

SCM 与 RCM?因果框架简要比较

滑翔闪

2026 年 10 月 11 日

复习笔记，为已经学完常见计量框架的经济学同学简要补充关于 pearl 因果框架的知识点。核心启发是控制变量的选择问题，也可以看作模型形式设定的问题。

主要参考资料：

- 《the book of why》(Judea Pearl)
- 《Causality: Models, Reasoning, and Inference》(Judea Pearl)
- 《因果推断实用计量方法 》(邱嘉平)
- 《[econometrics-and-causal-inference2025](#)》

1 因果框架

计量经济学常见的因果框架来源于 [Rubin, Donald B. \(1974\)](#), 其率先在论文里提出潜在事实结果 (Potential Outcomes), 建立了**潜在结果模型** (Potential Outcomes Framework, 简称 RCM)。

在 rubin 因果框架的影响下, 大部分计量教材, 将内生性问题归结为以下三点:

- 测量误差
- 双向因果
- 遗漏变量

因果分析因经济学而发扬光大, 但因果分析并不是经济学的专利 😊。计算机领域的 Judea Pearl 2018 年的《the book of why》使用图论的形式研究因果, 叫做**结构因果模型** (Structural Causal Model, 简称 SCM), 这个名称的结构并非理论数学模型意义上的结构, 而是**图论关系**上的结构, 因此, 这个因果框架形式上又可以叫做**有向无环图分析** (Directed Acyclic Graph, 简称 DAG)

SCM, RCM 都从经典的**相关不等于因果**这个思路出发阐释。Pearl 尝试将自己的工作看作 rubin 的延伸 (参见 pearl 的另外一本著作《Causality: Models, Reasoning, and Inference》^[1]), 并进一步纳入贝叶斯, 神经网络作为因果科学发展的一部分, 不过 rubin 自己并未回应过这个看法。

很多文献讨论对比过两个框架的差异, 例如 [Guido Imbens \(2020\)](#) 曾写过这方面的综述 [\[2\]](#), 认为图模型在实际经济学研究中用处不大。pearl 因此写了一个[长文](#)反驳并在评论区和 Imbens 展开辩论。

笔者的观点是, 两个框架存在交集, 但并不存在包含关系, 而后者可能是一个常见的误区: **认为估计偏误都能被测量误差、双向因果、遗漏变量三种内生性偏误所概括; 或者认为所有的估计偏误都能被图关系表达?** 但事实并非如此 🤖^[2]。三种内生性偏误是 RCM 框架总结出的典型, 但是**估计偏误不只有内生性偏误问题**, 例如假设违背、模型估计不当依旧会导致估计偏

¹2009 年, pearl 早期因果推断作品, 更加晦涩复杂。他在第七章脚注吐槽 [\(angrist 等, 1996\)](#) 明确拒绝引用图论表达是错误的选择。

²估计偏误是一个比内生性偏误大得多的问题, 而由于实证经济学的内生性重点拷打, 可能大部分人下意识把所有估计问题都往内生性偏误里套, 这是个不好的思维惯性。笔者自己就曾陷入这样的思维惯性误区

误。估计偏误，除去内生性问题，形式不当（指数、幂关系、高次数关系等）外，还有控制变量选择不当，其中包含了遗漏变量³和过度控制变量。而图关系能为模型的控制变量好与坏选择提供更直观的表达。

个人理解，仅供参考 😊。

2 DAG 的基本结构

一种错误的认知是 😞：因果关系就是有方向的相关关系；因果关系的必要条件是相关关系。后续我们将知道，由于遮蔽效应的存在，统计相关并不一定表现为因果的必要条件。

在现实应用中，相关只是一种统计关系上的相关，而因果关系，至少到现在，依旧不存在一个严谨逻辑框架下的表达概念。即便我们学习了 pearl 和 rubin 的因果框架，依旧发现他们的表达比较晦涩抽象，并不优雅，也没有真正统一。哲学上何为因果，数理形式上的因果表达，依旧存在很大的探索空间。

