应用微观计量经济学

Applied Microeconometrics

江　艇

中国人民大学经济学院

Last updated: March 18, 2019

Time:	周一晚 18:00-21:15
Venue:	明法 0502
Office:	明主 952
E-mail:	`ting.jiang@ruc.edu.cn`
Phone:	8250-0201
Grading:	期末闭卷考试

treatment effect

Lecture 3　潜在结果分析框架

causal effect

- 定义处理效应（因果效应）。

potential outcome

$$Y_i^0 = \text{不对个体 } i \text{ 实施某种处理的潜在结果}$$

$$Y_i^1 = \text{对个体 } i \text{ 实施某种处理的潜在结果}$$

$D_i = 1$

$Y_i = Y_i^1$

则个体处理效应为

$$\tau_i \equiv Y_i^1 - Y_i^0$$

$Y_{i0}^0 = Y_{i0}$ → Y^0 unobs.

- 观测到的处理状态

$Y_j = Y_j^0$

$$D_i = \begin{cases} 1 & \text{个体 } i \text{ 接受处理，进入处理组} \\ 0 & \text{个体 } i \text{ 不接受处理，进入控制组} \end{cases}$$

$D_i = 0$

观测到的结果

$Y_i = Y_i^0$

$$Y_i = (1 - D_i) \cdot Y_i^0 + D_i \cdot Y_i^1$$

Y^1 unobs.

Y_i

Y_i

treatment

- 关于这一定义的几点评论：
  - 因果效应的定义只取决于潜在结果，而不取决于究竟哪个潜在结果被实际观测到（实际发生）。
  - 因果效应永远是在同一时点上的不同结果之间的比较，而且处理必须发生在结果之前，特别地，事前事后比较不是因果效应。
  - 因果推断的基本难题就是数据缺失：我们永远只能至多观察到一个潜在结果。没有观察到的那个潜在结果叫做反事实。counterfactual
  - 因果分析的关键就是构造反事实，也就是个插值问题。而且反事实只能从实际观测到的、处理状态相反的其他结果中去寻找，要么是其他个体的潜在结果，要么是同一个体其他时期的潜在结果。我们把这样的其他个体或同一个体的其他时期称作对照组，因此也可以说因果分析的关键就是寻找好的对照组。 comparison.
  - 我们实际衡量因果效应时必须涉及多个个体或同一个体的不同时期，但无论是横截面比较还是事前事后比较，都不是因果效应的定义，尽管它们对于衡量因果效应很重要。

- 多个个体的存在本身并无助于解决因果推断的基本难题：处理水平 (treatment levels) 和潜在结果数量爆炸式增长。

  想象一下：二元处理，两个个体，将会存在四种处理水平，四个潜在结果，六种处理效应，但只有一个结果被观测到。

  所以我们才需要——

Stable unit treatment value assumption (SUTVA). 每个个体的潜在结果不会因为其他个体接受的处理水平的不同而不同；每个个体所接受到的每个处理都只会产生一种潜在结果。

这代表两层含义：

- No interference. 不存在同侪效应 (peer effects)、溢出效应 (spill-over effects)、一般均衡效应 (general equilibrium effects).
- No hidden variations in treatment. 实施 treatment 的方式不重要（不存在 Hawthorne effect, carry-over effect 等）。

A   B

$$D_A = \begin{cases} 1 & \text{if } A \text{ treated} \\ 0 & \text{if } A \text{ not treated} \end{cases} \qquad D_B = \begin{cases} 1 & \text{if } B \text{ treated} \\ 0 & \text{if } B \text{ not treated.} \end{cases}$$

4 treatment levels:

1. $D_A = 1, \; D_B = 1$
2. $D_A = 1, \; D_B = 0$
3. $D_A = 0, \; D_B = 1$
4. $D_A = 0, \; D_B = 0$

4 potential outcomes for each of A and B

$$Y_A^1 \quad Y_A^2 \quad Y_A^3 \quad Y_A^4$$

$$Y_B^1 \quad Y_B^2 \quad Y_B^3 \quad Y_B^4$$

$C_4^2 = 6$ causal effects for each of A and B

w/ A [illegible]

$$\tau_{A,12} = Y_A^1 - Y_A^2$$

$$\tau_{A,13} = Y_A^1 - Y_A^3$$

$$\vdots$$

- SUTVA 很关键，但很难检验。

示例 1. Fowlie et al (2012, *AER*)

