

# 第三章 未观测混杂

罗珊珊

March, 2023

# 目录

## ① 工具变量

同质性假定-无协变量

同质性假定-有协变量

同质性假定-其他

同质性假定，无效 IV 的情形

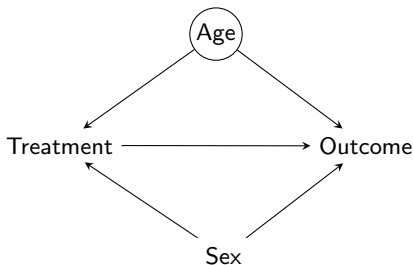
单调性假设

## ② 阴性对照

## ③ 双重差分

## 未观测混杂

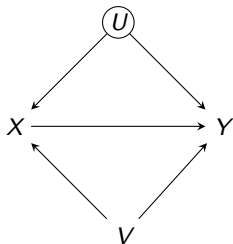
- 匹配、逆概加权、回归和双稳健估计方法的重要前提是可忽略性假定。
- 然而, 在实际研究中, 如果有重要背景变量未被观测、测量误差或者选择偏差, 就有潜在的未观测的混杂因素, 可忽略性假定可能不成立, 前一节介绍的统计推断方法在出现未观测的混杂因素时就有偏差。



## 未观测混杂

- 当存在未被观测的混杂因素时, 更合理的假定是潜在可忽略性:  
存在未被观测的变量  $U$  满足  $Y_x \perp\!\!\!\perp X \mid (U, V)$ , 其中  $V$  为观测的混杂因素.
- 潜在可忽略性假定:  $Y_x \perp\!\!\!\perp X \mid (U, V)$ . 在  $U$  是常数时, 此假定退化为可忽略性假定.  
在潜在可忽略性假定下,

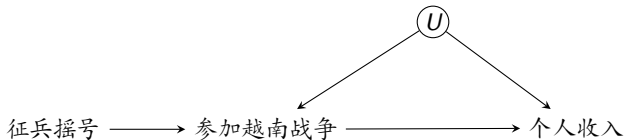
$$\mathbb{E}(Y_x) = \mathbb{E}\{\mathbb{E}(Y \mid X = x, U)\} \neq \mathbb{E}(Y \mid X = x)$$



- 如果  $U$  没有被观测,
  - 那么  $\mathbb{E}(Y | X = x, U)$  一般不能由观测数据识别, 因此,  $\mathbb{E}(Y_x)$  的识别性不能保证.
  - 如果用  $\mathbb{E}(Y | X = x)$  来估计  $\mathbb{E}(Y_x)$  就产生偏差.
- 在潜在可忽略性假定下, **辅助变量** 经常被用来帮助识别因果作用和消除混杂偏倚. 辅助变量通常只与  $(X, Y, U)$  三个变量的一个子集相关, 因此引入一些条件独立性帮助识别因果作用.
- 在潜在可忽略性假定下用来消除混杂偏差的两种方法,
  - 一种是常用的工具变量 (instrumental variable) 方法
  - 一种是最近引起人们注意的阴性对照变量 (negative control variable) 方法.

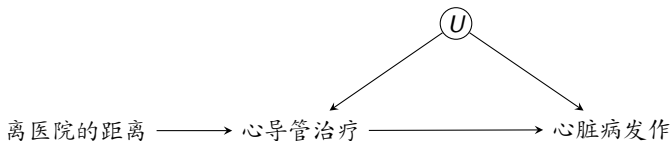
## 工具变量的例子

- J. Angrist (1990) 研究了参加越南战争对个人收入的影响. Angrist 指出, 越南战争的征兵摇号提供了一个自然试验 (natural experiment), 从而产生了一个工具变量. 年轻人摇号决定了他们是否会被征召去服役于越南战争 (研究生, 自愿参军).



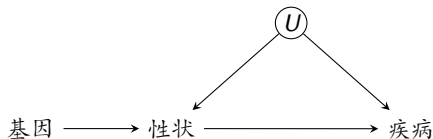
## 工具变量的例子

- McClellan et al. (1994, JAMA) 研究了心导管治疗对患者心脏病发作的影响. McClellan 指出一个人住的地方离先进医院有多近在很大程度上是随机的 (自然实验), 从而产生一个工具变量为, 指出可以使用两类医院的距离作为工具变量: the differential distance the patient lives from the nearest hospital that performs cardiac catheterization to the nearest hospital that does not perform it.



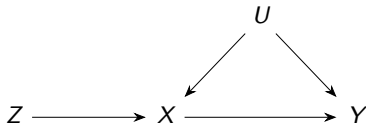
## 工具变量的例子

- 孟德尔随机化是工具变量分析的一种特殊类型，它与工具变量分析有着相同的基本原理和识别假设。孟德尔随机化以基因作为工具变量，对感兴趣的性状（原因变量）与疾病（结果变量）进行因果分析，并常使用基因作为工具变量。





## 工具变量 I



- $X$ : 处理变量 ( $X$  非随机化).
- $Y$ : 感兴趣的结果变量.
- $U$ : 未观测混杂 (协变量).
- $Z$ : 工具变量.

## 识别假设

三个必要的假设:

- 排除性限制:  $Z \perp\!\!\!\perp Y \mid (X, U)$
- 独立性:  $Z \perp\!\!\!\perp U$
- 相关性:  $Z \not\perp\!\!\!\perp X$

除了需要条件 (i)-(iii), 工具变量还需要额外的假定, 有两种常用的假定被用在工具变量的分析中:

- 一种是因果作用的同质性假定 (effect homogeneity),
- 一种是单调性假定 (monotonicity).

