第三章　未观测混杂

罗姗姗

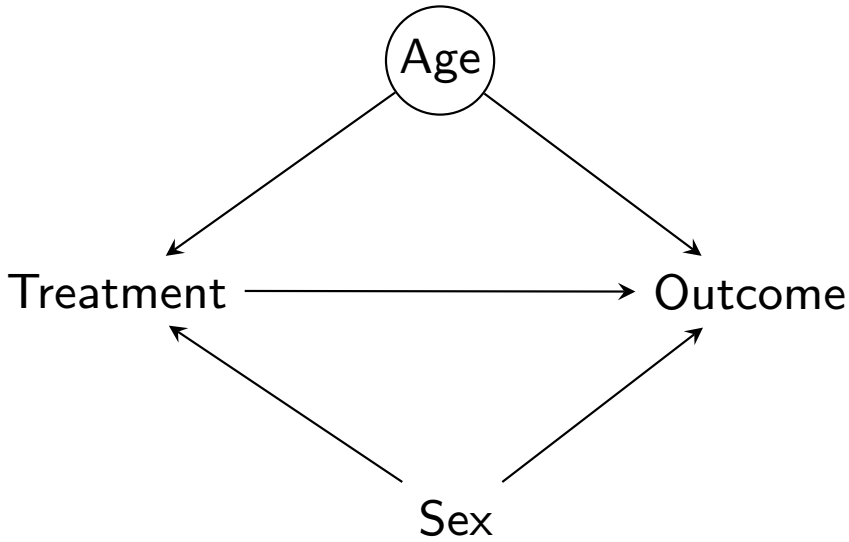
北京工商大学 数学与统计学院

因果推断课题组

目录

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

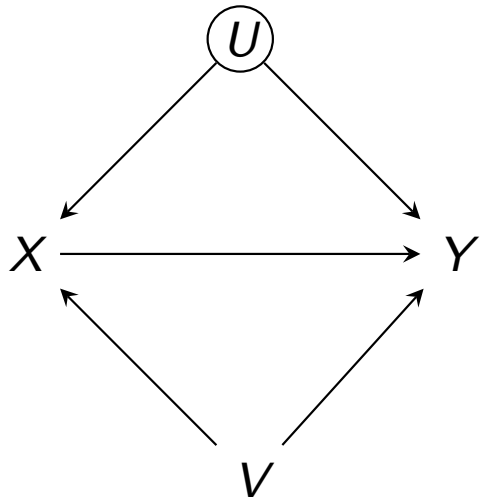
- 当存在未被观测的混杂因素时，更合理的假定是潜在可忽略性:

  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，

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



未观测混杂

- 如果 U 没有被观测，
  - 那么 $\mathbb{E}(Y \mid X = x, U)$ 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 $\mathbb{E}(Y \mid X = x)$ 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
  - 一种是阴性对照变量 (negative control variable) 方法.

Outline

记号 I

- 考虑一个由 $i = 1, \ldots, n$ 个个体组成的实验。
- 令 Z_i 表示分配的处理，1 代表处理组，0 代表对照组。
- 令 D_i 表示接受的处理，1 代表处理组，0 代表对照组。
- 当对于某些单位 i，$Z_i \neq D_i$ 时，会出现不依从问题。
- 不依从问题在鼓励设计中尤为常见。这是由于，我们无法强制参与者接受处理，而只能鼓励他们接受处理。

记号 II

- 令 Y_i 表示感兴趣的结果。考虑完全随机化的 Z_i，现在忽略协变量 X_i。
- 我们有处理接受的潜在值 $\{D_i(1), D_i(0)\}$，以及结果的潜在值 $\{Y_i(1), Y_i(0)\}$ 。
- 观察值为 $D_i = Z_i D_i(1) + (1 - Z_i) D_i(0)$ 和 $Y_i = Z_i Y_i(1) + (1 - Z_i) Y_i(0)$。
- 为了简化表示，我们假设 $\{Z_i, D_i(1), D_i(0), Y_i(1), Y_i(0)\}_{i=1}^n$ 是独立同分布的，有时会省略下标 i 而不引起混淆。

意向处理效应 I

- 我们首先考虑完全随机化实验，即 $Z \perp\!\!\!\perp \{D(1), D(0), Y(1), Y(0)\}$。
- 因为采用随机化处理分配，所以分配与结果的相关性度量可以表示因果效应。随机化允许我们识别关于 $Z \to D$ 和 $Z \to Y$ 的平均因果效应：

$$\tau_D = E\{D(1) - D(0)\} = E(D \mid Z = 1) - E(D \mid Z = 0)$$
$$\tau_Y = E\{Y(1) - Y(0)\} = E(Y \mid Z = 1) - E(Y \mid Z = 0).$$

- 分配 $Z = 1$ 的意向是使得患者接受处理 $D = 1$，分配 $Z = 0$ 的意向是使得患者接受对照 $D = 0$，因此称处理分配对结果的因果效应被称为意向处理效应（intention to treat; ITT）。
- 我们可以使用简单的均值差估计量 $\hat{\tau}_D$ 和 $\hat{\tau}_Y$ 来估计 τ_D 和 τ_Y。
- 在随机化试验下，ITT 可以评估分配机制对结果的影响，即 Z 对 Y 的影响。然而，它并不能完全地回答我们关心的问题，即实际接受的处理 D 对结果 Y 的因果效应。

依从组因果作用 I

D_1	D_0	G	Description
1	1	AT	Always-taker
1	0	CO	Complier
0	1	DE	Defier
0	0	NT	Never-taker

- 在文献 Angrist et al. (1996) 的基础上，我们根据 $\{D_i(1), D_i(0)\}$ 的联合潜在取值来对人群进行分层。因为 D 是二值的，我们有四种可能的组合：
- 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；$D = 1$ 表示完成了高中学业，$D = 0$ 表示没有. 令 Y 表示对数收入 (Card, 1993).

