

Causality, IV and DID Method

张剑

zhangjian@sicau.edu.cn

四川农业大学经济学院

January 7, 2020

目录

- 1 因果图谱
- 2 IV Method
- 3 DID Method

如何定义因果？

因果关系

潜在结果框架 (potential outcome framework) 或 Rubin 因果模型 (Rubin Causal Model, RCM)

因果关系和相关关系有很大区别，经济学更希望探索因果关系!!

潜在结果框架 (Potential outcome framework)

Rubin 因果模型 (Rubin Causal Model, RCM)

- 潜在结果 (potential outcome): 每中状态下的对应结果
- 分配机制 (assignment mechanism): 描述个体为何处于不同的状态。
- 稳定性假设 (Stable Unit Treatment Value Assumption, SUTVA): 是指任意一个个体的潜在结果不随其他个体是否接受处理而改变, 而且每个个体受到的处理和潜在结果之间是定义良好的函数

潜在结果框架

以大学教育对收入的影响为例，每个个体 i 都有两种状态：

$D_i = 1$ 表示完成大学教育

$D_i = 0$ 表示完成高中教育

每种状态下对应于一种潜在的结果： $Y_i(1)$ ，个体 i 完成大学教育时的收入， $Y_i(0)$ ，个体 i 未完成大学（完成高中）教育时的收入。

大学教育的因果效应：

$$\tau_i = Y_i(1) - Y_i(0)$$

如何定义因果？

但是，人不可能同时生活在两个世界，个体不可能同时处于两种状态。

观察结果是：

$$\begin{aligned} Y &= DY(1) + (1 - D)Y(0) \\ &= \begin{cases} Y(0) & \text{if } D = 0 \\ Y(1) & \text{if } D = 1 \end{cases} \end{aligned}$$

对任何个体， (Y_{1i}, Y_{0i}) 只以观察到一个，面临数据缺失问题！

因果推断的基本问题（Holland, 1986）。

因果推断的基本问题

只能观测到一种状态下的结果:

$$\begin{aligned} Y &= DY(1) + (1 - D)Y(0) \\ &= \begin{cases} Y(0) & \text{if } D = 0 \\ Y(1) & \text{if } D = 1 \end{cases} \end{aligned}$$

个体因果效应:

$$\tau_i = Y_i(1) - Y_i(0)$$

必须估计反事实结果 (counterfactual outcome), 即观测不到的

潜在结果。

如何估计? ——要用多个个体!

因果推断的基本问题

- 最简单的办法，找一个在各方面都一样的个体，只是状态不一样，作为反事实结果的估计。
- 假设个体 i 是大学毕业，状态为 $D=1$ ，观测收入 $Y_i = Y_i(1)$ ，大学教育对其收入的影响为

$$\tau_i = Y_i(1) - Y_i(0)$$

- $Y_i(0)$ 观测不到，是反事实结果。
- 如果能够找到一个 j 和 i 一模一样，只是仅完成高中教育，则 j 的观测收入 $Y_j = Y_j(0)$ 。可以用 $Y_j(0)$ 代替 $Y_i(0)$ 。
- 这实际上是控制方法 (conditioning)

因果效应参数

- 总体平均因果效应 (Average Treatment Effect, ATE)

$$\tau_{ATE} = E[Y_{1i} - Y_{0i}]$$

- 干预组平均因果效应 (Average Treatment Effect on the Treated, ATT)

$$\tau_{ATT} = E[Y_{1i} - Y_{0i} | D_i = 1]$$

- 控制组平均因果效应 (Average Treatment Effect on the control, ATC)

$$\tau_{ATC} = E[Y_{1i} - Y_{0i} | D_i = 0]$$

- 条件平均因果效应 (Conditional Average Treatment Effect, CATE)

$$\tau_{ATE}(x) = E[Y_{1i} - Y_{0i} | X_i = x]$$

分配机制 (Assignment Mechanism)

分配机制就是决定哪些个体在干预组，哪些个体在控制组的程序。或者，是决定哪个潜在结果会被显示出来的机制。

分配机制的分类：

- 随机化分配 (random assignment)
- 根据观测变量进行的选择 (selection on observables)
- 根据未观测变量进行的选择 (selection on unobservables)

CIA

根据观测变量进行的选择，也称为非混杂性 (unconfoundedness)，可忽略性 (ignorability)，或条件独立性 (CIA)。

因果图

- 点 nodes 代表随机变量
- 边 edges (箭头), 从一个变量指向另一个变量, 表示一个变量对另一个变量, 至少对 1 个个体有直接因果效应。
- 如果两变量间没有箭头, 表示总体上任何个体, 两个变量之间没有直接的因果影响。