2.1 DAG

DAG 使用无环，有向的路径图表达变量之间的关系。

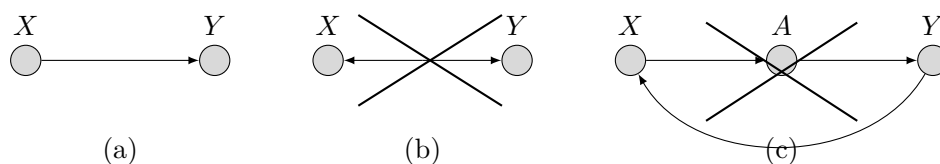


图 1: 无环路径图

路径图，意味着 (a) 的情况，从一个点指向另外一个点的单项路径。

无环，意味着两件事，点与点之间的路径不存在环的结构，因此，(b) 中的双向关系、(c) 中的环关系不能被有向无环图表达⁴。

³ 遗漏变量既可以看作内生性问题，也可以看作模型形式设定不当问题，是 RCM 与 SCM 的交叉点

⁴ 表达这些需要使用更复杂的路径图工具，可能就是因此，rubin 的潜在结果框架更方便表达双向因果存在的问题

2.2 因果路径

在以前，人们认为只要有 X 与 Y，加上时间关系，我们就可以推测因果。例如 X 与 Y 总是一起出现，而 X 总发生在 Y 之前，我们会推测是 X 引发了 Y。但是事实证明，这里面存在着诸多的谬误；即便没有其他变量带来的谬误，共时性的因果也注定了时间序列分析的局限。

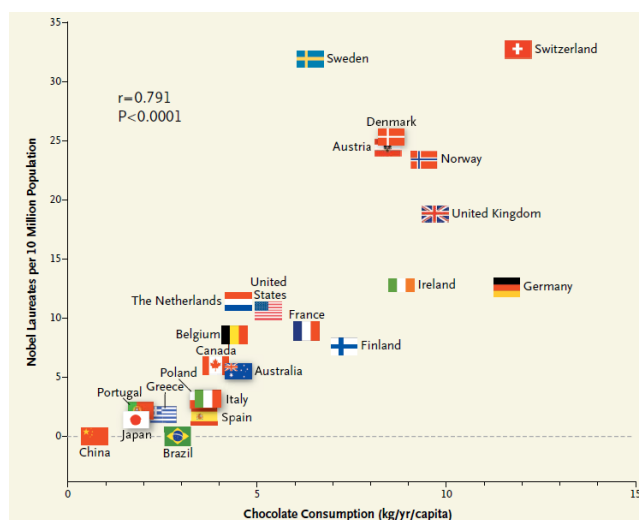


图 2: 诺奖与巧克力的关系

图片来源, Franz H. Messerli, M.D.2012 年的论文《Chocolate Consumption, Cognitive Function, and Nobel Laureates》

例如诺奖与巧克力相关性（图 2），巧克力越多，诺奖得主越多，X 与 Y 的关系图非常漂亮，而且我们知道，巧克力的产量一般也会在诺奖得主之前决定，此时依赖 X 与 Y 加时间关系的推断就走不通了，必须引入第三个变量⁵与 X、Y 的关系并加以分析。正是因为现实有数不清的 Z 变量在 X，Y 之间遮掩，所以 pearl 才说我们要定义的是关系图下的因果结构。