# Outline

## ① 工具变量

同质性假定-无协变量

同质性假定-有协变量

同质性假定-其他

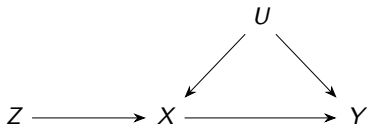
同质性假定，无效 IV 的情形

单调性假设

## ② 阴性对照

## ③ 双重差分

## 同质性假定-线性模型



- 同质性假定 (effect homogeneity) 最常用的版本即为在经济学和社会学中广泛应用的 结构方程模型 (structural equation model).
- 工具变量被用来估计结构方程中处理或暴露变量的回归系数. 我们首先考虑最简单的情形, 即没有其余的协变量, 此时连续型结果变量的线性模型为

$$Y = \beta_0 + \beta_1 X + U,$$

其中  $U$  是未观测的混杂因素,  $\beta_1$  表示在其他因素 ( $U$ ) 不变的情形下,  $X$  每增加一个单位对  $Y$  的作用 (ceteris paribus effect).

## 同质性假定-线性模型

- 这个方程实际上假定  $X$  对  $Y$  的作用在所有人当中是一个常数. 这个方程也可以等价地用潜在结果表示为  $Y_x = \beta_0 + \beta_1 x + U$ . 此模型隐含了潜在可忽略性假定.
- 在此模型下, **排除性限制假定**也满足, 即  $Z \perp\!\!\!\perp Y \mid (X, U)$ .
- 在此模型下,  $\beta_1$  代表平均因果作用  $\mathbb{E}(Y_1 - Y_0)$ . 由于存在未观测的混杂因素, 因此仅从此结构方程不能识别  $\beta_1$ .
- 例如, 当  $\mathbb{E}(U \mid X) \neq 0$  时,  $\beta_1$  的最小二乘估计有偏.

## 同质性假定-线性模型 I

在线性模型假设下, 我们可以识别参数  $\beta_1$ :

$$Y = \beta_0 + \beta_1 X + U,$$

证明.

$$\begin{aligned}\text{Cov}(Y, Z) &= \text{Cov}(\beta_0 + \beta_1 X + U, Z) \\ &= \beta_0 \text{Cov}(1, Z) + \beta_1 \text{Cov}(X, Z) + \text{Cov}(U, Z) \\ &= \beta_1 \text{Cov}(X, Z). \text{ 独立性假定} \\ \Rightarrow \beta_1 &= \frac{\text{Cov}(Y, Z)}{\text{Cov}(X, Z)} = \frac{\mathbb{E}(YZ) - \mathbb{E}(Y)\mathbb{E}(Z)}{\mathbb{E}(XZ) - \mathbb{E}(X)\mathbb{E}(Z)}. \text{ 相关性假定}\end{aligned}$$



## 同质性假定-线性模型 II

二值情形.

$$\begin{aligned}\beta_1 &= \frac{\mathbb{E}(YZ) - \mathbb{E}(Y)\mathbb{E}(Z)}{\mathbb{E}(XZ) - \mathbb{E}(X)\mathbb{E}(Z)} \\&= \frac{\mathbb{E}(YZ) - \mathbb{E}\{YZ + Y(1-Z)\}\mathbb{E}(Z)}{\mathbb{E}(XZ) - \mathbb{E}\{XZ + X(1-Z)\}\mathbb{E}(Z)} \\&= \frac{\mathbb{E}(YZ)\mathbb{E}(1-Z) - \mathbb{E}\{Y(1-Z)\}\mathbb{E}(Z)}{\mathbb{E}(XZ)\mathbb{E}(1-Z) - \mathbb{E}\{X(1-Z)\}\mathbb{E}(Z)} \\&= \frac{\frac{\mathbb{E}(YZ)}{\mathbb{E}(Z)} - \frac{\mathbb{E}\{Y(1-Z)\}}{\mathbb{E}\{(1-Z)\}}}{\frac{\mathbb{E}(XZ)}{\mathbb{E}(Z)} - \frac{\mathbb{E}\{X(1-Z)\}}{\mathbb{E}\{(1-Z)\}}} \\&= \frac{\mathbb{E}(Y \mid Z=1) - \mathbb{E}(Y \mid Z=0)}{\mathbb{E}(X \mid Z=1) - \mathbb{E}(X \mid Z=0)}.\end{aligned}$$



## 同质性假定-线性模型 III

- 在线性模型假设下,  $\beta_1$  的 IV 估计量是协方差之比:

$$\hat{\beta}_{1,\text{iv}} = \frac{\widehat{\text{Cov}}(Y_i, Z_i)}{\widehat{\text{Cov}}(X_i, Z_i)} = \frac{\frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})(Z_i - \bar{Z})}{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})(Z_i - \bar{Z})}$$

其中  $\bar{Y}, \bar{X}$  为样本均值.

- 对于二值工具变量  $Z$ , 这是 Wald 估计量:

$$\hat{\beta}_{1,\text{iv}} = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{X}_1 - \bar{X}_0},$$

其中  $\bar{Y}_z, \bar{X}_z$  是组  $z(=0, 1)$  上的样本均值.