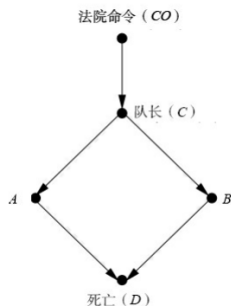


Figure 1: 一个例子

因果图的例子

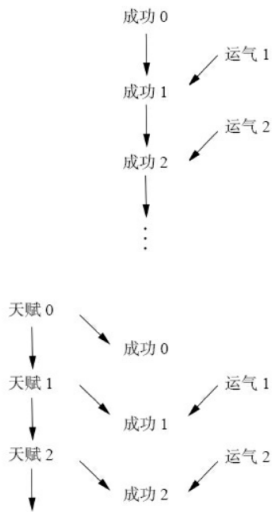


Figure 2: 更多例子

因果图的三种基本结构



Figure 3: 因果影响

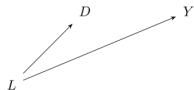


Figure 4: 共同原因

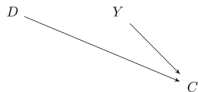


Figure 5: 共同结果

对上面三种关系的简单解释

- 第一种原因 D 是完全随机化分配的, D 到 Y 只有一条因果路径, 那么 D 对 Y 有因果, 两者肯定相关。因果有方向, 而相关没有方向。
- 第二种情况, 假设 L 为是否有抽烟行为, D 为是否随身带打火机, Y 为患肺癌的风险。
- 一般认为抽烟 L 会导致肺癌风险 Y 增加, 因而 L 对 Y 有因果影响。一个常识是抽烟的人 L , 经常带打火机 D 。那么我们能否推出以下结论?

混杂偏差 (confounding)?

由于我们观察到 D 和 Y 是高度相关的, 于是推出带打火机导致了肺癌风险增加?

对上面三种关系的简单解释

- 第三种情况，D 和 Y 之间没有箭头联系，表示两者没有直接的因果关系，但是它们共同决定变量 C，C 被称为共同结果变量
- 假想一个情景，D 是民族，Y 是学习能力，C 是是否被大学录取
- 但是在大学录取中，我国有少数民族优惠政策，少数民族学生会有一定的加分，这样在相同的高考成绩下，少数民族学生录取概率更大，从而 D 影响 C。同样显然 Y 也影响 C

下面的论述是真的嘛？

如果我们控制了 C，也就是以 C 为研究样本，发现 D 和 Y 高度相关，能否推出 D 是 Y 的因？

什么时候变量间会呈现相关性

- 第一种：即使没有因果关系，如果有共同的原因 (confounding)，我们关心的变量也会呈现相关性。
- 第二种：以通过结果为条件，也会打开相关路径，造成估计结果偏差，这被称为样本选择偏差 (Sample Selection Bias) Heckman 1979。
- 一个变量是另一个变量的原因： $D \rightarrow Y$

什么时候会阻断相关性是变量相互独立？

以下三种情况会使得原来相关的变量，变得独立：

- 以中介变量（mediator）为条件时， $D \rightarrow [B] \rightarrow Y$
- 以共同原因（Confounding）为条件时， $D \leftarrow [L] \rightarrow Y$
- 有共同结果，或者说不已共同结果为条件时， $D \rightarrow L \leftarrow Y$

两种偏差

1. 混杂偏差（confounding bias）——共同原因导致
2. 样本选择偏差（sample selection bias）——共同结果导致

后门路径 Back-door Criteria

定义

Pearl (2009) 证明因果效应可以通过以一些变量的集合 z 为条件进行识别，当且仅当所有的原因变量和结果变量之间的后门路径可以通过以 Z 为条件而阻断。

他证明所有的后门路径可以被 z 阻断，当且仅当每条后门路径满足：

- 包含一个中介路径 $A \rightarrow C \rightarrow B$, 其中中介变量 $C \in Z$
- 包含一个共同原因的路径 $A \leftarrow C \rightarrow B$, 其中 $C \in Z$
- 包含一个共同结果的路径 $A \rightarrow C \leftarrow B$, 但共同结果 C 及所有 C 的子孙不在 Z 中

