计量经济学因果识别方法详解笔记

Zircon

2026 年 5 月 26 日

目录

1 关于课程

1.1 总评构成

- 课堂表现：课堂提问、讨论和出勤，占 50%
- 课堂报告：汇报一个自己的创新性的研究题目、方法、数据，课堂汇报 15 分钟左右，占 50%
  - 需要想一个你认为有价值的题目，针对这个题目需要汇报研究的方法和数据；
  - 不需要真实做出结果，加入数据是为了说明研究的可行性；
  - 针对这个题目，需要使用哪些计量经济学方法来研究（使用 DiD, RDD, IV 中的一种或多种）
  - 可以有一些参考文献，但不需要太多
  - 最看重的是创新性以及对方法掌握的精确程度，以及是否能够找到相关的数据（利用现成数据库看是否有对应的列变量可以用于分析）

1.2 参考书籍

1. 因果推断实用计量方法：邱嘉平
2. Causality Models
3. Introduction to Econometrics
4. Econometric Analysis

1.3 内容提要

这是一门强调应用的课程，而不是讲述理论的课程。会重点结合实证的例子讲解。涉及的论文有百余篇。

1. 因果识别概述
   - 如果能够做到完全随机，那么 DID 就退化成 RCT。DID 和 RCT 的区别就在于“随机”。
   - RCT 是因果识别中的黄金标准，在方法上非常简单，随机之后就可以干预，之后收集数据即可，甚至只有一期就可以。
   - 但 RCT 的难处在于其可行性；同时，很多随机干预的问题在于强度不够，或者成本太高。医学、化学、生物学等实验学科中，RCT 应用广泛；在经济学中，通常并不能对经济实体做实验，因此 RCT 几乎不可能。
   - 随机分组要求两组的差别只在于分组，可观测变量在两组的平衡是最基本的要求，因此性别的平衡是最低的底线。在没有 RCT 的资源的情况下，才考虑本课程所讲的其他方法。在掌握这些方法后，RCT 也就很轻易了。

- 读文章时，判断文章好坏的标准并不是文章的复杂程度，而是文章的因果识别是否干净。

2. DID 专题
3. RD 专题
4. IV 专题

2 因果识别概述

2.1 辛普森悖论

现实生活中见到的多为相关关系，但真正值得关注的是因果关系。因果关系必然导致相关关系，但相关关系未必一定反映因果关系。在某些情况下用相关关系推导因果关系还会自相矛盾。

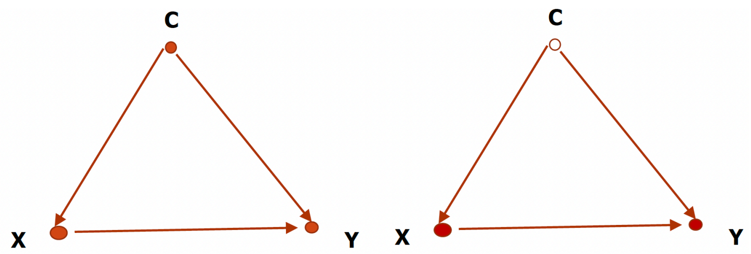
> 两个变量 X 和 Y 在每个分组中的关系是正（负），但在总体（所有组加总）中关系会发生逆转变成负（正）。

此外，符合时间的发生顺序的事件并不意味着因果关系，甚至可能不意味着相关关系（在宽的定义之下可以认为是相关关系）。

在做分析时，一定要注意遗漏变量问题，否则遗漏变量进入误差项会导致估计出的系数只是相关关系而不是因果关系，并且估计出的系数会有偏差。即便得到了看似满意的结果，也需要考虑控制其他变量，看结果是否可能变化。很多时候遗漏变量问题隐藏得很深，只有多尝试才能看出影响。当然，控制更多的变量不一定更好，可能导致的问题有多重共线性、自由度损失、中介变量。

2.2 变量关系图

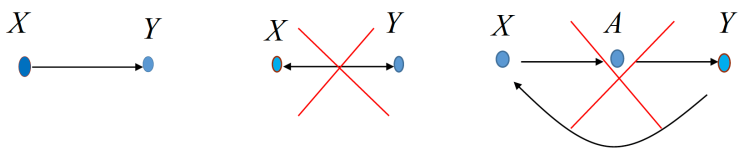
变量关系图也称为无环图（Direct Acyclic Graph），由节点和**单向**箭头组成，其中每个节点表示一个变量，用实心圆点表示观测得到的变量，用空心圆点表示观测不到的变量。当一个变量会通过因果影响另一变量时，画一条箭头线表示。因果关系是有方向且非循环的。路径图是一个有向无环图，因此不能用其去描述互为因果（simultaneous causation）和反馈循环（feedback loops）。