## 解释 1-两次回归 I

但是利用一个满足假定 (i)-(iii) 的工具变量  $Z$ , 我们可以识别  $\beta_1$ . 我们考虑如下简约方程:

$$Y_i = \pi_{10} + \pi_{11}Z_i + \varepsilon_{1i},$$

$$X_i = \pi_{20} + \pi_{21}Z_i + \varepsilon_{2i}.$$

间接最小二乘 (ILS) 估计量是  $\pi_{11}$  和  $\pi_{21}$  的最小二乘估计之比:

$$\hat{\beta}_{1,LS} = \hat{\pi}_{11} / \hat{\pi}_{21}.$$

在二值的随机试验中,  $\hat{\beta}_{1,LS}$  是 ITT 估计的比值 (Angrist, Imbens, Rubin, 1996), 又叫 LATE, 局部平均因果作用.

- 给定观测数据, 把样本协方差  $\hat{\pi}_{11}$  和  $\hat{\pi}_{21}$  代入, 即得到  $\beta_1$  的工具变量估计 (instrumental variable estimator).

## 解释 1-两次回归 II

- 即使  $\mathbb{E}(U | X) \neq 0$ , 在假定 (i)-(iii) 和一定的正则条件下, 可以证明  $\hat{\beta}_{1,LS}$  的相合性和渐近正态性.
- 工具变量方法有效地缓解了未观测的混杂因素导致的偏差.

## 解释 2-两阶段最小二乘

第 1 阶段：利用 IV 预测处理值  $\hat{X}_i$ ：

$$\hat{X}_i = \hat{\pi}_{20} + \hat{\pi}_{21}Z_i$$

第 2 阶段：在结果模型中使用第 1 阶段的处理变量的预测值：

$$Y_i = \beta_0 + \beta_1\hat{X}_i + \eta_i$$

从第 2 阶段通过 OLS 估计  $\beta_1$ ，得到  $\beta_1$  的两阶段最小二乘估计量： $\hat{\beta}_{1,2sls}$ 。

- 直觉：

$D$  虽然被混杂了，但我们可以使用 IV 来恢复  $D$  的“未混淆部分”并插入结果模型。

- 容易验证：

$$\hat{\beta}_{1, iv} = \hat{\beta}_{1, ils} = \hat{\beta}_{1, 2sls}$$

# Outline

## ① 工具变量

同质性假定-无协变量

同质性假定-有协变量

同质性假定-其他

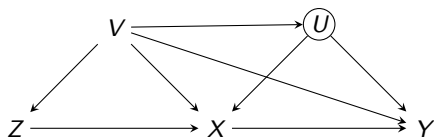
同质性假定，无效 IV 的情形

单调性假设

## ② 阴性对照

## ③ 双重差分

## 存在协变量时 I



当存在额外的协变量  $V$  时，此时三个必要的假设为：

- 排除性限制:  $Z \perp\!\!\!\perp Y \mid (X, U, V)$
- 独立性:  $Z \perp\!\!\!\perp U \mid V$
- 相关性:  $Z \not\perp\!\!\!\perp X \mid V$

## 同质性假定-toy example I

当存在协变量时，我们首先考虑以下两个线性模型，此时处理变量  $X$  要求是连续的：

$$X_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 V_i + U_i + \epsilon_{xi}.$$

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 V_i + U_i + \epsilon_{yi}.$$

其中  $\mathbb{E}(U_i | Z_i, V_i) = 0$ . 于是,

$$Y_i = \beta_0 + \beta_1(\alpha_0 + \alpha_1 Z_i + \alpha_2 V_i + U_i + \epsilon_x) + \beta_2 V_i + U_i + \epsilon_{yi}$$

$$= \beta_0 + \beta_1 \alpha_0 + \beta_1 \alpha_1 Z_i + (\beta_1 \alpha_2 + \beta_2) V_i + (\beta_1 + 1) U_i + \epsilon_{yi}$$

$$= \gamma_0 + \gamma_1 Z_i + \gamma_2 V_i + \gamma_3 U_i + \epsilon_{yi}.$$

$$X_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 V_i + U_i + \epsilon_x.$$

## 同质性假定-toy example II

(熟悉的同学第一段可以省略).

我们首先证明  $\alpha_1$  可识别,  $\gamma_1$  类似. 由于  $U$  无法观测, 我们只能求解最小二乘:

$$\min_{\alpha_0^*, \alpha_1^*, \alpha_2^*} \left\| \underbrace{\begin{pmatrix} 1 & Z_1 & V_1 \\ \vdots & \vdots & \vdots \\ 1 & Z_n & V_n \end{pmatrix}}_A \underbrace{\begin{pmatrix} \alpha_0^* \\ \alpha_1^* \\ \alpha_2^* \end{pmatrix}}_{\vec{\alpha}^*} - \underbrace{\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}}_X \right\|_2 = \min_b \|A\vec{\alpha}^* - X\|_2.$$

求解最小二乘, 我们有,

$$\begin{aligned} \vec{\alpha}^* &= \mathbb{E}\{(A^T A)\}^{-1} \mathbb{E}(A^T X) \\ &= \mathbb{E}\{(A^T A)\}^{-1} \mathbb{E}\{A^T (A\vec{\alpha} + U + \epsilon_x)\} \\ &= \alpha + \mathbb{E}\{(A^T A)\}^{-1} \mathbb{E}(A^T U) + \mathbb{E}\{(A^T A)\}^{-1} \mathbb{E}(A^T \epsilon_x) \\ &= \alpha + \mathbb{E}\{(A^T A)\}^{-1} \mathbb{E}(A^T U) + \vec{0}, \end{aligned}$$