一些后门路径的例子

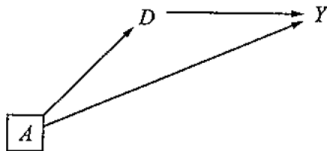


Figure 6: 教育收益率

- 其中 A 是个人能力，教育 D ，个人收入 Y ， D 和 Y 之间有两条路径，一条因果路径 $D \rightarrow Y$ ，一条后门路径 $D \leftarrow A \rightarrow Y$
- 那么 D 到 Y 的相关性保函了两部分，一部分是真正的因果关系，另一部分是混杂因素 A 造成的相关性
- 只要想办法把后门路径关闭即可得到真正的因果关系。
- 利用后门规制第二条，以共同原因 A 为条件，阻断教育和能力之间的非因果路径

一些后门路径的例子 2

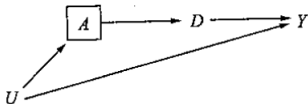


Figure 7: 班级规模与学习成绩

- D 是班级规模，Y 是成绩，A 是地方政府财力，U 是移民在社区中的比重
- 假设根据理论分析我们得到了上述的因果关系图，我们关心班级规模对学习成绩的因果关系
- 因果路径 $D \rightarrow Y$ 和后门路径 $D \leftarrow A \leftarrow U \rightarrow Y$
- 根据后门规制第一条或者第二条，以 A 或者 U 为条件均可以阻断后门路径，从而得到 D 对 Y 的因果

一些后门路径的例子 3

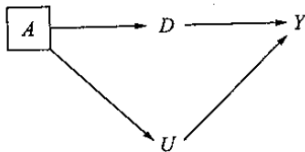


Figure 8: 幸福的决定因素

- A 为个人教育情况，D 是收入，Y 是幸福感，U 是宗教信仰
- 我们希望研究收入 D 对幸福感 Y 的影响
- 因果路径 $D \rightarrow Y$ ，和后门路径 $D \leftarrow A \rightarrow U \rightarrow Y$
- 根据后门规制第一条或者第二条，以 A 或者 U 为条件均可以阻断后门路径，从而得到 D 对 Y 的因果

一些后门路径的例子 4

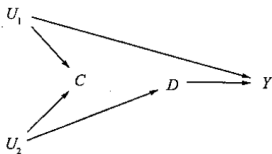


Figure 9: 样本选择偏差

- 两个变量的共同结果为**条件造成的**，这两个变量中一个原因变量或与原因变量相关，另一个变量是结果变量或与结果变量相关。
- 这一共同结果可以发展在结果之后，也可以发展在干预变量与结果变量之间，还可以发生在干预变量之前
- 上图 C 、 D 、 Y 的关系， U_1 、 U_2 为无法观测变量
- 如果以 C 为条件，不但不能阻断后门路径，反而打开后门路径，注意前面后面规则的第三条

共同结果——样本选择偏差的例子 1

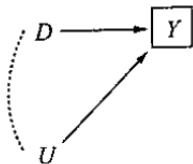


Figure 10: 样本选择偏差 2

- Hausman and Wise(1977) 考察教育 D 对个人收入 Y 的印象，但是使用了一个仅包括低收入者的样本。
- 如上图所示，教育 D 和其他因素 U 共同决定了 Y ，假设 D 与 U 独立，因果路径 $D \rightarrow Y$ 体现了教育对收入的因果，但是由于样本现对于低收入人群，相当于以低收入为条件
- Y 是 D 和 U 的共同结果，以 Y 为条件，使得 D 与 U 相关，用虚线表示。这时 D 到 Y 的相关性，除了反映因果路径 $D \rightarrow Y$ ，还反映了 $D - U \rightarrow Y$ ，打开了后门路径

共同结果——样本选择偏差的例子 1

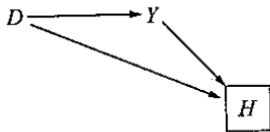


Figure 11: 农村教育回报率估计

- D 教育，Y 是收入，H 是户籍
- 由于户籍限制，在城市化进程中，农村中有能力的个体率先通过升学、参军等途径突破户籍进入城市体系，在调查样本中无法观测到这些已经成为了城市居民的**农村居民**，调查样本中只留下了那些没有办法突破户籍的农村居民，相当于以农村居民（户籍 H）为条件。
- 以 H 为条件，打开了后门路径 $D \rightarrow H \leftarrow Y$ ，产生样本选择偏差