2.2.1 基本路径

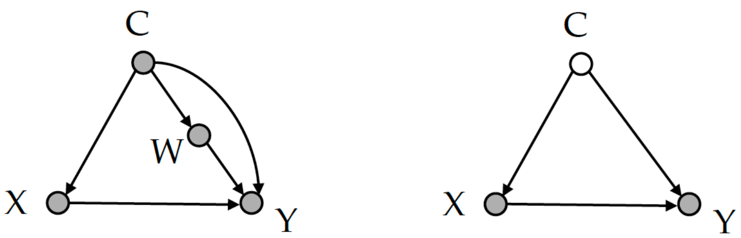
基本路径有因果路径、混淆路径和对撞路径。

因果路径 因果路径也称为链状路径 $A \to B \to C$。因果路径是一条连接两变量 X 和 Y 的所有箭头指向同一方向的路径，X 可以直接或间接地因果影响 Y。因果路径的特点是箭头指向同一方向。两个变量之间如果存在因果关系，它们就必然存在相关关系，所以因果路径为开放路径。



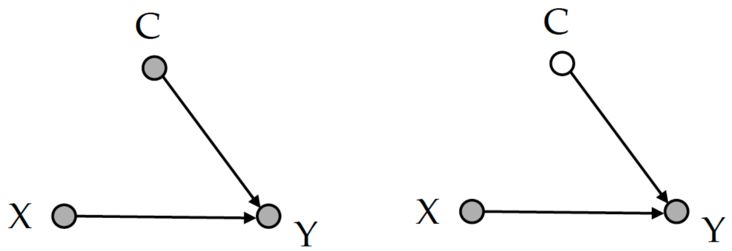
遗漏变量是不好的，过度控制也是不好的；遗漏变量必须控制，中介变量必须不控制。至于变量的归类于遗漏变量或中介变量，需要通过理论或者常识来判断。过度控制会让估计出的系数有衰减偏误，即便方向一般不变。

混淆路径 混淆路径也称为叉状路径 $A \leftarrow B \to C$，是指在解释变量 X 和被解释变量 Y 之间存在混淆变量的路径，混淆变量是同时影响 X 和 Y 的变量，混淆变量不一定直接影响 X 和 Y，也可以间接影响 X 和 Y。混淆变量的存在也会导致相关关系，所以混淆路径也是开放路径。



对撞路径 对撞路径也称为反交叉路径 $A \to B \leftarrow C$，是一条包含至少两个箭头指向同一个变量的路径。给定对撞变量以后（控制结果），两个（本不相关的）变量会产生相关性。对撞变量本身的存在并不会导致偏差，但给定对撞变量后会导致偏差，这个错误的相关性称为对撞误差；这正也解释了样本自选择偏误。

样本自选择模型的本质就是对撞路径。



2.3 因果关系估计误差来源

估计变量之间的因果关系的本质是**找到二者之间所有的因果路径**，同时去除二者间的非因果关系路径。

实际操作中会造成各种偏差，主要可以分为三类：混淆误差、过度控制误差、内生选择误差

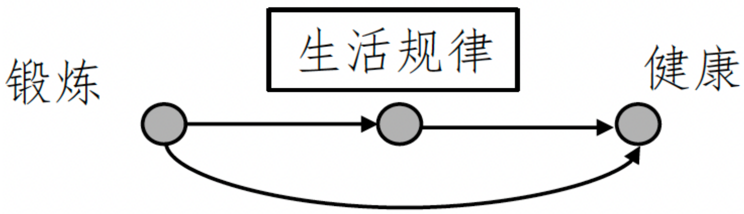
2.3.1 混淆误差

混淆误差是指在解释变量和被解释变量之间存在**未截断的混淆路径**，造成解释变量和被解释变量的相关性不仅包含因果关系，还包含非因果关系。截断混淆路径是通过给定混淆变量（conditional on confounding variable），从而排除混淆变量的干扰；给定混淆变量可以简单理解为固定混淆变量的值。在关系图中，加方框以表示该变量是给定的。当混淆变量给定时，解释变量和被解释变量的相关性就与混淆变量无关，二者的相关性就反映了因果关系。

常见的导致混淆误差的情形有模型误设、遗漏变量、测量误差。

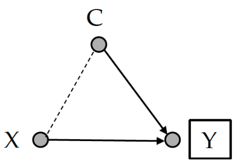
2.3.2 过度控制误差

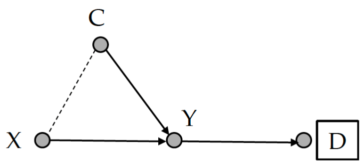
过度控制误差是指控制了因果路径上的变量（中介变量）而造成的误差。在研究中要避免控制受解释变量影响并且会影响被解释变量的中介变量，否则会导致过度控制偏差。