结果：污染物排放量

处理：是否进入排污权交易项目

Control group	Levels	Logs	RECLAIM facilities	Controls
Panel A. Change in NO_x emissions between periods 1 and 4				
Base specification	−20.59***	−0.25***	212	1,222
	(7.63)	(0.09)		
Exclude L.A. facilities	−23.50***	−0.34***	210	778
	(7.96)	(0.09)		
Exclude northern CA	−26.60***	−0.23**	210	767
	(7.58)	(0.11)		
Severe nonattainment only	−21.65**	−0.29**	208	475
	(7.89)	(0.11)		
Single facility only	−19.92**	−0.23**	210	781
	(7.60)	(0.10)		

- 逻辑似乎是：

$$\text{因为：SUTVA 不成立} \Rightarrow \text{事件 A 发生}$$
$$\text{所以：事件 A 未发生} \Rightarrow \text{SUTVA 成立}$$

- 实际上是：

$$\text{因为：事件 A 发生} \Rightarrow \text{SUTVA 不成立}$$
$$\text{所以：事件 A 未发生} \Rightarrow \text{SUTVA 成立}$$

- 因为我们并不能穷尽 SUTVA 不成立时可能发生的所有事件，当 SUTVA 不成立时，可能是事件 A 发生，也可能事件 A 不发生但事件 B 发生。所以对 SUTVA 的检验是一种"间接检验 (indirect test)"。

被估计量

- 平均处理效应 (average treatment effect, ATE)

$$\tau \equiv E\left(Y_i^1 - Y_i^0\right) = E(\tau_i)$$

- 处理组的平均处理效应 (average treatment effect on the treated, ATT)

$$\tau_1 \equiv E\left(Y_i^1 - Y_i^0 | D_i = 1\right)$$

- 控制组的平均处理效应 (average treatment effect on the untreated, ATU)

$$\tau_0 \equiv E\left(Y_i^1 - Y_i^0 | D_i = 0\right)$$

- 定义在协变量上的条件平均处理效应

$$\tau(x) \equiv E\left(Y_i^1 - Y_i^0 | X_i = x\right)$$
$$\tau_1(x) \equiv E\left(Y_i^1 - Y_i^0 | X_i = x, D_i = 1\right)$$
$$\tau_0(x) \equiv E\left(Y_i^1 - Y_i^0 | X_i = x, D_i = 0\right)$$

根据期望迭代定律,

$$\tau = E\left\{E\left(Y_i^1 - Y_i^0 | X_i = x\right)\right\} = \int \tau(x)\, dF_X$$
$$\tau_1 = E\left\{E\left(Y_i^1 - Y_i^0 | X_i = x, D_i = 1\right)\right\} = \int \tau_1(x)\, dF_{X|D=1}$$
$$\tau_0 = E\left\{E\left(Y_i^1 - Y_i^0 | X_i = x, D_i = 0\right)\right\} = \int \tau_0(x)\, dF_{X|D=0}$$

$E(Y|D=1,X)$
$-E(Y|D=0,X)$

- 为什么主要研究均值意义上的处理效应？因为 $E(\cdot)$ 是线性的。

$$E(Y^1 - Y^0) = E(Y^1) - E(Y^0)$$

例如，$Med(Y^1 - Y^0) \neq Med(Y^1) - Med(Y^0)$，前者 > 0 表示至少一半人受益，后者 > 0 表示 median person 受益（试想一例，所有人都受损，只有中产阶级受益）。目前关于后者 (treatment effect on distributions) 的研究较多，但前者 (distributional treatment effect) empirically more relevant（$Med(Y^1)$ 和 $Med(Y^0)$ 很可能不是同一个人）。

分配机制与识别假设

- 分配机制是指每个个体如何接受处理。理想的情况是随机分组。另一个极端是下面的“完美医生”的例子。

Unit	Potential Outcomes		Causal Effects
	Surgery	Drug	S-D
Patient #1	7	1	6
Patient #2	5	6	-1
Patient #3	5	1	4
Patient #4	7	8	-1
Average Effect			2

Unit	Treatment	Observed Outcome
Patient #1	S	7
Patient #2	D	6
Patient #3	S	5
Patient #4	D	8
Average Difference		-1