依从组因果作用 II

- 由全概率公式，我们可以

$$\begin{aligned}\tau_Y =& E\{Y(1) - Y(0) \mid U = \mathrm{a}\}\,\mathrm{pr}(U = \mathrm{a}) \\ &+ E\{Y(1) - Y(0) \mid U = \mathrm{c}\}\,\mathrm{pr}(U = \mathrm{c}) \\ &+ E\{Y(1) - Y(0) \mid U = \mathrm{d}\}\,\mathrm{pr}(U = \mathrm{d}) \\ &+ E\{Y(1) - Y(0) \mid U = \mathrm{n}\}\,\mathrm{pr}(U = \mathrm{n}).\end{aligned}$$

因此，τ_Y 是四个潜在子群效应的加权平均。我们将在下面更详细地探讨这些潜在群体。

依从组因果作用 III

Assumption 1 (单调性, monotonicity)

$\mathrm{pr}(U = \mathrm{d}) = 0$ 或 $D_i(1) \geq D_i(0)$，即不存在 *defiers*。

- 当对照组无法接触到处理时，即 $D_i(0) = 0$ 对所有个体都成立时，上述单调性假设会自然成立。在随机化及单调性假设下，我们会有：

$$\mathrm{pr}(D = 1 \mid Z = 1) \geq \mathrm{pr}(D = 1 \mid Z = 0).$$

- 上式可用于从数据中检验单调性，即单调性假设可被数据证伪。
- 但是，单调性假设比上述不等式要强很多。前者在个体水平上限制了 $D_i(1)$ 和 $D_i(0)$，而后者只在平均水平上限制了它们。
- 尽管如此，当这个可检验的不等式成立时，我们无法基于观察数据来反驳单调性假设。

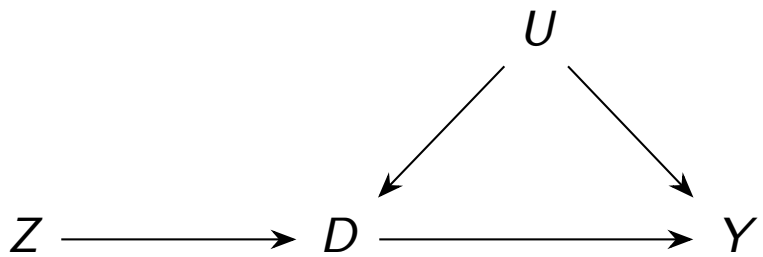
依从组因果作用 IV

Assumption 2 (排除限制, exclusion restriction)

对于 *always-takers* $(U_i = a)$ 和 *never-takers* $(U_i = \mathrm{n})$，有 $Y_i(1) = Y_i(0)$。

- ER 假设要求处理分配只在它对实际接受的处理有影响时，才会对结果有影响。
- 从生物学角度来看，在许多双盲临床试验中 ER 假设是合理的，因为结果仅依赖于实际接受的处理。
- 也就是说，如果处理分配不影响实际接受的处理，那么它也不会影响结果。
- 这一假设可能会在处理分配对结果产生直接影响而不是通过接受的处理时而被违反。例如，某些随机对照试验不是双盲的，处理分配可能会对结果产生一些未知的途径。

依从组因果作用 V



D_1	D_0	U	Description
1	1	AT	Always-taker
1	0	CO	Complier
0	1	DE	Defier
0	0	NT	Never-taker

依从组因果作用 VI

- 在单调性和 ER 假设成立的条件下，τ_Y 的分解只有第二项：

$$\begin{aligned}\tau_Y =& E\{Y(1) - Y(0) \mid U = \mathrm{a}\}\,\mathrm{pr}(U = \mathrm{a}) \\ &+ E\{Y(1) - Y(0) \mid U = \mathrm{c}\}\,\mathrm{pr}(U = \mathrm{c}) \\ &+ E\{Y(1) - Y(0) \mid U = \mathrm{d}\}\,\mathrm{pr}(U = \mathrm{d}) \\ &+ E\{Y(1) - Y(0) \mid U = \mathrm{n}\}\,\mathrm{pr}(U = \mathrm{n}) \\ =& E\{Y(1) - Y(0) \mid U = \mathrm{c}\}\,\mathrm{pr}(U = \mathrm{c}).\end{aligned}$$