2.3.3 内生选择误差

内生选择误差可以理解为，当给定两个变量共同的被解释变量（对撞变量）时（或者对撞变量的被解释变量），也即对撞变量被家定位条件时，两个变量之间会产生一个衍生路径。衍生路径会造成两个原本不相关的变量变为相关，或造成两个原本相关的变量的相关性发生改变。

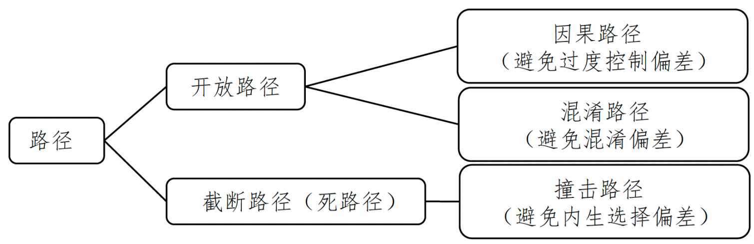




2.3.4 相关推测因果

由于因果关系通常无法被直接观测到，只能通过变量间的相关性去推测因果关系，因此从路径的角度上讲，分析因果关系的本质就是：

- 发现因果关系
  - 避免过度控制偏差
- 阶段混淆路径
  - 避免混淆偏差
- 避免对撞路径产生的衍生路径
  - 避免内生选择偏差



2.4 常用因果估计方法

在理想状态下，使用控制实验和准实验估计因果。而在实际研究中，我们通常面临观测数据，即数据产生不具备随机安排并且是个体自愿选择产生的。

1. 控制实验（随机分配）：通过控制实验达到随机化，使得处置组和控制组没有系统性差异，解释变量与任何其他可能的混淆变量都不相关，以达到估计因果效应的目的。
2. 准自然实验：自然或社会事件提供某种程度的随机化，事件的发生并不是个体自己所能选择的，加以正确的模型和假设可以估计因果效应。由于准自然实验法语控制实验不完全相似，因此对数据有一定的要求，并且要使用相应的估计方法。
   - 双重差分：平行趋势假设
   - 断点回归：局部连续性假设
3. 观测数据研究：在没有进行随机化时通过观测数据探究变量之间关系的研究
   - 回归方法
   - 匹配方法
   - 面板数据方法
   - 工具变量
   - 自选择模型

各种方法的区别与联系在于

- 准实验和随机控制实验的不同之处在于前者的干预行为不是随机的，故接受干预的处置组和未接受处置的控制组在干预以外的特征并完全不相同。（DID 中，随机性是不要求的，但外生性是要求的）

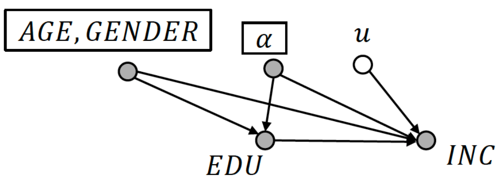
- 准实验和非实验的观测研究不同之处在于准随机试验的干预行为是外生的，它不受个体特征的影响，因此它不存在自选择问题。

- RDD 能解决 DID 和 IV 不能解决的因果识别。但 RDD 依赖于局部平衡假设，它的局限在于只能估计局部范围中的样本。DID 关心的是平均处理效应，对于全部样本都可以使用。当可以使用 DID 时，优先使用 DID

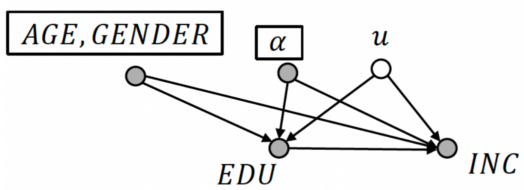
- 匹配方法直接把处理组和控制组在解释变量上进行匹配，然后比较二者在结果上的差异，避免了假设解释变量和被解释变量的方程关系。回归方法和匹配方法是类似的，只是在如何“控制”解释变量上有所区别；既然变量可观测，实际上直接控制即可，没必要匹配。回归法和匹配法的局限在于，它们都只能控制能观测到的因素，而没法去除无法观测的因素，无法解决遗漏变量问题。

