mathbb{E}_n \left[ \frac{\Delta Y \cdot 1\{D = 1\}}{\mathbb{E}[1\{D = 1\}]} \right] - \sqrt{n} \mathbb{E}_n \left[ \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}[1\{D = 1\}]} \frac{1\{D = 1\}}{\mathbb{E}[1\{D = 1\}]} \right] + o_p(1) \\ &= \sqrt{n} \mathbb{E}_n \left[ \frac{\Delta Y \cdot 1\{D = 1\}}{\mathbb{E}[1\{D = 1\}]} - \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}[1\{D = 1\}]} \frac{1\{D = 1\}}{\mathbb{E}[1\{D = 1\}]} \right] + o_p(1) \\ &= \sqrt{n} \mathbb{E}_n \left[ \frac{1\{D = 1\}}{\mathbb{E}[1\{D = 1\}]} \left( \Delta Y - \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}[1\{D = 1\}]} \right) \right] + o_p(1) \end{aligned}$$

# 面板数据下的 $2 \times 2$ 双重差分法估计量的渐近正态性

- 继续变形:

$$\begin{aligned} & \sqrt{n} \left( \frac{\mathbb{E}_n[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}_n[1\{D = 1\}]} - \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}[1\{D = 1\}]} \right) \\ &= \sqrt{n} \mathbb{E}_n \left[ \frac{1\{D = 1\}}{\mathbb{E}[1\{D = 1\}]} \left( \Delta Y - \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 1\}]}{\mathbb{E}[1\{D = 1\}]} \right) \right] + o_p(1) \\ &= \sqrt{n} \frac{1}{n} \sum_{i=1}^n \underbrace{\left( \frac{1\{D_i = 1\}}{\mathbb{E}[1\{G = 2\}]} \left( \Delta Y_i - \frac{\mathbb{E}[\Delta Y_i \cdot 1\{G = 2\}]}{\mathbb{E}[1\{G = 2\}]} \right) \right)}_{=\xi_{i,G=2}} + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_{i,G=2} + o_p(1), \end{aligned}$$

- $\xi_{i,G=2}$  即影响函数: 均值为零, 有限方差, 且独立同分布

## 面板数据下的 $2 \times 2$ 双重差分法估计量的渐近正态性

- 重复相同的步骤，可得：

$$\begin{aligned} & \sqrt{n} \left( \frac{\mathbb{E}_n[\Delta Y \cdot 1\{D = 0\}]}{\mathbb{E}_n[1\{G = 0\}]} - \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 0\}]}{\mathbb{E}[1\{G = 0\}]} \right) \\ &= \sqrt{n} \mathbb{E}_n \left[ \frac{1\{D = 0\}}{\mathbb{E}[1\{D = 0\}]} \left( \Delta Y - \frac{\mathbb{E}[\Delta Y \cdot 1\{D = 0\}]}{\mathbb{E}[1\{D = 0\}]} \right) \right] + o_p(1) \\ &= \sqrt{n} \frac{1}{n} \sum_{i=1}^n \underbrace{\left( \frac{1\{D_i = 0\}}{\mathbb{E}[1\{G = 0\}]} \left( \Delta Y_i - \frac{\mathbb{E}[\Delta Y_i \cdot 1\{G = 0\}]}{\mathbb{E}[1\{G = 0\}]} \right) \right)}_{=\xi_{i,G=0}} + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_{i,G=0} + o_p(1), \end{aligned}$$

## 2 × 2 双重差分估计量的统计推断

- 这部分我们主要关注面板数据，因为重复截面数据与之很类似
- 根据双重差分法的估计量得到平均处理效应 ATT，我们希望对此进行统计推断
  - › 我们要做假设检验
  - › 我们要构建置信区间
- 然后我们需要估计渐近方差

$$\hat{V}_{n,p} = \mathbb{E}_n \left[ \hat{\xi}_{G=2}^2 \right] + \mathbb{E}_n \left[ \hat{\xi}_{G=0} \right]$$

其中  $d \in \{0, 1\}$

$$\hat{\xi}_{i,G=g} = \frac{1 \{D_i = g\}}{\mathbb{E}_n[1 \{G = g\}]} \left( \Delta Y_i - \frac{\mathbb{E}_n[\Delta Y \cdot 1 \{G = g\}]}{\mathbb{E}_n[1 \{D_i = g\}]} \right)$$

- 我们怎么样才能做到呢？

## 2 × 2 双重差分估计量的统计推断

- 类比原则再度出击！
  - › 用样本均值替代总体期望！

- 渐近方差的估计量

$$\widehat{V}_{n,p} = \mathbb{E}_n \left[ \widehat{\xi}_{G=2}^2 \right] + \mathbb{E}_n \left[ \widehat{\xi}_{G=0} \right]$$

其中对于  $d \in \{0, 1\}$

$$\widehat{\xi}_{i,G=g} = \frac{1 \{D_i = g\}}{\mathbb{E}_n[1 \{G = g\}]} \left( \Delta Y_i - \frac{\mathbb{E}_n[\Delta Y \cdot 1 \{G = g\}]}{\mathbb{E}_n[1 \{D_i = g\}]} \right)$$

## 2 × 2 双重差分估计量的统计推断

- 易证，当  $n \rightarrow 0, \hat{V}_n \xrightarrow{p} V$  时，估计出的标准误如下：

$$\widehat{Se}_{n,p} \left( \hat{\theta}_n^{\text{DiD}} \right) = \sqrt{\frac{\hat{V}_{n,p}}{n}}$$