共同结果——样本选择偏差的例子 1

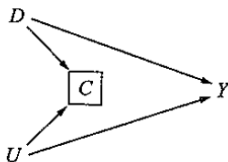


Figure 12: 研发投入因果图

- D 是研发， Y 是企业绩效， C 表示是否流失， U 是影响 Y 和 C 的无法观测变量。其中 C 的含义是，很多文献使用中国企业数据库很多企业干一干就失败了，离开了调研样本。
- 以 C 为条件，或者说以活下来的、继续统计的样本为条件（死掉的企业没有办法进入后续数据库），打开了后门路径 $D \rightarrow C \leftarrow U \rightarrow Y$ ，从而产生样本选择偏差

好的控制变量和差的控制变量

好的控制变量

好的控制变量应该是发生在原因变量之前或者不随原因变量变化的变量

不好的控制变量

坏的控制变量是那些受到原因变量影响的变量或者发生在原因变量滞后的变量。

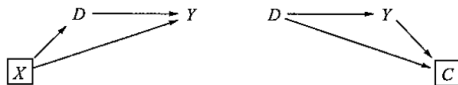


Figure 13: 好坏控制变量

Iv Method

$$Y_i = \alpha + \tau D_i + A_i' \gamma + v_i$$

- Y 是收入, D 为教育年限, A 是能力, v 是扰动项, 如果能力变量 A 可以观察, 那么要识别 D 对收入 Y 的影响, 直接利用回归控制 A , 以 A 为条件, 阻断后门路径 $D \leftarrow A \rightarrow Y$ 就可以得到 $D \rightarrow Y$ 的因果关系。
- 如果能力 A 不可观测, 这是大多数的情况, 那么我们就没有办法控制 A 。如果能找到一个工具变量满足下面两个基本条件可以识别 D

IV 的两个基本条件

1. 相关性。 $\text{Cov}(D_i, Z_i) \neq 0$
2. 外生性。工具变量不通过无法观测因素影响结果。
 $\text{Cov}(\eta_i, Z_i) = 0$

IV 图示

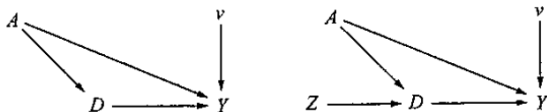


Figure 14: 工具变量法因果图

- Z 为工具变量，直接影响 D ，从而与 D 强相关（相关性要强）。
- Z 与 A 和 v 都不相关， Z 不会通过 A 和 v 影响 Y
- 如果我们没有办通过阻断 A 来识别， $D \rightarrow Y$ 。如果有工具变量 Z ，当 Z 变化时**直接通过影响 D 变化**，最终影响 Y 变化。
- 因此我们知道 Y 这时的变化，是因为 Z 的变化通过 D 影响到了 Y ，这时 D 和 Y 表现出的相关性就是 $D \rightarrow Y$ 的因果

IV 的简单推导

$$\text{Cov}(Y_i, Z_i) = \text{Cov}(\alpha, Z_i) + \tau \text{Cov}(D_i, Z_i) + \text{Cov}(\eta_i, Z_i)$$

- 显然 $\text{Cov}(\alpha, Z_i) = 0$
- IV 外生条件告诉我们 $\text{Cov}(\eta_i, Z_i) = 0$
- IV 相关性条件告诉我们 $\text{Cov}(D_i, Z_i) \neq 0$

于是整理得到：

$$\tau_{IV} = \frac{\text{Cov}(Y_i, Z_i)}{\text{Cov}(D_i, Z_i)} = \frac{\text{Cov}(Y_i, Z_i) / \text{Var}(Z_i)}{\text{Cov}(D_i, Z_i) / \text{Var}(Z_i)}$$

两阶段最小二乘--第一阶段

- 第一阶段: $D_i = \gamma + \delta z_i + u_i$, 找到一个工具变量 z , 将内生变量 D 对 z 进行 OLS 回归, 分离出 D 的外生部分
- 第一阶段拟合值: $\hat{D}_i = \hat{\gamma} + \hat{\delta} z_i$
- 第一阶段残差: $\hat{u}_i = D_i - \hat{D}_i$
- 第一阶段回归将内生的 D 分解为两部分: $D_i = \hat{D}_i + \hat{u}_i$; 其中的 \hat{D}_i 为外生部分, 其余残差 \hat{u}_i 为内生部分。