- 面板数据回归可以加入固定效应模型，控制不随时间变化的可观测或不可观测因素。但仍然无法去除无法观测但随时间变化的因素。

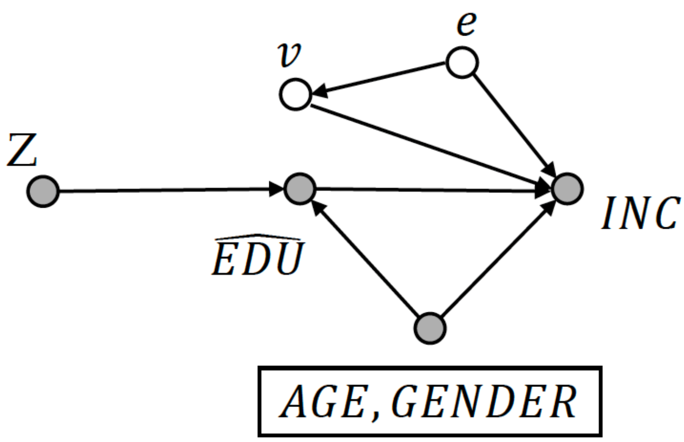
控制变量，个体 α_i，u 为无关变量



控制变量，个体 α_i，u 为混淆变量



- 工具变量的本质是通过其外生性（不受 u_{it} 干扰）将内生解释变量分解为内生部分和外生部分。

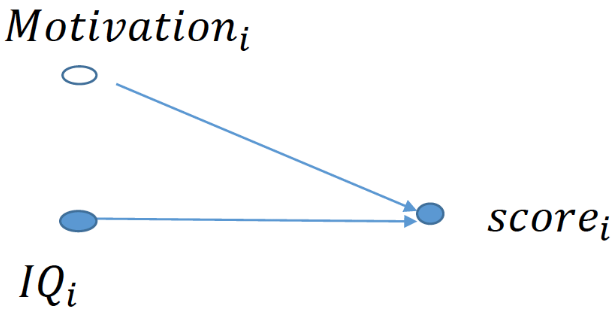


- Heckman 样本自选择模型。

  – 假设智商（IQ，可观测变量）和个人动力（motivation，不可观测变量）都会影响大学成绩（score，可观测），但 IQ 和 motivation 并没有相关性。

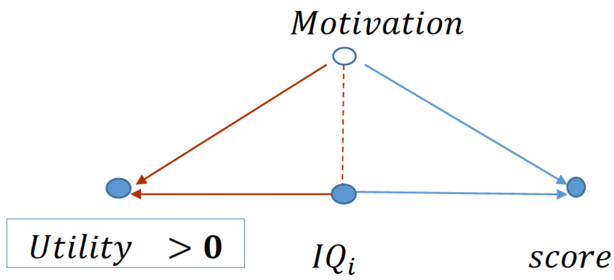
– 如果有总体数据或随机抽样数据，将 score 对 IQ 回归可以得到 IQ 对 score 的因果影响：

$$score_i = \alpha + \beta\, IQ_i + e_i$$



干扰项 e 包含了 motivation。由于 e 和 IQ 不相关，因此 β 的估计是正确的。

– 如果数据是自己选择上大学的大学生，假如上大学是由个人效用决定的，效用取决于 IQ 与 motivation。由于样本只包含了自己选择上大学的个体，假设只有当上大学的效用 $utility > 0$，个人才会选择上大学。使用自选择样本，相当于给定了对撞变量 $utility > 0$，产生了衍生路径。



在样本中 motivation 和 IQ 存在相关性，干扰项和 IQ 相关，回归 $score_i = \alpha + \beta\, IQ_i + e_i$ 得到的 IQ 系数将是不正确的。

- 最理想的方法是 RCT，其次是直接控制该变量，当变量无法被观察时，可以通过 DiD, RDD 以类似于 RCT 的方法控制。当无法观察，也无法控制时，只能寻找工具变量，工具变量是最终的方法。工具变量的假设容易被挑战，无法检验工具变量是外生的。最好寻找历史上形成的、不受现代经济社会因素影响的变量。或者寻找后期产生的，无法对前期造成影响的变量。对于工具变量，存在信与不信，可信度高于不高的区别。

  – 平行假设是可以检验的，通过事前的数据可以检验；DID 假设平行趋势会在事前和事后保持，在没有其他政策干预时。为了支持事后的平行，还要考虑加入其他控制变量。

  – 存在自选择问题时，只能使用工具变量方法。

方法	解决的因果关系估计中的偏差
简单回归法、匹配法	可观测因素造成的混淆偏差
面板数据分析法	可观测因素、不随时间变化的不可观测因素造成的混淆偏差
工具变量、DID、RDD	可观测因素、不可观测因素造成的混淆偏差

方法	解决的因果关系估计中的偏差
样本自选择模型	包含不可观测因素造成的内生选择性偏差

2.5 处置效应

2.5.1 潜在结果

如果个体 i 接受了某种处置行为（treatment）D_i，它的后果 Y_i 为 $Y_i(1)$。如果没有接受这种处置，它的后果为 $Y_i(0)$。这两种结果被称为潜在结果，表示为

$$\text{潜在结果} = \begin{cases} Y_i(0), \text{if } D_i = 0 \\ Y_i(1), \text{if } D_i = 1 \end{cases}$$