3 混淆变量、后门路径

3.1 混淆变量

在 rubin 的 RCM 框架下，遗漏变量会导致估计量的偏误。

⁵出于奥卡姆剃刀原则，一般只分析 Z,X,Y 的结构。更复杂的结构可以看作前面结构的组合

假设真实模型为：

$$Y = \beta X + \gamma Z + \epsilon$$

但是研究者遗漏了变量 Z，直接回归 Y 对 X，估计偏差为：

$$\text{Bias}(\hat{\beta}) = \gamma \frac{\text{Cov}(X, Z)}{\text{Var}(X)}$$

在 pearl 的 SM 框架中，这种情况是**混淆变量**存在的一种情况，若不控制变量，就会产生遗漏变量的问题。

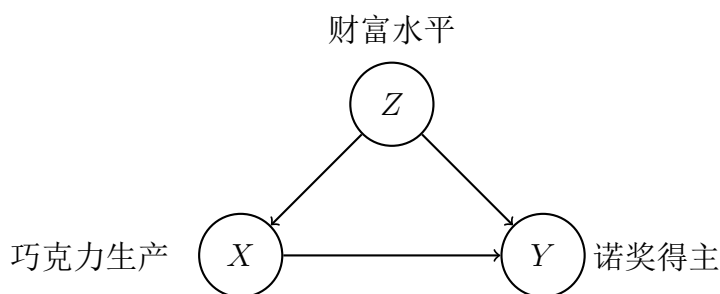


图 3: 后门路径

在 X 指向 Y 的路径中，第一个指向 X 的路径就是它的后门，应当加以控制，pearl 把这种关系叫做**阻断后门**，减少混淆变量的干扰。以图 2 为例子，就是遗漏了控制变量财富水平。同理，冰淇淋销量上升和鲨鱼伤人事件上升往往同步发生，导致这种统计观察的混淆变量是夏天。夏天到了，冰淇淋销量增加，人们也更愿意去海滩 🏖️。

3.2 辛普森悖论

更为经典的例子是**辛普森悖论**，总体与局部会呈现截然不同的结果。

例如统计收入 x 与咖啡消费 y 的关系，黑色代表男性，灰色代表女性，随着收入的增加，咖啡销量增多。但是如果我们把男生和女生的数据放在一起分析，却会得到收入越多，消费越低的情况 🐒。

我们可以用三种形式来表达辛普森悖论，但是结论殊途同归：

矩估计，基于协方差矩阵表示

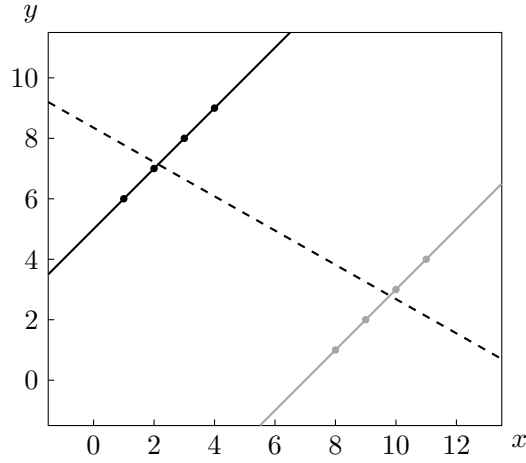


图 4: 辛普森悖论

例子来自范里安《微观经济学：现代观点》17 测度章节

$$\text{Cov}(x, y) = E[\text{Cov}(x, y | z)] + \text{Cov}(E[x | z], E[y | z])$$

其中**第一项**为组内趋势，代表控制性别后，各自性别内部的趋势；**第二项**为组间趋势，依旧是性别分组之间，男生组和女生组之间的差异。由于性别存在异质性，而组间趋势幅度超越组内趋势时，符号发生了变化。也就是此时系数含义的变化更多的反应了男性组转到女性组所带来的差异。

计量估计，基于 fw1 定理，遗漏变量带来的估计偏误：

$$E[\tilde{\beta}_1] = \beta_1 + (X_1' X_1)^{-1} X_1' X_2 \beta_2$$

贝叶斯到 DAG:

$$P(Y | X) = \sum_{z \in \text{男, 女}} P(Y | X, Z = z) \cdot P(Z = z | X)$$

其中 $P(Y | X, Z = z)$ 是特定性别下，收入对消费的影响； $P(Z = z | X)$ 是给定收入 X 下，判断某个人从属于性别后验概率。而后者基于的性别后验权重改变了整体方向的估计。例如，由于数据中，女性收入较低，可以推测收入较低都是女性，而女性的咖啡收入弹性比男性低，最终呈现出异常的结果。