两阶段最小二乘——第二阶段

- 既然第一阶段 \hat{D}_i 为外生部分，故而将 \hat{D}_i 替代元模型中的内生变量 D_i 为外生部分，既可以用 OLS 得到一致的估计

$$y_i = \alpha + \beta \hat{D}_i + (\varepsilon_i + \beta \hat{u}_i)$$

- 上式为第二阶段回归
- 如果包含外生控制变量

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

其中 x_{i1} 为内生， x_{i2} 为外生变量，需要注意的是，第一阶段回归时，需要把外生变量 x_{i2} 也放入其中。

IV 有效性问题

- 第一个：工具变量 Z 需要与内生变量 x 强相关。这个容易检验，只要看第一阶段的 F 值即可，一个拇指法则，第一阶段 F 值大于 15 以上。
- 第二个：工具变量外生性，如果工具变量“不干净”，**工具变量与不可观测的变量或者说扰动项相关**，那么通过第一阶段分离出来的外生性部分也必然“不干净”。

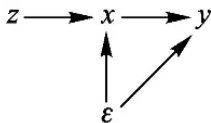


Figure 15: 工具变量法因果图

- 如果 z 满足外生性，那么 z **不可能通过除 x 以外的其他因素或者渠道影响 y** ，这叫做排他性约束 (exclusion restriction)

IV 中的排他性约束

- 在实践中需要将 z 影响 y 的所有可能渠道一一列出，然后将除 x 以外的渠道全部排除
- 如果发现 z 可能通过其他渠道，比如说 w 影响 y ，那么我们需要将 w 也作为控制变量，**以此来阻断后门路径**
- 如果 w 不可度量或无法控制，那么**工具变量 z 的外生性是可疑的**
- 因此在做论文时，**需要仔细讨论** z 的外生性。比如Miguel et al.(2004)就是一篇很好的例子。(自行点击链接查看论文)

IV 近乎外生情况下的推断

- 现实中很难找到绝对干净的 Z , 考虑下面的方程, 把工具变量 z 也放入原模型中

$$y_i = \alpha + \beta x_i + \gamma z_i + \varepsilon_i$$

- 如果工具变量 z 只通过内生变量 x 影响 y , 那么上式中, $\gamma = 0$ 应该成立, 如果 $\gamma \neq 0$, 则表明工具变量有轻微的内生性。
- 如果能够给出 γ 的取值范围或者先验分布, Conley et al. (2012) 表明可以根据 γ 信息购置估计系数 β 的置信区间, stata 给出了相应程序: `ssc install plausexog`
- 同样的 Berkowitz et al. (2012) 在 stata journal 上给出的 `far` 命令

DID Method

- 需要明确的是一般而言，DID 仅适用于面板数据，因此在只有截面数据时，基本无法使用 DID。
- DID 基准模型

$$Y_{it} = \alpha_0 + \alpha_1 du + \alpha_2 dt + \alpha_3 du \cdot dt + \varepsilon_{it}$$

- 其中 du 为政策干预虚拟变量，如果个体 i 受到政策干预——属于处理组， du 取值为 1；否则 i 属于对照组， du 取值为 0
- dt 为政策实施的时间虚拟变量，政策实施之前 dt 取值为 0，政策实施之后 dt 取值为 1
- $du \cdot dt$ 分组虚拟变量与政策实施虚拟变量的交互项，其系数 α_3 就反映了政策实施的净效应

DID 实施的两个基本条件

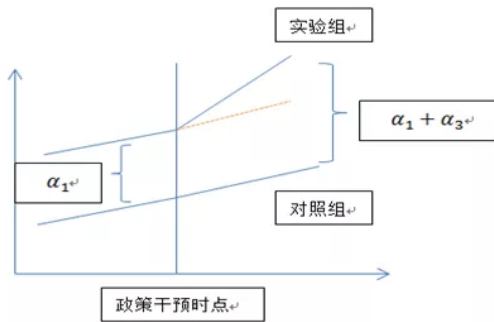
- 从 DID 的模型设置来看，要想使用 DID 必须满足以下两个关键条件：
- 一是必须存在一个具有试点性质的政策冲击，这样才能找到处理组和对照组，那种一次性全铺开政策并不适用于 DID 分析
- 二是必须具有一个相应的至少两年（政策实施前后各一年）的面板数据集

DID 中 α_3 的含义解读

	政策实施前	政策实施后	Difference
处理组	$\alpha_0 + \alpha_1$	$\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3$	$\alpha_2 + \alpha_3$
对照组	α_0	$\alpha_0 + \alpha_2$	α_2
Difference	α_1	$\alpha_1 + \alpha_3$	α_3 (D-in-D)