其中 $D_i = 0$ 表示个体 i 没有接受处置，$D_i = 1$ 表示个体 i 接受了处置。

处置行为 D_i 对个体 i 的处置效应（treatment effect）是个体 i 接受处置 $(D_i = 1)$ 的潜在结果和没接受处置 $(D_i = 0)$ 的潜在结果的差异。

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

处置效应 γ_i 即处置行为 D_i 对 Y_i 的因果效用（causal effect）。

然而实际上，个体只能接受或不接受处置，解决办法是为 $Y_i(0)$ 或 $Y_i(1)$ 找到一组替换。

2.5.2 平均处置效应

接受处置个体的平均值处置效应（average treatment effect on the treated, ATT）

$$ATT = E[Y_i(1) - Y_i(0)|D_i = 1] = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]$$

通常情况下，ATT 是我们最关注的结果，因为这是处置行为的直接后果。

未接受处置个体的平均处置效应（average treatment effect on the untreated, ATU）

$$ATU = E[Y_i(1) - Y_i(0)|D_i = 0] = E[Y_i(1)|D_i = 0] - E[Y_i(0)|D_i = 0]$$

ATU 衡量一项处置行为对未接受处置个体（如果他们接受了处置）的平均处置效应。

总体平均处置效应

$$ATE = E[Y_i(1) - Y_i(0)] = E[Y_i(1)] - E[Y_i(0)]$$

ATE 衡量一项处置行为对所有（包含已接受处置和未接受处置）个体的平均处置效应。ATE 是 ATT 和 ATU 的加权平均，权重是接受处置和未接受处置个体占总体数分别的比例。

对于个体 i，不可能同时观测得到两种潜在结果，因此无法直接计算处置效应，这是因果推断的根本难点。我们观测到的结果只是个体根据它的接受处置状态而显现出来的对应的潜在结果，称之为观测结果。观测结果可以表示为潜在结果和处置状态的函数：

$$Y_i = Y_i(0) + [Y_i(1) - Y_i(0)] \times D_i$$

所有的因果方法都只是为了还原对照组，RCT 要求随机分组，从而认为对照组和实验组在接受处置前相同；DID 要求平行趋势假设，即除了处置之外的因素对两组的影响平行；RDD 要求局部连续性假设，由此认为断点附近样本同质；工具变量用到外生性假设，背后即为还原外生分组。

2.5.3 反事实结果

对于处置组的个体，观测到了潜在的结果 $Y_1(1), \cdots, Y_k(1)$，但没有观测到潜在结果 $Y_1(0), \cdots, Y_k(0)$；对于控制组的个体，观测到了潜在的结果 $Y_1(0), \cdots, Y_k(0)$，但没有观测到潜在结果 $Y_1(1), \cdots, Y_k(1)$。我们把观测结果所对应的未观测到的潜在结果称为反事实结果（counterfactural outcome）。如何估计反事实结果是估计处置效应的关键。

2.6 回归分析

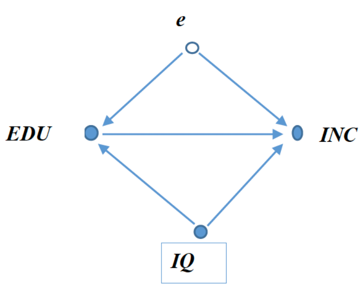
2.6.1 线性回归模型与因果推断

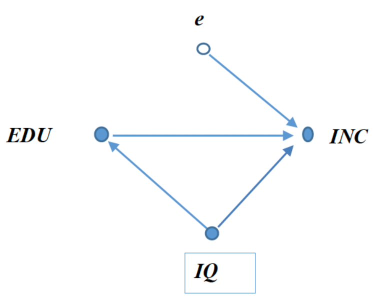
线性关系假设 假设要研究的问题是教育（EDU）对收入（INC）的因果影响，线性关系如下

$$INC = \alpha + \beta_1 EDU + \beta_2 IQ + e$$

其中 EDU 是希望估计的对 INC 因果影响程度的变量，也称为处置变量。处置变量不一定是哑变量，可以是任意关心的解释变量。对应地，IQ 在此处为控制变量。方程中右边的变量都被称为解释变量，左侧的为被解释变量。

常干扰项条件均值假设 因为干扰项是不可观测的，因此即使观测到了三个变量之间的变化关系（即三者相互的相关关系），也不能明确多少 INC 的变化是由于 EDU 变化造成的（即系数 β_1），因为它们之间变化的相关性可能是由于干扰项 e 造成的。





在可观测的混淆变量控制之后，为了保证干扰项和解释变量不存在混淆路径，需要要求干扰项条件均值独立于解释变量，

$$E(e|\boldsymbol{x}) = E(e) = c$$

因为线性回归模型包含了常数项，因此可以假设

$$E(e|\boldsymbol{x}) = E(e) = c$$

当线性回归不包括常数项时，干扰项条件独立于解释变量退化为常均值假设。

需要注意解释变量和误差项的相关的方向（相关是具有方向性的），从而确定误差项中的因素是否应该被控制。因此如果说解释变量和误差项有相关，那么不一定能推出 EDU 是内生变量。解释变量和误差项不相关是变量外生的充分不必要条件。

