第三章　未观测混杂

罗姗姗

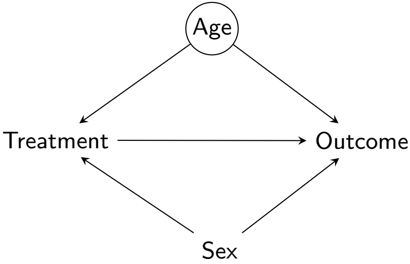
北京工商大学 数学与统计学院

因果推断课题组

目录

未观测混杂

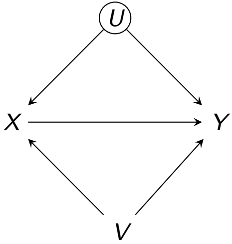
- 匹配、逆概加权、回归和双稳健估计方法的重要前提是可忽略性假定.
- 然而, 在实际研究中, 如果有重要背景变量未被观测、测量误差或者选择偏差, 就有潜在的未观测的混杂因素, 可忽略性假定可能不成立, 前一节介绍的统计推断方法在出现未观测的混杂因素时就有偏差.



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，

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



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。

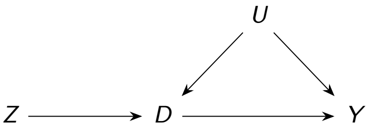
依从组因果作用 IV

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：

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

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

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

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：

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

- 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量如下（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，$\hat{\tau}_{\mathrm{c}}$ 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 $\hat{A}_i = Y_i - \hat{\tau}_{\mathrm{c}} D_i (i = 1, \ldots, n)$；
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量 $\hat{V}_{\hat{A}} / \hat{\tau}_D^2$。

估计 III

- 在零假设下，即 $\tau_{\mathrm{c}}=0$，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 估计器乘以 $1/\hat{\tau}_D > 1$，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。

... 后续内容待调整...

Outline

1. 含未知混杂下因果推断的本质困难
2. 工具变量-经济学家视角下的 de-confounding
   - 单调性假设
   - 线性模型
3. 阴性对照-非参数的识别思想

工具变量的关键假设

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:

- 除了前一节所提到的单调性假定 (monotonicity).
- 另一种是因果作用的同质性假定 (effect homogeneity).

同质性假定-线性模型

$$\begin{tikzcd} & & U \arrow[dl] \arrow[dr] & \\ Z \arrow[r] & X \arrow[rr] & & Y \end{tikzcd}$$

- 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, β_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, β_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 β_1.
- 例如, 当 $\mathbb{E}(U \mid X) \neq 0$ 时, β_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 β_1:

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

证明.

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

□

同质性假定-线性模型 II

二值情形.

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

□

同质性假定-线性模型 III

- 在线性模型假设下, β_1 的 IV 估计量是协方差之比:

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

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:

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

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 β_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 π_{11} 和 π_{21} 的最小二乘估计之比:

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

在二值的随机试验中,$\widehat{\beta}_{1,\mathrm{LS}}$ 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 $\widehat{\pi}_{11}$ 和 $\widehat{\pi}_{21}$ 代入, 即得到 β_1 的工具变量估计 (instrumental variable estimator).

解释 1-两次回归 II

- 即使 $\mathbb{E}(U \mid X) \neq 0$, 在假定 (i)-(iii) 和一定的正则条件下, 可以证明 $\widehat{\beta}_{1,\mathrm{LS}}$ 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 $\widehat{X}_i$:

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

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:

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

从第 2 阶段通过 OLS 估计 β_1, 得到 β_1 的两阶段最小二乘估计量: $\widehat{\beta}_{1,\text{2sls}}$.

- 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:

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

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 $\operatorname{plim} \beta_1^{\mathrm{iv}}$ 表示当样本量趋于无穷时 β_1^{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

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

- 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 σ_{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)

$$
\begin{tikzcd}
Z \arrow[dr] & & U \arrow[ll] \arrow[rr] \arrow[dl] \arrow[dr] & & W \arrow[dl] \\
& X \arrow[rr] & & Y &
\end{tikzcd}
$$

阴性对照 I

$$\begin{tikzcd}
(NCE) & & & & (NCO) \\
Z & & U \arrow[ll] \arrow[rr] \arrow[dl] \arrow[dr] & & W \arrow[dl] \\
& X \arrow[ul] \arrow[rr] & & Y &
\end{tikzcd}$$

- 测量误差: Kuroki & Pearl (2014)

$$\begin{tikzcd}
(NCE) & & & & (NCO) \\
Z & & U \arrow[ll] \arrow[rr] \arrow[dl] \arrow[dr] & & W \\
& X \arrow[rr] & & Y &
\end{tikzcd}$$

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：

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

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

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

□

阴性对照识别性 II

证明.

- 类似地，我们定义

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

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

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

- 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，

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

□

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.