为避免混杂因子产生影响，控制后门路径，此时需要将 $P(Z = z | X)$ 切换为边缘概率 $P(z)$ ，也就是不依赖于 x 进行后验修正，全局的性别占比概率，这时候式子就变成了：

$$P(y | \hat{x}) = \sum_z P(y | x, z)P(z)$$

满足一致性时，前面三者的表达可以相互转换，因此 pearl 认为 SCM 和 RCM 的表达是可以相互转换的 $\bigcirc$ 。

更为特殊的混淆变量就是无法观测的时间效应、个体效应，这些在计量中被各种方式进行处理。

4 中介变量、前门准则

4.1 前门准则

有些混淆因子可以被控制（性别、行业、季节等），有些则很难被控制（文化、个体素质、能力）。因此 pearl 从图论的角度退一步放宽现实情形进行探讨。

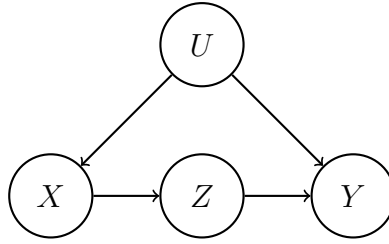


图 5: 前门准则

如图 5 所示，X 与 Y 之间有无法控制的混杂因子 U，但是 X 只通过 M 作用于 Y。此时就可以使用中介两步法研究因果关系

$$Z = \beta_1 X + u$$

$$Y = \alpha_1 M + \alpha_2 X + v$$

第二个式子需要控制 X 的原因是，如果我们单独考察 Y 与 Z 的关系，此时依旧存在 U 同时影响 Z 和 Y ，而此时控制了 X ，就能截断 U 到 Z 的路径，也就是后门控制。

此时 $\beta_1 * \alpha_1$ 可以看作 X 与 Y 的因果效应。但是这个应用条件极为苛刻，其需要 X 与 Y 之间只有一个机制变量 Z ，且在第二个回归式子中，需要控制 Z 与 Y 之间的所有后门路径。

5 过度控制、中介路径

5.1 可能过度控制的 DAG

X 与 Y 之间可能存在直接路径与间接路径，例如如下图 6 所示

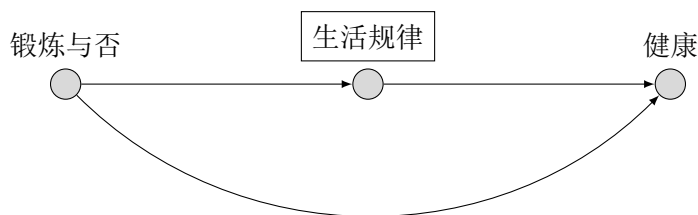


图 6: 过度控制例子 1

此时 X 到 Y 其实是有两条路径，因果效应有锻炼与体质的直接路径；锻炼与生活作息的间接路径，在直接路径和间接路径方向相同的情况下，如果控制了生活规律就会导致低估。

但是，如果符号相反，就存在遮蔽效应，且可能产生中介谬误。

5.2 遮蔽效应、中介谬误

最开始我们提到过，人们的一种错误认识是——认为因果需要在统计相关的前提下进一步排除，但由于遮蔽效应的存在，我们不能这样认为。

我个人赞同 zhao 等 (2010) 的总结。他们把中介效应分为互补中介、竞争性中介、只有间接中介的情况。

- 只有间接中介时，其实就是前门准则（章节 4.1）可以使用的场景。

- **互补中介时**，直接路径和间接路径方向一致，控制中介变量只是低估，在计量实操上反而更加严谨。不过关键在于，研究是要分析总体效应，还是间接机制，还是直接机制。
- **竞争性中介时**，直接和间接符号可能相反，此时会产生各种意外。