线性条件期望函数 一个能够识别出解释变量和被解释变量的因果关系的线性回归模型要满足如下两个假设

1. 线性关系假设：$INC = \alpha + \beta_1 EDU + \beta_2 IQ + e$
   - ”线性” 的含义是线性于参数，而不是线性于变量。
2. 干扰项条件均值为零假设：$E(e|EDU, IQ) = 0$

线性回归模型的两个假设对应的是线性条件期望函数（conditional expectation function, CEF）

$$E(INC|EDU, IQ) = \alpha + \beta_1 EDU + \beta_2 IQ$$

可以对 CEF 两端分别求偏导数得到解释变量的偏效应：

$$\frac{\partial E(INC|EDU, IQ)}{\partial EDU} = \beta_1; \frac{\partial E(INC|EDU, IQ)}{\partial IQ} = \beta_2$$

- β_1 的含义是，当 IQ 不变时，INC 的期望值（均值）随 EDU 如何变化
- β_2 的含义是，当 EDU 不变时，INC 的期望值（均值）随 IQ 如何变化

在 $E(INC|EDU, IQ) = 0$ 的假设之下，β_1 和 β_2 分别给出了 EDU 和 IQ 对 INC 的均值的因果影响。因此条件期望函数 $E(INC|EDU, IQ) = \alpha + \beta_1 EDU + \beta_2 IQ$ 称为因果关系 CEF 或因果关系回归函数。

遗漏变量 假设只观测到了 EDU 和 INC，并没有观测到 IQ，故只能把 IQ 作为干扰项的一部分

$$INC = \alpha + \beta_1 EDU + \varepsilon; \varepsilon = \beta_2 IQ + e$$

因为干扰项 $\varepsilon = \beta_2 IQ + e$ 包含了 IQ，并且 IQ 和 EDU 存在相关性，ε 和解释变量 EDU 也就产生了相关性，即零条件均值假设不再成立

$$E[\varepsilon|EDU_i] = E[\beta_2 IQ + e|EDU] = \beta_2 E[IQ|EDU] \neq 0$$

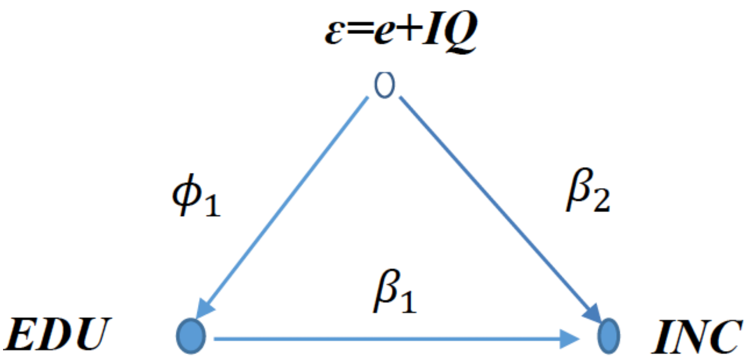
假设 INC 只对 EDU 回归的方程为 $INC = \gamma_0 + \gamma_1 EDU + u, E[u|EDU] = 0$，对应地，CEF 为

$$E[INC|EDU] = \gamma_0 + \gamma_1 EDU \frac{\mathrm{d}E[INC|EDU]}{\mathrm{d}EDU} = \gamma_1$$

γ_1 反映了 INC 的期望值如何随 EDU 变化，但并没有保持 IQ 不变。

对于真正的总体回归方程求导可以得到 γ_1 和 β_1 的关系，同时假设 EDU 和 IQ 有线性相关关系 $E[IQ|EDU] = \phi_0 + \phi_1 EDU$，

$$\frac{\mathrm{d}E[INC|EDU]}{\mathrm{d}EDU} = \beta_1 + \beta_2 \frac{\mathrm{d}E[IQ|EDU]}{\mathrm{d}EDU} \iff \gamma_1 = \beta_1 + \beta_2 \phi_1$$



由于系数 γ_1 反映的是教育和收入的相关性，并非纯粹的教育对收入的因果影响，因此说 $E[INC|EDU] = \gamma_0 + \gamma_1 EDU$ 是个相关关系 CEF 或相关关系回归方程。

2.6.2 最小二乘法

假设要估计的线性回归模型为

$$Y = \boldsymbol{X}'\boldsymbol{\beta} + \varepsilon\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)' E(Y|\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{\beta}$$