6 7

- 因果识别依赖于假设，因果推断的识别假设通常是关于分配机制的假设。

$E(Y^1|D)=E(Y^1)$

- 识别假设 I：潜在结果均值独立于处理状态。

$E(Y^1|D=1)=E(Y^1|D=0)$

$$E(Y^d|D) = E(Y^d),\ d = 0, 1$$

$E(Y^0|0)=E(Y^0)$

  - 此时组间均值差异能够识别 ATE。

$E(Y^0|D=1)=E(Y^0|D=0)$

$$\begin{aligned}
& E(Y|D=1) - E(Y|D=0) \\
=& E(Y^1|D=1) - E(Y^0|D=0) \\
=& E(Y^1) - E(Y^0) \\
=& E(Y^1 - Y^0)
\end{aligned}$$

识别假设

  - 要想识别 ATT，需要假定 $E(Y^0|D) = E(Y^0)$。
    要想识别 ATU，需要假定 $E(Y^1|D) = E(Y^1)$。
  - 这一假设是随机实验研究的识别假设。

$$\hat{\tau} = \bar{Y}_T - \bar{Y}_C$$

估计

$$\xrightarrow{p} E(y|D=1) - E(y|D=0)$$

$\|$ 识别

$$\tau = E(Y^1 - Y^0)$$

(ATE)

$$\bar{Y}_T = \frac{1}{n_T} \sum_{i \in T} y_i \xrightarrow{p} E(y|D=1)$$

$$\bar{Y}_C = \frac{1}{n_C} \sum_{i \in C} y_i \xrightarrow{p} E(y|D=0)$$

ATT:

$$E(Y|D=1) - E(Y|D=0)$$
$$= E(Y^1|D=1) - E(Y^0|D=0)$$
$$= E(Y^1|D=1) - E(Y^0|D=1)$$
$$\tau_1 = E(Y^1 - Y^0|D=1)$$

obs.

unobs.

ATU

$$E(Y|D=1) - E(Y|D=0)$$
$$= E(Y^1|D=1) - E(Y^0|D=0)$$
$$= E(Y^1|D=0) - E(Y^0|D=0)$$
$$\tau_0 = E(Y^1 - Y^0|D=0)$$

obs

unobs.

(只)关于可以识别
变量的选择性

$$E(Y^d|D) \neq E(Y^d) \quad E(Y^d|D=1) \neq E(Y^d|D=0)$$

- 识别假设 II：给定可观测变量，潜在结果均值独立于处理状态。

$$E\left(Y^{d}|D,X\right)=E\left(Y^{d}|X\right),\ d=0,1$$

– 此时可观测变量相同条件下的组间均值差异能够识别 $\tau(x)$.

$\hat{\tau}=\bar{Y}_T(X=x)-\bar{Y}_C(X=x)$

$$\begin{aligned}&E(Y|X,D=1)-E(Y|X,D=0)\\=&E(Y^1|X,D=1)-E(Y^0|X,D=0)\\=&E(Y^1|X)-E(Y^0|X)\\=&E(Y^1-Y^0|X)\\=&\tau(X)\end{aligned}$$

ATE(X)

进而能够识别 ATE

$$\tau=E_X\left[\tau(x)\right]$$

– 要想识别 $\tau_1(x)$, 进而识别 ATT, 需要假定 $E\left(Y^0|D,X\right)=E\left(Y^0|X\right)$。
  要想识别 $\tau_0(x)$, 进而识别 ATU, 需要假定 $E\left(Y^1|D,X\right)=E\left(Y^1|X\right)$。

– 这一假设是匹配方法（以及作为其特例的 OLS 方法）的识别假设。

$$E(Y^d \mid D, X) = E(Y^d \mid X)$$

$$E(Y^1 \mid D, X) = E(Y^1 \mid X)$$

✓ ① $$E(Y^1 \mid X=x, D=1) = E(Y^1 \mid X=x, D=0)$$

$$E(Y^0 \mid D, X) = E(Y^0 \mid X)$$

✓ ② $$E(Y^0 \mid X=x, D=1) = E(Y^0 \mid X=x, D=0)$$

		D=1	D=0			
$X=1$	$n_1 =$	n_{1t}	$+ n_{1c}$	$\tau(x=1)$	$\tau_1(x=1)$	$\tau_0(x=1)$
$X=2$	$n_2 =$	n_{2t}	$+ n_{2c}$	$\tau(x=2)$	$\tau_1(x=2)$	$\tau_0(x=2)$
$X=3$	$n_3 =$	n_{3t}	$+ n_{3c}$	$\tau(x=3)$	$\tau_1(x=3)$	$\tau_0(x=3)$