## 同质性假定-toy example IV

其中

(熟悉的同学第一段可以省略).

$$\mathbb{E}(A^T U) = \mathbb{E}\{A^T \mathbb{E}(U | A)\} = \vec{0},$$



- 在  $\alpha_1$  和  $\gamma_1$  可识别的基础上, 我们可以识别

$$\beta_1 = \frac{\gamma_1}{\alpha_1}.$$



## 同质性假定-线性模型 I

不失一般性，我们中心化所有数据，省略截距项。实际上，我们并不要求  $X$  与  $Z$  之间是线性模型，即以下模型并不需要：

$$X = Z\alpha_1 + V + U\alpha_2 + \epsilon_x.$$

我们考虑以下的第二阶段的线性模型即可：

$$Y = X\beta_1 + V\beta_2 + U + \epsilon_y.$$

其中

$$\mathbb{E}(U \mid Z, V) = \mathbb{E}(U \mid V) = 0. \quad (1)$$

## 同质性假定-线性模型 II

证明.

$$\mathbb{E}(U \mid Z, V) = \mathbb{E}(Y - X\beta_1 - V\beta_2 \mid Z, V) = 0.$$

基于上述矩约束条件 (moment restriction), 我们可以构建如下等式:

$$\mathbb{E}(YZ - XZ\beta_1 - VZ\beta_2) = 0,$$

$$\mathbb{E}(YV - XV\beta_1 - V^2\beta_2) = 0.$$

令  $A = (Z, V)^T$ , 上式等价于

$$\mathbb{E}\{A(Y - X\beta_1 - V\beta_2)\} = 0. \quad (2)$$

两个方程, 两个未知数  $(\beta_1, \beta_2)$ , 恰好可识别. □

## 同质性假定-线性模型 I

- 当工具变量个数比“被混杂”的变量多时（粗糙的说，IV 个数多于处理个数时），此时方程个数将大于未知数个数。
- 利用广义矩估计 (generalized method of moments, GMM) 思想，我们依旧可以证明参数  $(\beta_1, \beta_2)$  可识别。

## Two stage least square I

计算 IV 估计值的一种计算方法是两阶段最小二乘 (2SLS 或 TSLS)。

第 1 阶段：利用 IV 预测处理值  $\hat{X}$ , ( $X = A^T \delta + \epsilon_x$ ):

$$\hat{\delta} = (AA^T)^{-1} AX$$

预测值为：

$$\hat{X} = A^T \hat{\delta} = A^T (AA^T)^{-1} AX = P_A X,$$

其中  $P_A$  是投影矩阵  $P_A = A^T (AA^T)^{-1} A$ .

## Two stage least square II

第 2 阶段：在结果模型中使用第 1 阶段的处理变量的预测值：

$$Y = \hat{X}\beta_1 + V\beta_2 + \eta_y.$$

上述回归的有效性是由于

①  $Y = \hat{X}\beta_1 + V\beta_2 + \underbrace{(X - \hat{X})\beta_1 + U}_{\epsilon},$

② 由回归分析知残差  $X - \hat{X}$  与预测值  $\hat{X}$  不相关，即  $\text{Cov}(\hat{X}, X - \hat{X}) = 0$ 。

③ 由于  $\hat{X}$  是  $X$  的预测值，是  $A$ （即  $Z$  和  $V$ ）的线性函数，故  $\text{Cov}(\hat{X}, U) = 0$ 。

④ 2 与 3 结合可知，1 中  $\text{Cov}(\hat{X}, \epsilon) = 0$ ，故  $Y$  可以直接对于预测值  $\hat{X}$  和  $V$  回归得到相合估计。

# Outline

## ① 工具变量

同质性假定-无协变量

同质性假定-有协变量

同质性假定-其他

同质性假定，无效 IV 的情形

单调性假设

## ② 阴性对照

## ③ 双重差分

## 同质性假定- Control function I

- 不失一般性, 我们将数据中心化. 假设

$$X = Z\alpha + U + \varepsilon_1 \quad (3)$$

$$Y = X\beta + \gamma U + \varepsilon_2, \quad (4)$$

$$\Rightarrow \tilde{U} = X - Z\alpha = U + \varepsilon_1,$$

注意第一阶段的模型(3)此时要求是可加模型.

- 此时会有  $X = \tilde{U} + Z\alpha$  及  $U = \tilde{U} - \varepsilon_1$ , 由于  $Z \perp \varepsilon_1$ , 我们会有  $X \perp U \mid \tilde{U}$ . 从而,

$$\begin{aligned} Y_x &= x\beta + \gamma U + \varepsilon_2 \\ &= x\beta + \gamma(\tilde{U} - \varepsilon_1) + \varepsilon_2 \\ &= x\beta + \gamma\tilde{U} - \gamma\varepsilon_1 + \varepsilon_2, \end{aligned}$$

## 同质性假定- Control function II

- 此时虽然  $U$  无法看见, 但我们构造了另一个可见的混杂  $\tilde{U}$  充分控制 (control) 住了  $U$  的混杂信息.
- 在模型(4)下, 我们有  $Y_x \perp\!\!\!\perp X \mid \tilde{U}$ , 且识别性可以如下得到保证,

$$E(Y_x) = E\left\{E\left(Y_x \mid \tilde{U}\right)\right\} = E\{E(Y \mid X = x, \tilde{U})\}.$$