使用总体最小二乘法求解系数 $\hat{\boldsymbol{\beta}}$，就是最小化被解释变量值 Y 和解释变量线性预测值 $\hat{Y} = \boldsymbol{X}'\boldsymbol{\beta}$ 的残差 $\hat{\varepsilon} = Y - \hat{Y}$ 的平方的期望值，即求解

$$\hat{\boldsymbol{\beta}} = \text{argmin}_b E(Y - \boldsymbol{X}'\boldsymbol{\beta})^2 \Longrightarrow E(\boldsymbol{X}(Y - \boldsymbol{X}'\hat{\boldsymbol{\beta}})) = \mathbf{0} \iff E[\boldsymbol{X}\hat{\varepsilon}] = \mathbf{0}$$

可见最小二乘法的本质是求解系数 $\hat{\boldsymbol{\beta}}$ 使得解释变量 $\boldsymbol{X}$ 和残差 $\hat{\varepsilon}$ 不相关。

如果求解的条件期望函数是因果关系 CEF，最小二乘法求得的系数 $\hat{\boldsymbol{\beta}}$ 就具备因果关系解释。如果求解的条件期望函数是相关关系 CEF，最小二乘法得到的系数 $\hat{\boldsymbol{\beta}}$ 就只具备相关关系解释。

干扰项包含了除解释变量之外的其他影响被解释变量的因素。干扰项是无法观测到的，因此干扰项和解释变量是否相关是无法检验的，这只能依靠经济理论和实证经验去判断。残差是从最小二乘法计算出来的，并且根据算法总是和解释变量不相关，$E(\boldsymbol{X}(Y - \boldsymbol{X}'\hat{\boldsymbol{\beta}})) = \mathbf{0} \iff E[\boldsymbol{X}\hat{\varepsilon}] = \mathbf{0}$。但残差和解释变量不相关并不代表着模型中的干扰项和解释变量不相关；只有当干扰项和解释变量不相关时，最小二乘法得到的残差才是干扰项的正确估计，因此系数也是正确的估计。

2.7 回归模型内生性来源

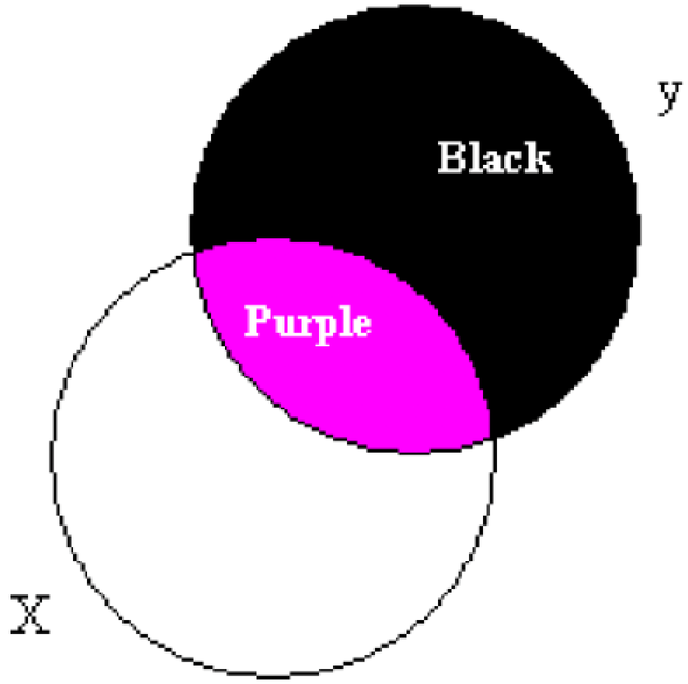
造成干扰项和解释变量相关，即 $E(e,[X]k) \neq 0$ 的原因有三种情况：

- 遗漏变量
  - 遗漏变量和解释变量的相关性导致了包含遗漏变量的干扰项和解释变量相关
- 解释变量的测量误差
  - 使用有测量误差的解释变量造成了干扰项里包含测量误差，进而导致干扰项和有测量误差的解释变量相关
- 互为因果
  -
$$\begin{cases} y_1 = \beta_1 x_1 + \phi_1 y_2 + e_1 \\ y_2 = \beta_2 x_2 + \phi_2 y_1 + e_2 \\ Cov(x_i, e_i) = 0, \forall i = 1, 2 \\ Cov(e_1, e_2) = 0 \end{cases} \Longrightarrow y_1 = \frac{\beta_1}{1 - \phi_1\phi_2} x_1 + \frac{\beta_2\phi_1}{1 - \phi_1\phi_2} x_2 + \frac{e_1}{1 - \phi_1\phi_2} + \frac{e_2\phi_1}{1 - \phi_1\phi_2} \Longrightarrow Cov(y_1, e_2$$
  - 当干扰项 e_1 发生变化，被解释变量 y_1 变化；由于存在逆向因果，y_1 成了解释变量，其变化造成了被解释变量 y_2 的变化，进而导致干扰项 e_1 的变化和解释变量 y_2 的变化形成了相关性。