依从组因果作用 VII

同样，我们可以将 D 的平均因果效应分解为四个项：

$$\begin{aligned}\tau_D =& E\{D(1) - D(0) \mid U = \mathrm{a}\} \operatorname{pr}(U = \mathrm{a}) \\ &+ E\{D(1) - D(0) \mid U = \mathrm{c}\} \operatorname{pr}(U = \mathrm{c}) \\ &+ E\{D(1) - D(0) \mid U = \mathrm{d}\} \operatorname{pr}(U = \mathrm{d}) \\ &+ E\{D(1) - D(0) \mid U = \mathrm{n}\} \operatorname{pr}(U = \mathrm{n}) \\ =& 0 \times \operatorname{pr}(U = \mathrm{a}) + 1 \times \operatorname{pr}(U = \mathrm{c}) + (-1) \times \operatorname{pr}(U = \mathrm{d}) + 0 \times \operatorname{pr}(U = \mathrm{n}) \\ =& \operatorname{pr}(U = \mathrm{c}).\end{aligned}$$

CACE 与 LATE I

- 我们发现一个有趣的事情，在完全随机化条件下，依从者的比例 π_c 是可识别的，而且等于处理分配对 D 的平均因果效应！
- 尽管我们无法基于观察数据找到所有的依从者，但我们可以根据 τ_D 的识别表达式来确定他们在整个人群中的比例。

定理 1

在单调性和 *ER* 假设下，我们有

$$E\{Y(1) - Y(0) \mid U = c\} = \frac{\tau_Y}{\tau_D} = \frac{E(Y \mid Z = 1) - E(Y \mid Z = 0)}{E(D \mid Z = 1) - E(D \mid Z = 0)}.$$

如果 $\tau_D \neq 0$。

CACE 与 LATE II

- 上述因果参数 $E\{Y(1) - Y(0) \mid U = \mathrm{c}\}$ 表示的是依从组上的实际接受的处理 D 对结果变量 Y 的影响，所以常被称为 Complier Average Causal Effect（CACE）。
- 但上述因果参数也反映了一个局部子人群的因果作用，故又称为 Local Average Treatment Effect（LATE）。也可等价地表示如下：

$$\begin{aligned} \tau_{\mathrm{c}} &= E\{Y(1) - Y(0) \mid D(1) = 1, D(0) = 0\} \\ &= E\{Y(1) - Y(0) \mid D(1) > D(0)\}. \end{aligned} \tag{1}$$

- 此外，上述定理表明 CACE 或 LATE 等于 $Z \to Y$ 的平均因果效应与 $D \to Y$ 的平均因果效应之比。

估计 I

- 根据上述结论，我们可以通过简单的比值来估计 τ_c：

$$\hat{\tau}_c = \frac{\hat{\tau}_Y}{\hat{\tau}_D}$$

这被称为 Wald 估计量或 IV 估计量。

- 我们可以获得方差估计如下（参见示例 A1.3）：

$$\hat{\tau}_c - \tau_c = \frac{(\hat{\tau}_Y - \tau_c \hat{\tau}_D)}{\hat{\tau}_D} \approx \frac{(\hat{\tau}_Y - \tau_c \hat{\tau}_D)}{\tau_D} = \frac{\hat{\tau}_A}{\tau_D}.$$

其中 $\hat{\tau}_A$ 是 pseudo-outcome $A_i = Y_i - \tau_c D_i$ 的均值之差。

估计 II

因此，$\hat{\tau}_c$ 的渐近方差接近于 $\hat{\tau}_A$ 的方差除以 τ_D^2。具体地方差估计分为以下步骤：

1. 获得已调整结果 $\hat{A}_i = Y_i - \hat{\tau}_c D_i (i = 1, \ldots, n)$；
2. 基于已调整结果获得 Neyman 类型方差估计：

$$\hat{V}_{\hat{A}} = \frac{\hat{S}^2_{\hat{A}}(1)}{n_1} + \frac{\hat{S}^2_{\hat{A}}(0)}{n_0},$$

其中 $\hat{S}^2_{\hat{A}}(1)$ 和 $\hat{S}^2_{\hat{A}}(0)$ 分别是处理和对照组下的 $\hat{A}_i$ 的样本方差；

3. 获得最终的方差估计 $\hat{V}_{\hat{A}}/\hat{\tau}_D^2$。

估计 III

- 在零假设下，即 $\tau_c = 0$，我们可以简单地用 $\hat{V}_Y/\hat{\tau}_D^2$ 来近似方差，其中 $\hat{V}_Y$ 是 Y 的均值差异的 Neyman 类型方差估计。
- 如果真实的 τ_c 不为零，这种方差估计量并不相合。因此，它适用于简单测试而并非准确估计。
- 然而，它为 ITT 估计量和 Wald 估计量提供了有趣的解释视角。

估计 IV

- ITT 估计量 $\hat{\tau}_Y$ 具有估计的标准误差 $\sqrt{\hat{V}_Y}$。Wald 估计量 $\hat{\tau}_Y/\hat{\tau}_D$ 本质上等于 ITT 估计量乘以 $1/\hat{\tau}_D > 1$，它在数量上更大，但同时它的估计标准误差也以相同的因子增加。τ_Y 和 τ_c 的置信区间如下：

$$\hat{\tau}_Y \pm z_{1-\alpha/2}\sqrt{\hat{V}_Y}$$

$$\frac{\hat{\tau}_Y}{\hat{\tau}_D} \pm z_{1-\alpha/2}\frac{\sqrt{\hat{V}_Y}}{\hat{\tau}_D} = \frac{\hat{\tau}_Y \pm z_{1-\alpha/2}\sqrt{\hat{V}_Y}}{\hat{\tau}_D}.$$