例如我们研究员工年龄可能有如下图7关系：

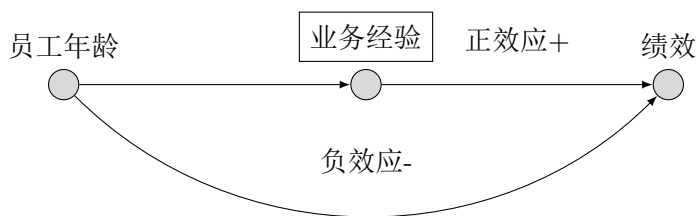


图 7: 遮蔽例子 1

若正负效应互相竞争，效应低消，那么从统计相关上看，员工年龄和绩效并不存在相关关系，也就是像忍者一样隐藏起来，是为遮蔽效应，这就是前文说的**相关性未必是因果的必要条件**。但如果控制了员工业务经验，可能会呈现负向相关关系。

而直接效应和间接效应的区分，经济上的解释也意义重大：

1975 年，学生起诉加州大学伯克利分校（UC Berkeley）研究生录取存在性别偏见。女生录取率显著低于男生。[Bickel 等 \(1975\)](#) 用数据分析了具体情况，简洁例子如下：

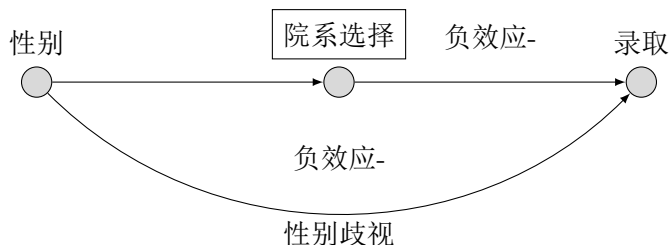


图 8: 遮蔽例子 2

在没有控制院系固定效应的情况下，女性录取率显著低于男性。但是一旦控制院系固定效应后，直接路径（性别歧视路径）不再显著。原因在于女性更倾向于申请难度较大，录取激烈的院系。

中介变量的探讨在实证故事中至关重要，因为最后我们需要通过某个机制尝试干预，一旦干预理论前提错误就可能造成严重后果。

📖 1911 年至 1912 年间人类历史上最著名的“南极极点冲刺赛”上演过一场悲剧。挪威探险家罗尔德·阿蒙森（Roald Amundsen）团队与英国探险家罗伯特·斯科特（Robert Falcon Scott）团队展开生死竞争。

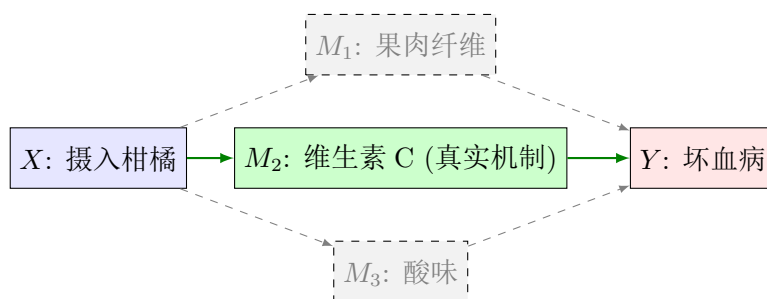


图 9: 坏血病治疗的经验观测

海上的水手常常产生坏血病，而经验证据表明橘子利于防止坏血病，但究竟是哪个中介变量发挥作用？人们不得而知。维生素 c 直到 1912 年底才被提出。

斯科特团队携带饮用橙汁，但维生素 C 很快被时间、光照、加工国产破坏，最终饱受坏血病折磨；阿蒙森团队向因纽特人学习了原住民的极地生存经验，坚持吃新鲜的半生肉，而其中包含人体需要的维生素 C，无意间契合了科学 🧐。