- 其中标准误聚类到横截面单元层面：允许同一个体 不同时期内存在任意的相关性
- 基于渐近正态性，平均处理效应 ATT 的 95% 置信区间如下：

$$\hat{\theta}_n^{\text{DiD}} \pm 1.96 \cdot \widehat{Se}_{n,p} \left( \hat{\theta}_n^{\text{DiD}} \right)$$

- 假设检验的原假设  $H_0 : \text{ATT} = c$ ，对于任意  $c \in \mathbb{R}$  通过 T 检验：

$$t - \text{stat} = \frac{\hat{\theta}_n^{\text{DiD}} - c}{\widehat{se}_{n,p} \left( \hat{\theta}_n^{\text{DiD}} \right)}$$



# Stata实例

```
. diff fte, t(nj) p(after) robust
```

## DIFFERENCE-IN-DIFFERENCES ESTIMATION RESULTS

Number of observations in the DIFF-IN-DIFF: 698

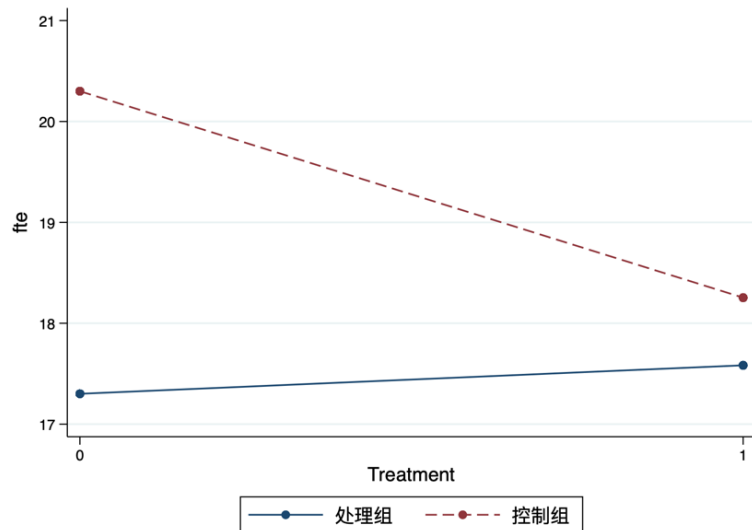
	Before	After		
Control:	65	65	130	
Treated:	284	284	568	
	349	349		
Outcome var.	fte	S. Err.	t	P> t
Before				
Control	20.300			
Treated	17.301			
Diff (T-C)	-2.999	1.591	-1.88	0.060*
After				
Control	18.254			
Treated	17.584			
Diff (T-C)	-0.670	1.093	0.61	0.540
Diff-in-Diff	2.329	1.931	1.21	0.228

R-square: 0.01

\* Means and Standard Errors are estimated by linear regression

\*\*Robust Std. Errors

\*\*Inference: \*\*\* p<0.01; \*\* p<0.05; \* p<0.1



## 参考文献

- Abbring, Jaap H. and Gerard J. van den Berg, “The nonparametric identification of treatment effects in duration models,” *Econometrica*, 2003, 71 (5), 1491–1517.
- Callaway, Brantly and Pedro H. C. Sant’Anna, “Difference-in-Differences with Multiple Time Periods,” *Journal of Econometrics*, 2021, (2), 200–230.
- Card, David and Alan Krueger, “Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania,” *American Economic Review*, 1994, 84 (4), 772–793.
- Carey, Colleen M., Sarah Miller, and Laura R. Wherry, “The Impact of Insurance Expansions on the Already Insured: The Affordable Care Act and Medicare†,” *American Economic Journal: Applied Economics*, 2020, 12 (4), 288–318.
- Currie, Janet, Henrik Kleven, and Esmée Zwieters, “Technology and Big Data Are Changing Economics: Mining Text to Track Methods,” *AEA Papers and Proceedings*, May 2020, 110, 42–48.

## 参考文献

- Dube, Arindrajit, T William Lester, and Michael Reich, “Minimum Wage Effects across State Borders: Estimates Using Contiguous Counties,” *The Review of Economics and Statistics*, 2010, 92 (4), 945–964.
- , T. William Lester, and Michael Reich, “Minimum wage shocks, employment flows, and labor market frictions,” *Journal of Labor Economics*, 2016, 34 (3), 663–704.
- Malani, Anup and Julian Reif, “Interpreting pre-trends as anticipation: Impact on estimated treatment effects from tort reform,” *Journal of Public Economics*, 2015, 124, 1–17.
- Meyer, Bruce D., W. Kip Viscusi, and David L. Durbin, “Workers’ Compensation and Injury Duration: Evidence from a Natural Experiment,” *The American Economic Review*, 1995, 85 (3), 322–340.
- Sianesi, Barbara, “An evaluation of the Swedish system of active labor market programs in the 1990s,” *The Review of Economics and Statistics*, 2004, 86 (1), 133–155.
- Venkataramani, Atheendar S., Sachin J. Shah, Rourke O’Brien, Ichiro Kawachi, and Alexander C. Tsai, “Health consequences of the US Deferred Action for Childhood Arrivals (DACA) immigration programme: a quasi-experimental study,” *The Lancet Public Health*, 2017, 2 (4), e175–e181.