- 这些置信区间提供了相同的定性结论，因为它们都将包括零或不包括零。从某种意义上说，工具变量分析提供了与 Y 的意向处理效应 (ITT) 分析相同的定性信息，尽管它涉及更复杂的估计过程。

Outline

Ordinary Least Squares I

- 在讨论工具变量的经济计量视角之前，我们将首先回顾统计学中的普通最小二乘法 (OLS)。
- OLS 是统计学中的经典话题，但它有不同的数学表达及解释。
- 第一种观点是基于投影的。给定任意一对具有有限二阶矩的随机变量 (D, Y)，定义总体 OLS 系数为

$$\beta = \arg\min_b E\left(Y - D^{\mathrm{T}} b\right)^2 = E\left(DD^{\mathrm{T}}\right)^{-1} E(DY),$$

然后定义残差为 $\varepsilon = Y - D^{\mathrm{T}}\beta$。根据定义，$Y$ 可以分解为

$$Y = D^{\mathrm{T}}\beta + \varepsilon$$

Ordinary Least Squares II

这必须满足

$$E(D\varepsilon) = 0.$$

- 基于 $(D_i, Y_i)_{i=1}^n \overset{\text{IID}}{\sim} (D, Y)$，OLS 估计量 β 为

$$\hat{\beta} = \left(\sum_{i=1}^n D_i D_i^{\mathrm{T}} \right)^{-1} \sum_{i=1}^n D_i Y_i$$

- 由于

$$\hat{\beta} = \left(\sum_{i=1}^n D_i D_i^{\mathrm{T}} \right)^{-1} \sum_{i=1}^n D_i \left(D_i^{\mathrm{T}} \beta + \varepsilon_i \right) = \beta + \left(\sum_{i=1}^n D_i D_i^{\mathrm{T}} \right)^{-1} \sum_{i=1}^n D_i \varepsilon_i,$$

我们可以证明由于 $E(\varepsilon D) = 0$，$\hat{\beta}$ 是一致估计 β 的。

经典 EHW 方差估计量 I

- 经典的 EHW (Eicker-Huber-White Robust Variance Estimator) 方差估计量用于 $\operatorname{cov}(\hat{\beta})$，表示如下：

$$\hat{V}_{\mathrm{EHW}} = \left(\sum_{i=1}^{n} D_i D_i^{\mathrm{T}}\right)^{-1} \left(\sum_{i=1}^{n} \hat{\varepsilon}_i^2 D_i D_i^{\mathrm{T}}\right) \left(\sum_{i=1}^{n} D_i D_i^{\mathrm{T}}\right)^{-1}$$

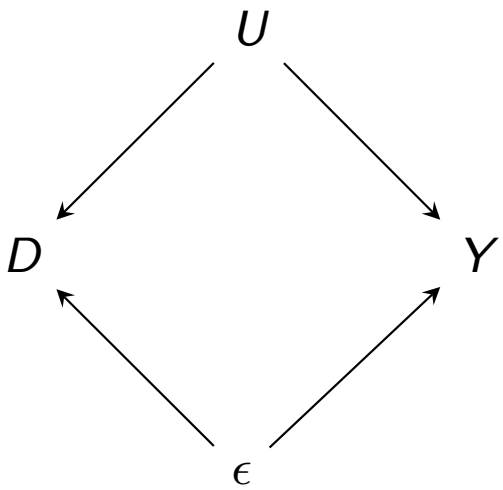
其中 $\hat{\varepsilon}_i = Y_i - D_i^{\mathrm{T}} \hat{\beta}$ 是残差。

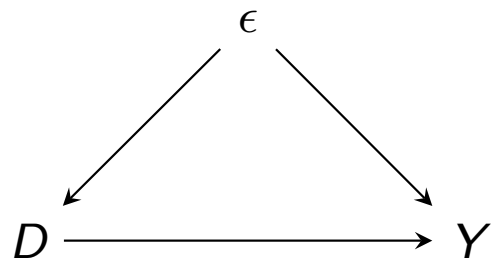
- 第二种观点是将

$$Y = D^{\mathrm{T}} \beta + \varepsilon \tag{2}$$

视为数据生成过程的真实模型。也就是说，给定随机变量 (D, ε)，我们基于线性方程(2)生成 Y。重要的是，在数据生成过程中，ε 和 D 可能相关，即 $E(D\varepsilon) \neq 0$。

经典 EHW 方差估计量 II





- 这与第一种观点不同，其中根据总体 OLS 的定义，$E(\varepsilon D) = 0$。因此，OLS 估计量可能并不相合：

$$\hat{\beta} \to \beta + E\left(DD^{\mathrm{T}}\right)^{-1} E(D\varepsilon) \neq \beta$$