- 我们可以如下通过回归估计参数  $\beta$ ,

$$Y \sim \hat{\beta}X + \hat{\gamma}\tilde{U}$$



## 同质性假定-非参数方法 I

- 非参数工具变量回归 (Newey & Powell, 2003)

$$E(Y | X, U) = g(X) + U, \quad E(U | Z) = 0,$$

$$E(Y | Z) = E\{g(X) | Z\}.$$

- 非参数工具变量识别 (Wang & Tchetgen, 2018)

- 关于原因变量的模型里我们要求没有  $U$  和  $Z$  的交互项:

$$E(X | Z = 1, U) - E(X | Z = 0, U) = E(X | Z = 1) - E(X | Z = 0).$$

- 关于结果变量的模型里我们要求没有  $U$  和  $X$  的交互项:

$$E(Y_1 - Y_0 | U) = E(Y_1 - Y_0).$$

# Outline

## ① 工具变量

同质性假定-无协变量

同质性假定-有协变量

同质性假定-其他

同质性假定，无效 IV 的情形

单调性假设

## ② 阴性对照

## ③ 双重差分

## 同质性假定，无效 IV 的情形 I

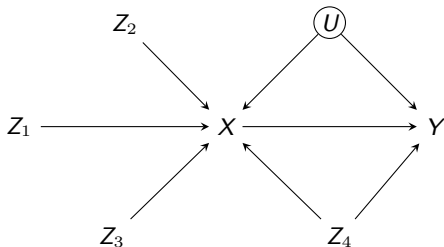


图:  $Z_1, Z_2, Z_3$  为有效工具变量,  $Z_4$  为无效工具变量

- 在实际情况中, 工具变量可以进行随机化试验, 从而  $(Z_1, \dots, Z_p) \perp\!\!\!\perp U$ , 但仍难以避免  $Z_j$  对  $Y$  有直接作用。
- 实际中, 我们并不知道哪些工具变量为有效工具变量。
- Kang et al. (2016) 提出大多数原则 (majority rule): 当有效工具变量超过一半仍可识别因果作用。

## 同质性假定, 无效 IV 的情形 II

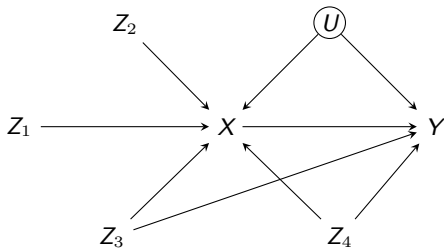


图:  $Z_1, Z_2$  为有效工具变量,  $Z_3, Z_4$  为无效工具变量

- 实际中, 我们并不知道哪些工具变量为有效工具变量。
- Guo et al. (2018) 提出众数原则 (plurality rule): 当有效工具变量的估计量为众数时, 因果作用也是可识别的。

# Outline

## ① 工具变量

同质性假定-无协变量

同质性假定-有协变量

同质性假定-其他

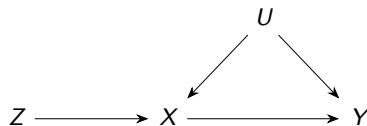
同质性假定，无效 IV 的情形

单调性假设

## ② 阴性对照

## ③ 双重差分

# 单调性假设-依从组的平均因果作用 (LATE) I



$X_1$	$X_0$	$G$	Description
1	1	AT	Always-taker
1	0	CO	Complier
0	1	DE	Defier
0	0	NT	Never-taker

- 例子 令  $Z = 1$  表示成长于大学附近,  $Z = 0$  表示没有成长于大学附近;  $X = 1$  表示完成了高中学业,  $S = 0$  表示没有. 令  $Y$  表示对数收入 (Card, 1993).
- 单调性是指:  $X_1 \geq X_0$ .
- 以上三个假定 (i)-(iii) 下, 我们得到:

# 单调性假设-依从组的平均因果作用 (LATE) I

证明.

$$\begin{aligned} & E(Y_{X=1}) - E(Y_{X=0}) \\ &= \text{pr}(X_{Z=1} = 1, X_{Z=0} = 0) E(Y_{Z=1} - Y_{Z=0} \mid X_{Z=1} = 1, X_{Z=0} = 0) \\ &\quad + \text{pr}(X_{Z=1} = 0, X_{Z=0} = 0) E(Y_{Z=1} - Y_{Z=0} \mid X_{Z=1} = 0, X_{Z=0} = 0) \\ &\quad + \text{pr}(X_{Z=1} = 1, X_{Z=0} = 1) E(Y_{Z=1} - Y_{Z=0} \mid X_{Z=1} = 1, X_{Z=0} = 1) \\ &= \text{pr}(X_{Z=1} = 1, X_{Z=0} = 0) E(Y_{Z=1} - Y_{Z=0} \mid X_{Z=1} = 1, X_{Z=0} = 0) \\ &\quad + \text{pr}(X_{Z=1} = 0, X_{Z=0} = 0) E(Y_{X=0} - Y_{X=0} \mid X_{Z=1} = 0, X_{Z=0} = 0) \\ &\quad + \text{pr}(X_{Z=1} = 1, X_{Z=0} = 1) E(Y_{X=1} - Y_{X=1} \mid X_{Z=1} = 1, X_{Z=0} = 1) \\ &= \text{pr}(X_{Z=1} = 1, X_{Z=0} = 0) E(Y_{Z=1} - Y_{Z=0} \mid X_{Z=1} = 1, X_{Z=0} = 0) \\ &= \text{pr}(X_{Z=1} = 1, X_{Z=0} = 0) E(Y_{X=1} - Y_{X=0} \mid X_{Z=1} = 1, X_{Z=0} = 0) \end{aligned}$$