$$\hat{\tau} = \bar{Y}_T(x=x) - \bar{Y}_C(x=x)$$

$$\frac{n_1 \cdot \tau(x=1) + n_2 \cdot \tau(x=2) + n_3 \cdot \tau(x=3)}{n_1 + n_2 + n_3} = \tau$$

$$\frac{n_{1t} \cdot \tau(x=1) + n_{2t} \cdot \tau(x=2) + n_{3t} \cdot \tau(x=3)}{n_{1t} + n_{2t} + n_{3t}} = \tau_1$$

$$\frac{n_{1c} \cdot \tau(x=1) + n_{2c} \cdot \tau(x=2) + n_{3c} \cdot \tau(x=3)}{n_{1c} + n_{2c} + n_{3c}} = \tau_0$$

ATT: $\tau_1 = E(Y^1 - Y^0 \mid D=1)$

$$\Uparrow$$

$$\tau_1(X) = E(Y^1 - Y^0 \mid D=1, X)$$

$$= E(Y^1 \mid D=1, X) - \boxed{E(Y^0 \mid D=1, X)}$$

$$= E(Y^1 \mid D=1, X) - \boxed{E(Y^0 \mid D=0, X)}$$

$$E(Y \mid D=1, X) - E(Y \mid D=0, X)$$

ATU: $\tau_0 = E(Y^1 - Y^0 \mid D=0)$

$$\Uparrow$$

$$\tau_0(X) = E(Y^1 - Y^0 \mid X, D=0)$$

$$\Uparrow$$

$$E(Y \mid D=1, X) - E(Y \mid D=0, X)$$

如何在潜在结果框架下理解 OLS？

- 写下一个线性模型：

$$Y_i = \alpha + \beta D_i + U_i$$

$$Y_i^1 = \underbrace{E(Y^1)}_{\alpha_1} + u_i^1$$

$$Y_i^0 = \underbrace{E(Y^0)}_{\alpha_0} + u_i^0$$

- 定义潜在结果：

$$Y_i^d = E(Y_i^d) + U_i^d \equiv \alpha_d + U_i^d,\ d = 0, 1$$

- 斜率系数就是 ATE。

$$\begin{aligned}
Y_i =& (1 - D_i)Y_i^0 + D_i Y_i^1 \\
=& (1 - D_i)(\alpha_0 + U_i^0) + D_i(\alpha_1 + U_i^1) \\
=& \alpha_0 + \underbrace{(\alpha_1 - \alpha_0)}_{\text{ATE}} D_i + \boxed{[D_i U_i^1 + (1 - D_i)U_i^0]}
\end{aligned}$$

$$\alpha_1 - \alpha_0 = E(Y^1) - E(Y^0) = E(Y^1 - Y^0)$$

$$Cov(X, Y) = E(X \cdot Y) - E(X)E(Y)$$

- OLS 估计的一致性要求

$$Cov(D, DU^1 + (1-D)U^0) = 0$$

$$\begin{aligned}
& LHS \\
&= E(DU^1) - E(D)E(DU^1 + (1-D)U^0) \\
&= (1-E(D))E(DU^1) - E(D)E((1-D)U^0) \\
&= (1-E(D))E(U^1|D=1)P(D=1) - E(D)E(U^0|D=0)P(D=0) \\
&= E(D)\big(1-E(D)\big)\Big[E\big(U^1|D=1\big) - E\big(U^0|D=0\big)\Big] = 0
\end{aligned}$$

- 如果识别假设 I 成立，$E\big(Y^d|D\big) = E\big(Y^d\big)$，则 $E(U^d|D) = E(U^d) = 0$，因此 $E\big(U^1|D=1\big) - E\big(U^0|D=0\big) = 0$，即 $LHS = 0$ 成立。**识别假设 I 等价于线性回归模型 OLS 的识别假设。**

$$\begin{aligned}
\hat{\beta} &= \overline{Y}_{|D=1} - \overline{Y}_{|D=0} \\
&\to_p \beta + \frac{Cov(U, D)}{Var(D)} \\
&= (\alpha_1 - \alpha_0) + \Big[E\big(U^1|D=1\big) - E\big(U^0|D=0\big)\Big]
\end{aligned}$$