内生和外生回归变量 I

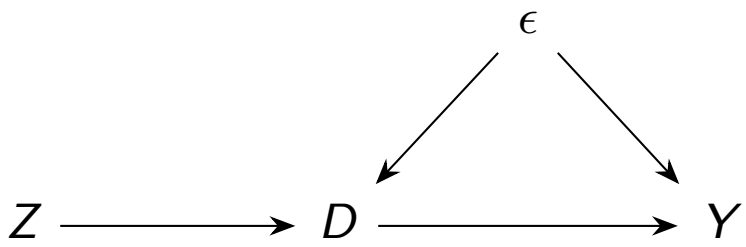
- 我们根据(2)的定义，对内生和外生回归变量进行如下定义：

定义 1

当 $E(\varepsilon D) \neq 0$ 时，回归变量 D 被称为内生的；当 $E(\varepsilon D) = 0$ 时，回归变量 D 被称为外生的。

- 计量经济学常采用定义 1 中的术语。当 $E(\varepsilon D) \neq 0$ 时，我们也常说 D 有内生性；当 $E(\varepsilon D) = 0$ 时，我们也常说 D 有外生性。
- 在 OLS 的第一种视角中，内生性和外生性的概念并不起作用，因为根据 ϵ 的定义，总会有 $E(\varepsilon D) = 0$。
- 因此，我们采用第二种视角理解线性模型下的工具变量。

线性工具变量模型 I



- 当 D 是内生的时，OLS 估计量不一致。我们必须使用额外的信息来构建 β 的相合估计量。我们将重点关注以下线性工具变量模型：

$$Y = D^{\mathrm{T}}\beta + \varepsilon, \tag{3}$$

其中 ε 和 Z 满足

$$E(\varepsilon Z) = 0. \tag{4}$$

模型(3)允许 $E(\varepsilon D) \neq 0$，但需要一个替代的矩条件(4)。通过加入截距项，考虑到 $E(\varepsilon) = 0$，新的条件表明 Z 与误差项 ε 不相关。

线性工具变量模型 II

- 但是，任何随机生成的噪声都与 ε 不相关，因此 Z 必须满足额外的条件，以确保 Z 对于估计 β 是有用的。直观地说，额外的条件要求 Z 与 D 相关。

- 模型(3)和(4)看似简单。然而，在实证研究中，找到满足上述条件的变量比较困难。由于上述条件涉及未观测变量 ε，通常我们无法从数据中检验这个条件。

恰好识别的情形 I

- 首先考虑 Z 和 D 维数相同且 $E(ZD^{\mathrm{T}})$ 满秩的情况。当 $E(ZD^{\mathrm{T}})$ 不是退化的，条件 $E(\varepsilon Z)=0$ 意味着

$$\begin{aligned} E\left\{Z\left(Y-D^{\mathrm{T}}\beta\right)\right\}=0 &\Longrightarrow E(ZY)=E\left(ZD^{\mathrm{T}}\right)\beta \\ &\Longrightarrow \beta=E\left(ZD^{\mathrm{T}}\right)^{-1}E(ZY) \end{aligned} \tag{5}$$

上述 β 的样本形式可以表示为：

$$\hat{\beta}_{\mathrm{IV}}=\left(\sum_{i=1}^{n}Z_iD_i^{\mathrm{T}}\right)^{-1}\sum_{i=1}^{n}Z_iY_i. \tag{6}$$

- 此外，OLS 估计是工具变量的一种特殊情况，即如果 $E(\varepsilon D)=0$，D 本身可以视为充当了自身的工具变量。

恰好识别的情形 II

- 在存在截距项，以及 D 和 Z 都是一维（标量）的情形下，我们有

$$\begin{cases} Y = \alpha + \beta D + \varepsilon, \\ E(\varepsilon) = 0, \quad \operatorname{cov}(\varepsilon, Z) = 0, \end{cases}$$

这意味着

$$\operatorname{cov}(Z, Y) = \beta \operatorname{cov}(Z, D) \Longrightarrow \beta = \frac{\operatorname{cov}(Z, Y)}{\operatorname{cov}(Z, D)}.$$

- 分子分母同除以 $\operatorname{var}(Z)$，我们得到

$$\beta = \frac{\operatorname{cov}(Z, Y)/\operatorname{var}(Z)}{\operatorname{cov}(Z, D)/\operatorname{var}(Z)},$$

这等于两个 OLS 回归（$Y \sim Z$ 和 $D \sim Z$）的回归系数之比。

恰好识别的情形 III

- 当 Z 是二值的，这些系数是 difference-in-means，进而 β 可以简化为

$$\beta = \frac{E(Y \mid Z = 1) - E(Y \mid Z = 0)}{E(D \mid Z = 1) - E(D \mid Z = 0)}.$$

- 这依从组上的平均因果作用识别公式(1)相同。也就是说，使用二值 IV Z 和二值处理 D，IV 估计量可以恢复依从组上的平均因果作用 (Angrist et al., 1996)。