## 单调性假设-依从组的平均因果作用 (LATE) II

- 单调性假定 (monotonicity) 使得  $X$  的潜在结果的组合只有三种;
- 排斥限制 (exclusion restriction) 使得上面分解的后两个式子为 0 .
- 接下来考虑  $\text{pr}(X_{Z=1} = 1, X_{Z=0} = 0)$  的识别性:

$X_1$	$X_0$	$G$	Description
1	1	AT	Always-taker
1	0	CO	Complier
0	1	DE	Defier
0	0	NT	Never-taker



## 单调性假设-依从组的平均因果作用 (LATE) III

$$\begin{aligned}\text{pr}(X = 1 \mid Z = 1) &= \text{pr}(X_{Z=1} = 1) \\ &= \text{pr}(X_{Z=1} = 1, X_{Z=0} = 1) + \text{pr}(X_{Z=1} = 1, X_{Z=0} = 0) \\ &= \text{pr}(G = ss) + \text{pr}(G = s\bar{s}), \\ \text{pr}(X = 1 \mid Z = 0) &= \text{pr}(X_{Z=0} = 1) \\ &= \text{pr}(X_{Z=0} = 1, X_{Z=1} = 1) + \text{pr}(X_{Z=0} = 1, X_{Z=1} = 0) \\ &= \text{pr}(G = ss) + 0, \\ \text{pr}(X = 0 \mid Z = 1) &= \text{pr}(X_{Z=1} = 0) \\ &= \text{pr}(X_{Z=0} = 1, X_{Z=1} = 0) + \text{pr}(X_{Z=0} = 0, X_{Z=1} = 0) \\ &= 0 + \text{pr}(G = \bar{s}\bar{s}).\end{aligned}$$

## 单调性假设-依从组的平均因果作用 (LATE) IV

- $(Z_1 = 1, Z_0 = 0)$  一类人的比例就是  $Z$  对  $X$  平均因果作用:

$$\begin{aligned}\text{ATE}(Z \rightarrow X) &= \text{pr}(X_{Z=1} = 1, X_{Z=0} = 0). \\ &= \text{pr}(X = 1 \mid Z = 1) - \text{pr}(X = 1 \mid Z = 0).\end{aligned}$$

- 因此,

$$\begin{aligned}\text{LATE} &= E(Y_{Z=1} - Y_{Z=0} \mid Z_1 = 1, Z_0 = 0) \\ &= E(Y_{X=1} - Y_{X=0} \mid Z_1 = 1, Z_0 = 0) \\ &= \frac{\text{ATE}(Z \rightarrow Y)}{\text{ATE}(Z \rightarrow X)}.\end{aligned}$$

- 上面的式子被定义为 LATE 是有理由的. 它表示的是子总体  $(Z_1 = 1, Z_0 = 0)$  中, 随机化对于结果的因果作用; 由于这类人中随机化和接受的处理是相同的, 它也表示处理对结果的因果作用.

## 单调性假设-依从组的平均因果作用 (LATE) V

- 这类人接受处理与否完全由于是否接受鼓励而定, 他们被成为“依从者”(complier), 因为这类人群中的平均因果作用又被成为“依从者平均因果作用”(CACE: complier average causal effect); 计量经济学家称它为“局部处理作用”(LATE: local average treatment effect) .
- 由于  $Z$  是随机化的, 它对于  $X$  和  $Y$  的平均因果作用都是显而易见可以得到的. 因为  $\widehat{ATE}(Z \rightarrow X) = \bar{X}_1 - \bar{X}_0$ ,  $\widehat{ATE}(Z \rightarrow Y) = \bar{Y}_1 - \bar{Y}_0$ , CACE 的一个矩估计便是

$$\hat{\beta}_{IV} = \frac{\widehat{ATE}(Z \rightarrow Y)}{\widehat{ATE}(Z \rightarrow X)} = \frac{\hat{\sigma}_{yz}}{\hat{\sigma}_{xz}}.$$

- 由此可见工具变量估计量的因果含义. 上面的讨论既显示了工具变量对于识别因果作用的有效性.
- 局限性: 我们只能识别某个子总体的平均因果作用; 而通常情况下, 我们并不知道某个个体具体属于哪个子总体.

## 工具变量的缺点

- 工具变量估计对假定 (i)-(iii) 很敏感.
  - 假定 (i) 需要有专业知识保证工具变量对结果变量没有直接作用.
  - 条件 (ii) 难以验证, 因为混杂因素  $U$  没有观测到.
  - 条件 (iii) 可以用观测数据检验
- 用  $\text{plim } \beta_1^{\text{iv}}$  表示当样本量趋于无穷时  $\beta_1^{\text{iv}}$  依概率收敛的极限. 当条件 (ii) 不满足时, 工具变量估计的渐近偏差是

$$\text{plim } \beta_1^{\text{iv}} - \beta_1 = \frac{\sigma_{uz}}{\sigma_{xz}}.$$

- 由此可见, 当  $\sigma_{uz} \neq 0$  时, 工具变量估计不相合, 而且偏差会由于  $\sigma_{xz}$  过小而被放大很多倍. 使用阴性对照变量 (negative control variable) 方法解决这些问题.