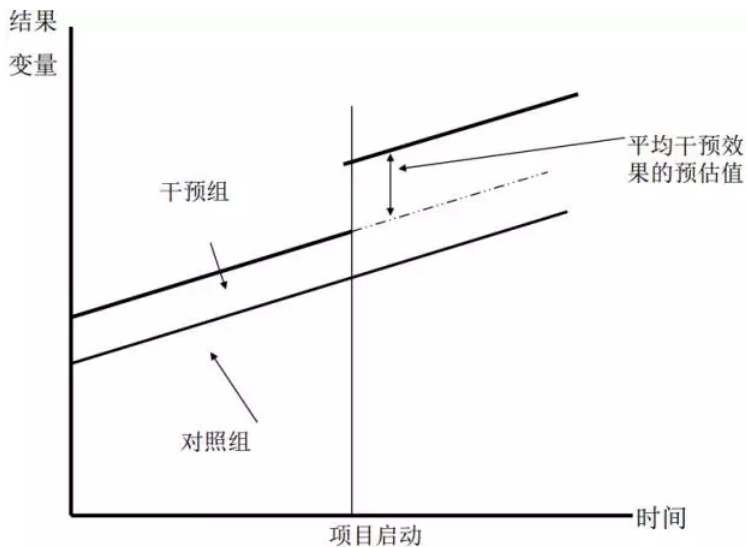
双重差分法的基本思想就是通过对政策实施前后对照组和处理组之间差异的比较构造出反映政策效果的双重差分统计量

图示 DID



- 图中红色虚线表示的是假设政策并未实施时处理组的发展趋势, **共同趋势 (Common Trends)** : 处理组和对照组在政策实施之前必须具有相同的发展趋势
- DID 的使用不需要什么政策随机以及分组随机, 只要求 CT 假设, 因此用 DID 做论文时必须对该假设进行验证

图示 DID2



多年面板数据下的 DID 基准模型

- 如果数据集包含多年的面板，那么基准 DID 模型也可以写成：

$$Y_{it} = \alpha_0 + \alpha_1 du \cdot dt + \lambda_i + v_t + \varepsilon_{it}$$

- 这两个模型本质上是一样的，更多时候我们是使用这一页的模型。 λ_i 控制了个体固定效应， v_t 控制了时间固定效应
- 这也是 DID 方法一定程度上可以减轻遗漏变量偏误的原因

多年面板数据下的 DID 基准模型 2

- 为了防止遗漏变量导致的偏误，如果样本量比较大的时候我们还可以控制更多的固定效应，一般来说我们可以控制省份 \times 年份固定效应，还可以控制行业 \times 年份，模型就跟着改：

$$Y_{it} = \alpha_0 + \alpha_1 du \times dt + year_t \times ind_i + prov_i \times year_t + \lambda_i + v_t + \varepsilon_{it}$$

- 这个模型和上面模型也没啥区别，无外乎控制了更多的省或者行业随着时间变化的效应用以捕捉那些行业或者地区随着时间变化的不可观测的因素
- 记住 DID 就是面板，面板那一套都可以用在这里

DID 的稳健性假设——重要步骤

- 共同趋势的检验。一些文献 DID 时只有政策实施前后各一年的数据，这样的话根本无法验证政策实施前的趋势问题。（这种文章发好刊基本没戏）
- 即便处理组和对照组在政策实施之前的趋势相同，仍要担心是否同时发生了其他可能影响趋势变化的政策——处理变量对结果变量作用机制需要排他。
- 一些解决办法：A 安慰剂检验，即通过虚构处理组进行回归，具体可以：a) 选取政策实施之前的年份进行处理。比如原来的政策发生在 2008 年，研究区间为 2007-2009 年，这时可以将研究区间前移至 2005-2007 年，并假定政策实施年份为 2006 年，然后进行回归

一些解决办法

- b) 选取已知的并不受政策实施影响的群组作为处理组进行回归。如果不同虚构方式下的 DID 估计量的回归结果依然显著，说明原来的估计结果很有可能出现了偏误。
- 可以选取一个完全不受政策干预影响的因素作为被解释变量进行回归，如果 DID 估计量的回归结果依然显著，说明原来的估计结果很有可能出现了偏误。
- 可以利用不同的对照组进行回归，看研究结论是否依然一致。（更换分组标准）

DID 代码示例

相关 DID 代码示例, [链接在这里](#), html 文件下载后用浏览器打开

谢谢!