过度识别的情况

- 上一节侧重讨论恰好识别的情况。当 Z 的维度低于 X 且 $E\left(ZD^{\mathrm{T}}\right)$ 的列不满秩时，以下方程有无穷多个解。

$$E(ZY) = E\left(ZD^{\mathrm{T}}\right)\beta$$

- 上述等式表明此时因果作用不可识别，即使有工具变量 Z，系数 β 也无法唯一确定。在实际情形中，我们需要工具变量的维数不少于内生回归变量的维数。

两阶段最小二乘估计（TSLS）

- 在过度识别的情况下，一种常用的计算技巧是两阶段最小二乘估计（TSLS）：
  1. D 对 Z 进行 OLS 回归，得到拟合值 $\hat{D}_i(i=1,\ldots,n)$。如果 D_i 是向量，则需要逐分量 (component-wise) 运行 OLS 以获得 $\hat{D}_i$。将拟合向量 $\hat{D}_i^{\mathrm{T}}$ 放入矩阵 $\hat{D}$ 的第 i 行;
  2. 运行 OLS，将 Y 对 $\hat{D}$ 进行回归，得到系数 $\hat{\beta}_{\mathrm{TSLS}}$。

TSLS 的有效性 I

- 为了理解 TSLS 的有效性，我们需要引入更多的代数推导。将 TSLS 更详细地写出如下：

$$
\begin{aligned}
\hat{\beta}_{\mathrm{TSLS}} &= \left(\sum_{i=1}^{n} \hat{D}_i \hat{D}_i^{\mathrm{T}}\right)^{-1} \sum_{i=1}^{n} \hat{D}_i Y_i \\
&= \left(\sum_{i=1}^{n} \hat{D}_i \hat{D}_i^{\mathrm{T}}\right)^{-1} \sum_{i=1}^{n} \hat{D}_i \left(D_i^{\mathrm{T}} \beta + \varepsilon_i\right) \\
&= \left(\sum_{i=1}^{n} \hat{D}_i \hat{D}_i^{\mathrm{T}}\right)^{-1} \sum_{i=1}^{n} \hat{D}_i D_i^{\mathrm{T}} \beta + \left(\sum_{i=1}^{n} \hat{D}_i \hat{D}_i^{\mathrm{T}}\right)^{-1} \sum_{i=1}^{n} \hat{D}_i \varepsilon_i .
\end{aligned}
$$

TSLS 的有效性 II

- 第一阶段 OLS 可以保证 $D_i = \hat{D}_i + \check{D}_i$，其中 $\check{D}_i$ 可以被视为第 i 个个体的估计残差，并满足

$$\sum_{i=1}^{n} \hat{D}_i \check{D}_i^{\mathrm{T}} = 0$$

是一个与 D_i 具有相同维度的零方阵。上述正交性也意味着

$$\sum_{i=1}^{n} \hat{D}_i D_i^{\mathrm{T}} = \sum_{i=1}^{n} \hat{D}_i \hat{D}_i^{\mathrm{T}},$$

- 进一步推导可得：

$$\hat{\beta}_{\mathrm{TSLS}} = \beta + \left(\sum_{i=1}^{n} \hat{D}_i \hat{D}_i^{\mathrm{T}} \right)^{-1} \sum_{i=1}^{n} \hat{D}_i \varepsilon_i. \tag{7}$$

TSLS 的有效性 III

- 第一阶段 OLS 拟合还保证

$$\hat{D}_i = \hat{\Gamma}^{\mathrm{T}} Z_i$$

这意味着

$$\hat{\beta}_{\mathrm{TSLS}} = \beta + \left\{ \hat{\Gamma}^{\mathrm{T}} \left(n^{-1} \sum_{i=1}^{n} Z_i Z_i^{\mathrm{T}} \right) \hat{\Gamma} \right\}^{-1} \hat{\Gamma}^{\mathrm{T}} \left(n^{-1} \sum_{i=1}^{n} Z_i \varepsilon_i \right).$$

- 基于上式，我们可以看到 TSLS 估计的相合性，因为

$$n^{-1} \sum_{i=1}^{n} Z_i \varepsilon_i \longrightarrow E(Z\varepsilon) = 0.$$

当 Z 和 D 具有相同维度时，可以验证上式 $\hat{\beta}_{\mathrm{TSLS}}$ 与式(6)$\hat{\beta}_{\mathrm{IV}}$ 相同。

TSLS 的标准误差

- 基于(7)，我们可以如下计算标准误差。首先得到残差 $\hat{\varepsilon}_i = Y_i - \hat{\beta}_{\mathrm{TSLS}}^{\mathrm{T}} D_i$，然后得到方差估计如下：

$$\hat{V}_{\mathrm{TSLS}} = \left(\sum_{i=1}^{n} \hat{D}_i \hat{D}_i^{\mathrm{T}} \right)^{-1} \left(\sum_{i=1}^{n} \hat{\varepsilon}_i^2 \hat{D}_i \hat{D}_i^{\mathrm{T}} \right) \left(\sum_{i=1}^{n} \hat{D}_i \hat{D}_i^{\mathrm{T}} \right)^{-1}.$$

- 重要的是，$\hat{\varepsilon}_i$ 不是第二阶段 OLS 的残差 $Y_i - \hat{\beta}_{\mathrm{TSLS}}^{\mathrm{T}} \hat{D}_i$，所以 $\hat{V}_{\mathrm{TSLS}}$ 与第二阶段 OLS 的方差估计不同。