## 阴性对照 I

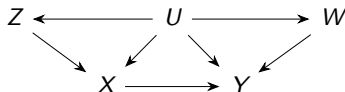
- 阴性对照变量是与混杂因素  $U$  相关, 但与处理  $X$  或结果变量  $Y$  无因果关系的辅助变量.
- 阴性对照变量分为两种: 阴性对照暴露和阴性对照结果.
- 前者是一个辅助的暴露变量, 但是对关心的结果没有直接的因果作用;
- 后者是一个辅助的结果变量, 但是不受暴露变量的影响.
- 这些特点可以严格地表述如下.
  - (阴性对照暴露, negative control exposure) 一个暴露变量  $Z$  称为一个阴性对照暴露, 如果它满足  $Z \perp\!\!\!\perp Y \mid (U, X)$  和  $Z \perp\!\!\!\perp W \mid (U, X)$ .
  - (阴性对照结果, negative control outcome) 一个结果变量  $W$  称为一个阴性对照结果, 如果它满足  $W \perp\!\!\!\perp X \mid U$  和  $W \not\perp\!\!\!\perp U$ .

## 阴性对照 I

- 阴性对照框架需要研究者将收集到的协变量划分为三类：
  - ▶ 可观测混杂：同时影响处理变量和结果变量的协变量  $C$ ;
  - ▶ 阴性对照处理：仅与处理变量和未观测混杂相关的协变量  $Z$ ;
  - ▶ 阴性对照结果：仅与结果变量和未观测混杂相关的协变量  $W$ .
- 独立性条件：
  - 阴性对照暴露变量:  $Z \perp\!\!\!\perp Y \mid (X, U)$ ,  $Z \perp\!\!\!\perp W \mid (U, X)$
  - 阴性对照结果变量:  $W \perp\!\!\!\perp X \mid U$ ,  $W \not\perp\!\!\!\perp U$

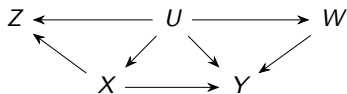
(NCE)

(NCO)



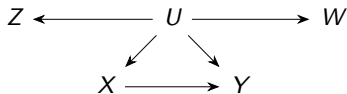
## 阴性对照 I

(NCE) (NCO)



- 测量误差: Kuroki & Pearl (2014)

(NCE) (NCO)



## 阴性对照 II

- 除了要求阴性对照变量  $Z$  和  $W$  与处理或结果变量无直接的因果关系, 上面的定义还要求  $Z$  与  $(W, Y)$  之间的混杂因素和  $X$  与  $(W, Y)$  之间的混杂因素相同. 当存在完全观测的协变量  $V$  时, 上述定义中的条件独立性需要给定  $V$ .
- 阴性对照暴露的定义类似工具变量中的无直接作用条件, 但是对  $(Z, U)$  的相关性不做要求, 因此, 工具变量可看作阴性对照暴露的特例.



## 阴性对照识别性 I

### 证明.

- 在潜在可忽略性假设 (latent ignorability) 下:

$$P(Y_x = y) = \sum_u P(Y = y \mid U = u, x) P(U = u).$$

- 我们首先考虑  $V, W$  和  $Z$  都是  $k$  个水平的离散变量。引入以下记号:

$$P(W \mid u) = \{P(w_1 \mid u), \dots, P(w_k \mid u)\}^\top$$

$$P(w \mid U) = \{P(w \mid u_1), \dots, P(w \mid u_k)\}$$

$$P(W \mid U) = \{P(W \mid u_1), \dots, P(W \mid u_k)\}$$



## 阴性对照识别性 II

### 证明.

- 类似地, 我们定义

$$P(U | v, x) = \{P(u_1 | v, x), \dots, P(u_k | v, x)\}^T$$

$$P(U | V, x) = \{P(U | v_1, x), \dots, P(U | v_k, x)\}$$

$$P(y | V, x) = \{P(y | v_1, x), \dots, P(y | v_k, x)\}$$

- 由于  $W \perp\!\!\!\perp (V, X) | U$  及  $V \perp\!\!\!\perp Y | (U, X)$ ,

$$P(W | V, x) = P(W | U)P(U | V, x)$$

$$P(y | V, x) = P(y | U, x)P(U | V, x)$$



## 阴性对照识别性 III

### 证明.

- 当矩阵  $P(W | V, x)$  可逆时, 我们有

$$P(U | V, x) = P(W | U)^{-1} P(W | V, x),$$

$$P(y | V, x) = P(y | U, x) P(W | U)^{-1} P(W | V, x)$$

$$P(y | U, x) = P(y | V, x) P(W | V, x)^{-1} P(W | U)$$

- 最后,

$$P(Y_x = y) = P(y | U, x) P(U)$$

$$= P(y | V, x) P(W | V, x)^{-1} P(W | U) P(U)$$

$$= P(y | V, x) P(W | V, x)^{-1} P(W) \quad (\text{Identifiable!})$$



### 例 1 (Card & Krueger, 1994)

目标：评估新泽西州最低工资变化前后，新泽西州快餐店的平均就业人数。

- 在 1992 年 4 月 1 日，新泽西州将州最低工资从 4.15 美元每小时提高到 5.05 美元每小时
- 在同一时期，宾夕法尼亚的最低工资一直维持在 4.25 美元每小时。
- 在 1992 年 2 月份，宾夕法尼亚州快餐店就业人数高于新泽西州，但是在 11 月份却下降了。