6 对撞路径、选择偏差

混淆变量可能使得两个变量产生虚假相关，需要加以控制；但是如果修改了混淆变量和 X,Y 的方向 (图 10)，此时如果控制了变量 U 才会产生虚假相关。

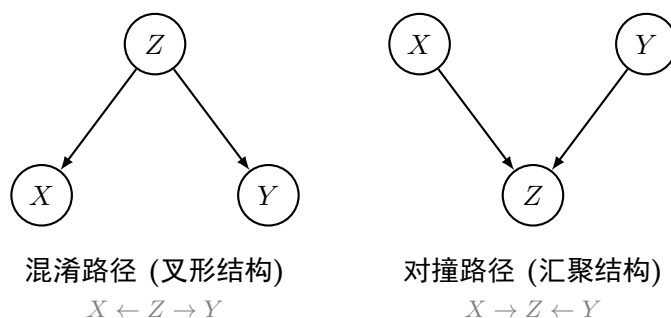


图 10: 混淆路径与对撞路径

6.1 幸存者偏差

经典例子就是幸存者偏差：我们能观察到的结果是可能经过筛选的样本。

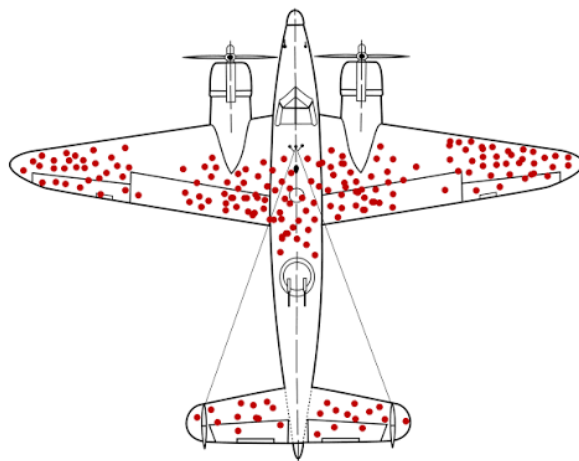


图 11: 幸存者偏差

“二战”期间，为了加强对战机的防护，英美军方调查了作战后幸存飞机上弹痕的分布，决定哪里弹痕多就加强哪里。然而统计学家沃德（Wald）⁶力排众议，指出更应该注意弹痕少的部位，因为这些部位受到重创的战机，

⁶Wald 检验量提出者。Wald 最终死于一场空难。1950 年 12 月 13 日，他和妻子受邀前往印度进行学术演讲。他们乘坐的印度航空（Air India）客机在印度南部的尼尔吉里丘陵（Nilgiri Mountains）坠毁，沃尔德夫妇双双在这次事故中遇难，终年 48 岁。