特殊情况：一维工具变量和一维内生处理 I

- 本节讨论一种简单情况，即工具变量和内生处理变量都是一维的情形，
- 考虑以下结构方程：

$$\begin{cases} Y_i = \beta_0 + \beta_1 D_i + \beta_2^{\mathrm{T}} X_i + \varepsilon_i \\ D_i = \gamma_0 + \gamma_1 Z_i + \gamma_2^{\mathrm{T}} X_i + \varepsilon_{2i} \end{cases}$$

- 其中，D_i 是一个标量内生处理变量，代表感兴趣的处理变量（满足 $E(\varepsilon_i D_i) \neq 0$）；
- Z_i 是一个标量工具变量，代表 D_i 的工具（满足 $E(\varepsilon_i Z_i) = 0$）；
- X_i 包含其他外生回归变量（即 $E(\varepsilon_i X_i) = 0$）。
- 上述结构方程是前面考虑的恰好识别的情形的特殊情况，其中(5)中的 D 被替换为 $(1, D, X)$，(5)中的 Z 被替换为 $(1, Z, X)$。

特殊情况：一维工具变量和一维内生处理 II

- 我们仍然可以采用 TSLS 进行估计:
  1. D 对 $(1, Z, X)$ 进行 OLS 回归，得到拟合值 $\hat{D}_i (i = 1, \ldots, n)$。如果 D_i 是向量，则需要逐分量 (component-wise) 运行 OLS 以获得 $\hat{D}_i$。将拟合向量 $\hat{D}_i^{\mathrm{T}}$ 放入矩阵 $\hat{D}$ 的第 i 行;
  2. 运行 OLS，将 Y 对 $(1, \hat{D}, X)$ 进行回归，得到系数 $\hat{\beta}_{1,\mathrm{TSLS}}$。

特殊情况：一维工具变量和一维内生处理 III

- 我们还可以采用 indirect 的回归估计法:
- 关于 Y 的模型也可以表示为:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 \left(\gamma_0 + \gamma_1 Z_i + \gamma_2^{\mathrm{T}} X_i + \varepsilon_{2i}\right) + \beta_2^{\mathrm{T}} X_i + \varepsilon_i \\ &= (\beta_0 + \beta_1\gamma_0) + \beta_1\gamma_1 Z_i + (\beta_2 + \beta_1\gamma_2)^{\mathrm{T}} X_i + (\varepsilon_i + \beta_1\varepsilon_{2i}) \end{aligned}$$

- 定义 $\Gamma_0 = \beta_0 + \beta_1\gamma_0, \Gamma_1 = \beta_1\gamma_1, \Gamma_2 = \beta_2 + \beta_1\gamma_2$，以及 $\varepsilon_{1i} = \varepsilon_i + \beta_1\varepsilon_{2i}$。我们得到以下方程:

$$\begin{cases} Y_i = \Gamma_0 + \Gamma_1 Z_i + \Gamma_2^{\mathrm{T}} X_i + \varepsilon_{1i} \\ D_i = \gamma_0 + \gamma_1 Z_i + \gamma_2^{\mathrm{T}} X_i + \varepsilon_{2i} \end{cases}$$

特殊情况：一维工具变量和一维内生处理 IV

- 这常被称为 reduced form。感兴趣的参数等于两个系数的比值：

$$\beta_1 = \frac{\Gamma_1}{\gamma_1}$$

- 在 reduced form 中，等式左侧是依赖变量 Y 和 D，等式右侧是外生变量 Z 和 X，且满足以下条件：

$$E(Z\varepsilon_{1i}) = E(Z\varepsilon_{2i}) = 0, \quad E(X\varepsilon_{1i}) = E(X\varepsilon_{2i}) = 0$$

特殊情况：一维工具变量和一维内生处理 V

- reduced form 表明，两个 OLS 系数的比值 $\hat{\Gamma}_1$ 和 $\hat{\gamma}_1$ 即可以估计 β_1。这被称为间接最小二乘（ILS）估计量：

$$\hat{\beta}_{1,\text{ILS}} \equiv \frac{\hat{\Gamma}_1}{\hat{\gamma}_1}$$

- 有趣的是，上述估计量与 TSLS 估计量数值上是相同的。

线性模型下 Control Function Estimator

1. D 对 Z 进行 OLS 回归，得到残差 $\check{D}_i (i = 1, \ldots, n)$。如果 D_i 是向量，则需要逐分量 (component-wise) 运行 OLS 以获得 $\check{D}_i$。将拟合向量 $\check{D}_i^{\mathrm{T}}$ 放入矩阵 $\hat{D}$ 的第 i 行;
2. 运行 OLS，将 Y 对 D 和 $\check{D}$ 进行回归，得到 D 的回归系数 $\hat{\beta}_{\mathrm{CF}}$。

   可以证明：$\hat{\beta}_{\mathrm{CF}} = \hat{\beta}_{\mathrm{TSLS}}$

目录

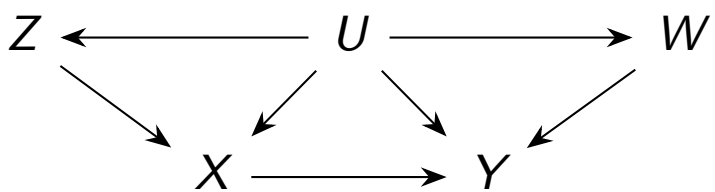
阴性对照 I

- 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 $W \not\perp\!\!\!\perp U$.