## 双重差分 II

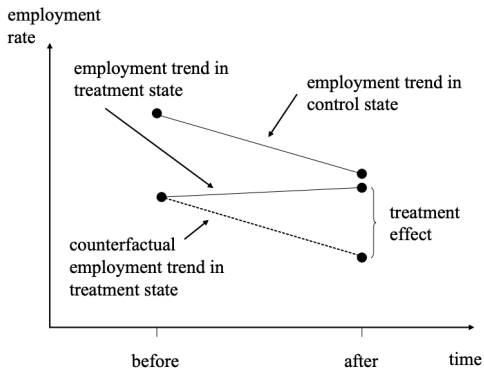


Figure 5.2.1: Causal effects in the differences-in-differences model

提高最低工资使新泽西州的就业趋势偏离了原有的趋势。

## 双重差分 III

- 双重差分法适用于面板数据或重复观测的横截面数据：两个或更多组别
- 每个组别里的个体有至少有两个观测时间点；在某些组别的某些时间点，个体接受处理
- 我们以两个组别和两个时间点为例介绍双重差分法。
- $G = 1$  表示处理组， $G = 0$  表示对照组
- 协变量  $X$
- 用  $T = t, T = t + 1$  分别表示两个时刻，我们假设  $T = t + 1$  时刻施加处理
- 定义处理状态  $Z = G * I(T = t + 1)$ ，也即仅在处理组、 $t + 1$  时刻， $Z = 1$ ； $T$  时刻的潜在结果  $Y_T(z)$ ，其中  $z \in \{0, 1\}$
- 两个时刻的结果变量分别记为  $Y_t, Y_{t+1}$ ；

## 双重差分 IV

	处理前	处理后
对照组 $G = 0$	$Y_t(0)$	$Y_{t+1}(0)$
处理组 $G = 1$	$Y_t(0)$	$Y_{t+1}(1)$

- 感兴趣的因果参数:  $\tau_{ATT} = E\{Y_{t+1}(1) - Y_{t+1}(0) \mid G = 1\} \triangleq \mu_1 - \mu_0$ , 其中
  - $\mu_1 = E\{Y_{t+1}(1) \mid G = 1\} = E(Y_{t+1} \mid G = 1)$ .
  - $\mu_0 = E\{Y_{t+1}(0) \mid G = 1\} = E(Y_{t+1}(0) \mid G = 1)$ . (待识别)
- 关键识别假定, **Parallel trends**:

$$\begin{aligned} & - E\{Y_{t+1}(0) \mid G = 0\} \\ & = E\{Y_t(0) \mid G = 1\} - E\{Y_t(1) \mid G = 0\}. \end{aligned}$$

- 在并行趋势下，DID 估计是非参数识别的：

$$\tilde{\mu}_{0,\text{DID}} = \mathbb{E}(Y_{it} \mid G_i = 1) + \mathbb{E}(Y_{i,t+1} \mid G_i = 0) - \mathbb{E}(Y_{it} \mid G_i = 0)$$

DID estimator takes the difference between the change in outcomes for treated individuals and the change for control individuals - Equivalently, DID estimator can be written as the difference between the before-after estimators of treatment and of control groups

$$\begin{aligned}\hat{\tau}_{\text{DID}} &= (\bar{Y}_{1,t+1} - \bar{Y}_{1,t}) - (\bar{Y}_{0,t+1} - \bar{Y}_{0,t}) \\ &= \hat{\tau}_1^{BA} - \hat{\tau}_0^{BA}\end{aligned}$$



## 双重差分 VI

where  $\bar{Y}_{g,t}$  are mean outcomes for group  $g$  at time  $t$  DID works? Assuming SUTVA, we have

$$Y_t = Y_t(0), \quad Y_{t+1} = (1 - G)Y_{t+1}(0) + GY_{t+1}(1),$$

Therefore

$$\begin{aligned}\mathbb{E}(\hat{\tau}_0^{BA}) &= \mathbb{E}(\bar{Y}_{0,t+1} - \bar{Y}_{0,t}) \\ &= \mathbb{E}[Y_{t+1}(0) - Y_t(0) \mid G = 0] \quad (\text{SUTVA}) \\ &= \mathbb{E}[Y_{t+1}(0) - Y_t(0) \mid G = 1] \quad (\text{par. trends}) \\ &= \mathbb{E}[Y_{t+1}(0) \mid G = 1] - \mathbb{E}[Y_t \mid G = 1] \quad (\text{SUTVA})\end{aligned}$$

Similarly, we have

$$\mathbb{E} \left( \hat{\tau}_1^{BA} \right) = \mathbb{E} [Y_{t+1}(1) \mid G = 1] - \mathbb{E} [Y_t \mid G = 1]$$

Subtracting the two, we have

$$\mathbb{E} (\hat{\tau}_{\text{DID}}) = \mathbb{E} \left( \hat{\tau}_1^{BA} - \hat{\tau}_0^{BA} \right) = \mathbb{E} [Y_{t+1}(1) - Y_{t+1}(0) \mid G = 1] = \tau_{\text{ATT}}$$