很难有机会返航，而这部分数据被忽略了👤。事实证明，沃德是正确的^[7]。

将幸存者偏差化为路径图就是一个对撞路径，由于只能观测到返航的飞机，但是中弹与非致命才能导致返航，因此导致虚假相关。

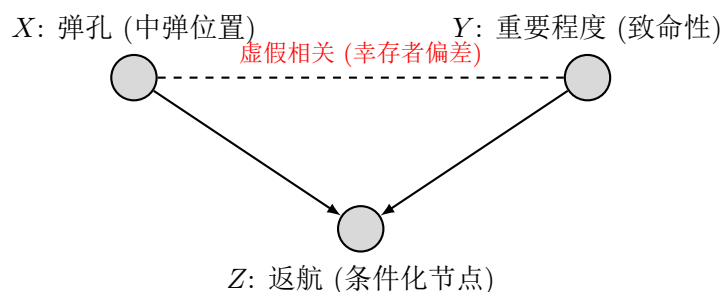


图 12: 沃尔德幸存者偏差

更贴合幸存者偏差的典型例子是下图^[13]。若中枪和中风者必死，那么我们观测活人样本时必发现，中风和中枪正相关。因为只有既没有中风，又没有中枪👤的人才能活着。

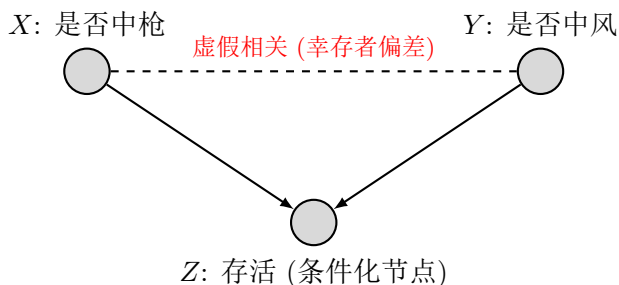


图 13: 幸存者偏差

拥有演技和长的漂亮的人都可能成为明星，因此通过对撞路径可能得到所谓流量明星的虚假相关——越漂亮，演技越差。

^[7]以上材料来自 2018 年高考语文（全国卷）的作文题目

7 小结

🕒 由于因果关系通常无法被直接观测到，我们只能通过变量间的相关性去推测因果关系，因此从路径的角度讲，因果关系分析的本质就是发现因果路径，截断混淆路径，避免对撞路径产生的衍生路径。

——邱嘉平《因果推断实用计量方法》

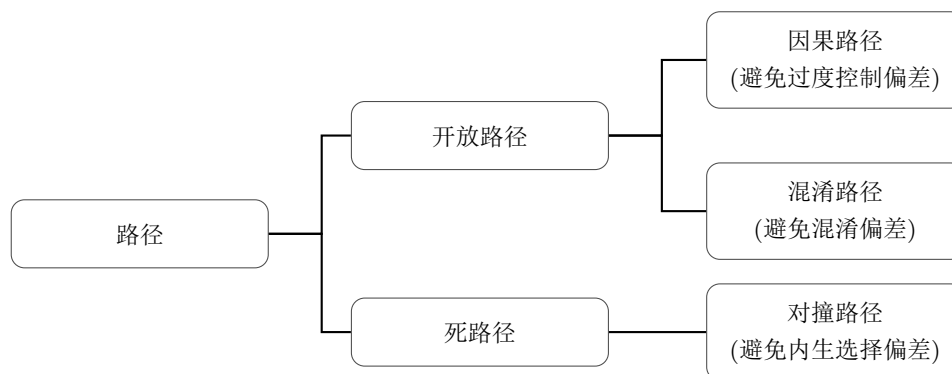


图 14: 小节

8 工具变量

在前门准则部分⁵，若我们保证 Z 完全中介唯一，且 X 控制后就是控制了所有的后门路径，那么可以使用前门准则分解 X 、 Y 的因果效应。此时方向是 X 指向 Z ；那么，在 Z 指向 X 的情况下，我们可以使用另外一个思路。事实上，Joshua Angrist 就是因为这个贡献与 David Card、Guido Imbens 一起获得了 2021 年的诺奖。