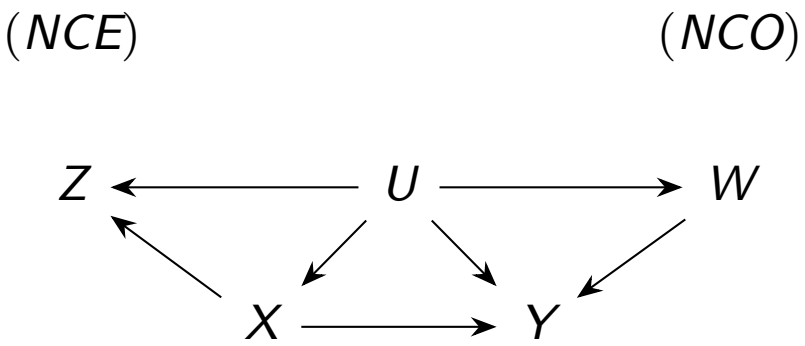
阴性对照 I

- 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: $W \perp\!\!\!\perp X \mid U$, $W \not\perp\!\!\!\perp U$

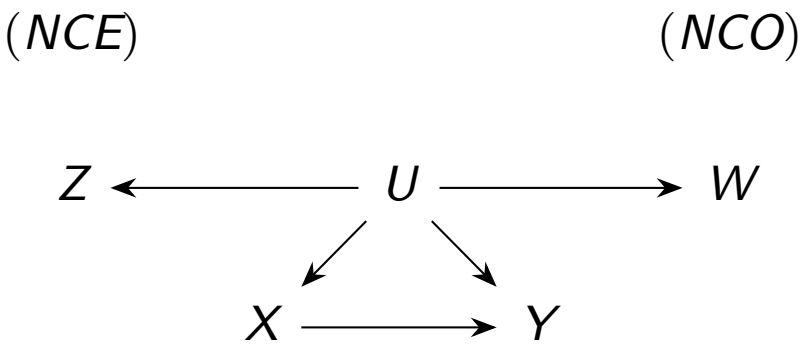
(NCE) (NCO)



阴性对照 I



- 测量误差: Kuroki & Pearl (2014)



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 2 个水平的离散变量。引入以下记号：

$$P(W \mid u) = \{P(w_1 \mid u), P(w_2 \mid u)\}^{\mathrm{T}}$$

$$P(w \mid U) = \{P(w \mid u_1), P(w \mid u_2)\}$$

$$P(W \mid U) = \{P(W \mid u_1), P(W \mid u_2)\}$$

□

阴性对照识别性 II

证明.

- 类似地，我们定义

$$P(U \mid v, x) = \{P(u_1 \mid v, x), P(u_2 \mid v, x)\}^{\mathrm{T}}$$

$$P(U \mid V, x) = \{P(U \mid v_1, x), P(U \mid v_2, x)\}$$

$$P(y \mid V, x) = \{P(y \mid v_1, x), P(y \mid v_2, x)\}$$

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

$$P(W \mid V, x) = P(W \mid U)P(U \mid V, x)$$

$$P(y \mid V, x) = P(y \mid U, x)P(U \mid V, x)$$

□

阴性对照识别性 III

证明.

- 当矩阵 $P(W \mid V, x)$ 可逆时，我们有

$$
\begin{aligned}
P(U \mid V, x) &= P(W \mid U)^{-1} P(W \mid V, x), \\
P(y \mid V, x) &= P(y \mid U, x) P(W \mid U)^{-1} P(W \mid V, x) \\
P(y \mid U, x) &= P(y \mid V, x) P(W \mid V, x)^{-1} P(W \mid U)
\end{aligned}
$$

- 最后，

$$
\begin{aligned}
P\left(Y_x = y\right) &= P(y \mid U, x) P(U) \\
&= P(y \mid V, x) P(W \mid V, x)^{-1} P(W \mid U) P(U) \\
&= P(y \mid V, x) P(W \mid V, x)^{-1} P(W) \quad \text{(Identifiable!)}
\end{aligned}
$$

□

References I

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. Technical report, National Bureau of Economic Research, Inc.

作业 I

1. 上机作业：计算所给数据的双稳健估计量（nnet 及 lasso 方法), 匹配估计量（采用两种方法: L_2 范数匹配和倾向得分匹配)
2. 证明题: 验证工具变量的两阶段最小二乘回归, 可参考文献: Joshua D. Angrist and Jörn-Steffen Pischke. Mostly Harmless Econometrics: An Empiricist's Companion. 中文版, P84 - 92.
3. 作业扫描成 pdf 发邮箱：shan3_luo@163.com
4. 上机作业要求用 Rmarkdown 输出 pdf, 注意注释及格式
5. 截止日期：下周三, 2023.03.15.

作业 I

计算所给数据的

1. 两阶段最小二乘估计
2. 依从组的比例
3. 工具变量对结果变量的因果作用
4. 依从组上的因果作用

- 作业扫描成 pdf 发邮箱：shan3_luo@163.com
- 上机作业要求用 Rmarkdown 输出 pdf, 注意注释及格式
- 截止日期：下周三, 2023.03.22.