在计量中我们都知道，一个好的 IV 需要两个条件：

- 相关性： X 与 Z 密切相关，否则 IV 估计会偏大⁸
- 排他性： Z 不能与 Y 产生关系

⁸如何理解 IV 和 OLS？伍德里奇《计量经济学导论》曾强调过，IV 和 OLS 不能叫做不同的模型，只是不同的估计方法

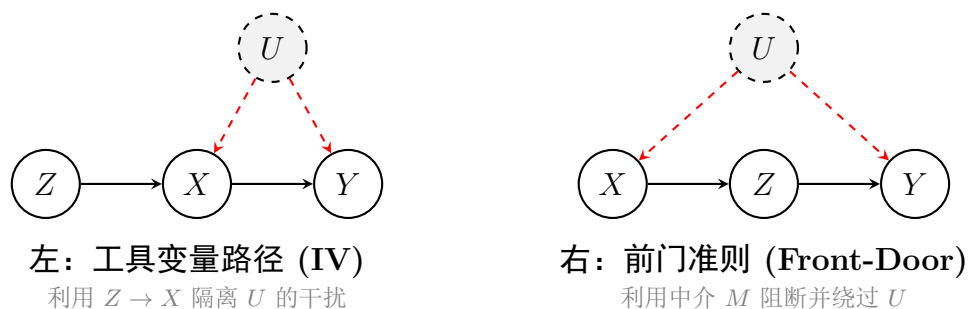


图 15: IV 与因果推断

在实践中，IV 的排他性和前门准则的完全唯一中介都是非常困难的事情，因为**唯一性**要求排除其他所有可能。

- 通过 DAG，我们可以看出， U, X, Z 在工具变量路径中是一个对撞路径，因此只要不直接控制 X ，就可以避免 U 和 Z 的虚假相关。
- 要求 Z 不能直接影响 Y 的原因，则是多中介时，我们需要区分 X 作为唯一中介的影响路径，而不是估计中，把 Z 对 Y 的直接路径也估计进入，导致高估。

9 前置和后置变量

参考资料：[例子](#)。

在选择模型形式时，关心的显然是可测量变量的组合。控制变量的好坏，可能存在遗漏变量和过度控制的问题。是否过度控制，取决于问题想研究整体效应还是具体的某个中介机制，此外，还有处置前后的变量问题。

9.1 处置后变量

此时变量的真实关系为：

$$Y = \alpha X + u_y$$

$$W = b_{yw}y + u_w$$

此时在模型中控制 W_i ：

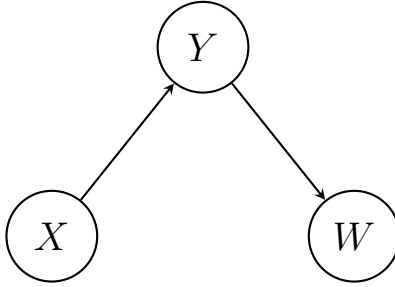


图 16: 处置后变量

$$Y_i = \beta X_i + \gamma W_i + u_i, \quad u_i = u_{yi} - \gamma W_i$$

代入整理可得：

$$Y_i = \frac{\beta}{1 - \delta b_{yw}} X_i + \frac{\delta}{1 - \delta b_{yw}} u_{wi} + \frac{1}{1 - \delta b_{yw}} u_i$$

也就是若控制了后置变量，系数会受到额外干扰。

这也是为什么最近很多论文，在控制 DID 事前趋势时，只使用基期或者基期前数据控制的原因，而不再和以前一样加入时间趋势、虚拟变量和控制变量的交互——因为一些变量可能是 Y 的处置后变量。

9.2 处置前变量

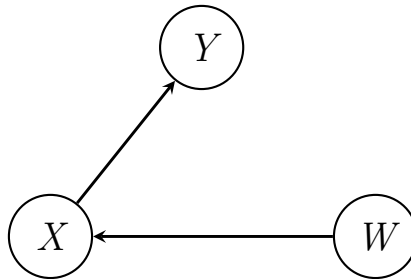


图 17: 处置前变量

由于 W 基本可以视为随机常数，因此加入控制不影响估计数值。

9.3 Y 的其他直接因素

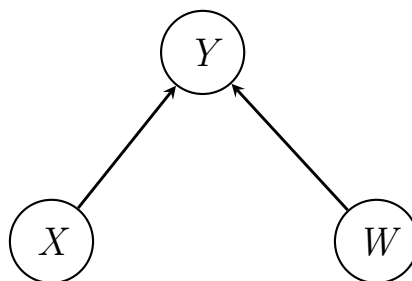


图 18: Y 的其他直接因素

由于 W 基本可以视为随机常数，因此加入控制不影响估计数值。