2.8 韦恩图

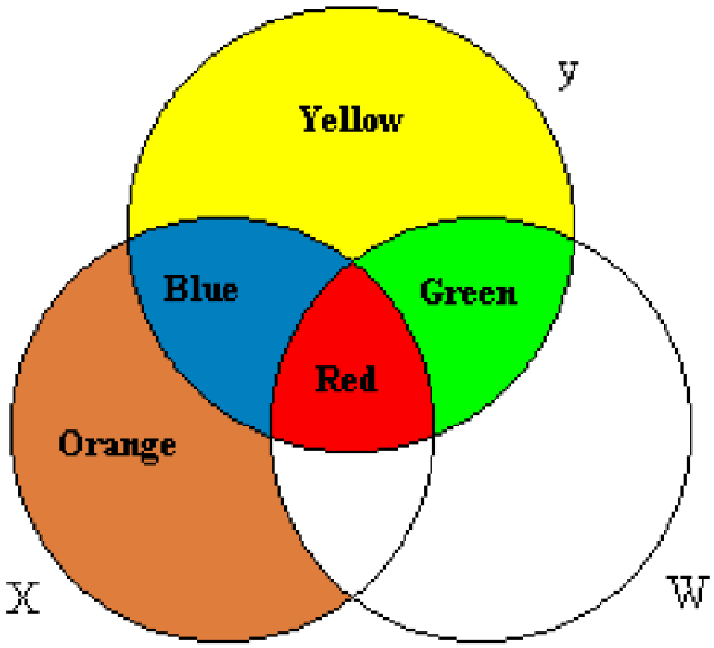
韦恩图可以拓展到经典线性回归模型的方差和偏差的阐述。若线性回归模型表示为

$$y = Xb + \varepsilon$$



两个圆分别表示两个变量的全部变异。两圆的重叠部分表示两个变量共有的变异，在这个区域内，两个变量共同变动。OLS 方法利用相关部分的共同变动来估计 b_x，即 X 的斜率系数。更大的紫色区域意味着在估计中使用了更多的信息，这意味着 b_x 估计值的方差较小。黑色区域代表不能由 X 解释的 y 的变异，因此该区域的大小代表 OLS 对误差项方差的估计值的大小。

考虑多元线性回归 $y = b_x x + b_w w + e$，



X 和 w 作为对 y 的解释变量，与 y 交叠部分记为解释变量能参与解释的 y 的变异性，它与 y 圆比值即为回归的 R^2。蓝色区域被用于估计 x 对 y 的影响，绿色区域被用于 w 对 y 的影响；三重交叠区域不参与估计 x 和 w 的影响，否则因果识别不干净。

红色区域越大，蓝色区域越小，估计的系数的标准误越大，这代表着更少的信息用于估计偏效应。当红色区域较大时，就会存在较为严重的多重共线性，但这时系数的估计仍然是无偏的，R^2 也不会变化。不过系数估计的准确度需要警惕，巨大的标准误让无偏的系数估计在本质上失去了识别的意义；这也启发我们要结合系数和标准误来看 R^2，有可能存在变量之间强相关而不能估计系数的问题。同时，多重共线性在计量上并没有很好的解决办法。

如果遗漏变量，那么估计系数的标准误会变小，因为估计时用到了更多信息，即便这些信息是不准确的。遗漏变量后带来的显著，实际上是利用了遗漏变量的无效信息。

3 DID

3.1 基本 DID 模型

	处理前	处理后	变化
处理组	Y_{11}	Y_{12}	$\Delta Y_T = (Y_{12} - Y_{11})$
控制组	Y_{21}	Y_{22}	$\Delta Y_C = (Y_{22} - Y_{11})$
DID			$DID = \Delta Y_T - \Delta Y_C$

回归形式为

$$Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 (Treat_s \times Post_t) + \varepsilon_{st}$$

- β_0: 在处理前对照组的均值
- β_1: 在处理前，处理组和对照组的差异
- β_2: 对照组在处理前后的差异
- β_3: 处理后，处理组与平行趋势的差异

处理前，两条线的平行趋势关系是可以检验的，而处理后两条线是否平行无法检验。因果识别的最大问题是无法观测到处理组的反事实结果，DID 正是在无法进行随机实验的现实条件下作为一种补救方法，从而模拟出识别因果关系中的反事实结果以分析处置效应。无法进行随机试验的最重要的原因有

- 样本太小
  - 样本太小以致于两组的各个特征并不平衡
- 成本过高
  - 分组会造成切实的政治经济社会影响

现实生活中，DID 适用于大部分的外生政策，也适用于自然灾害。

3.1.1 拓展到多年数据

$$Y_{st} = T_{st}\beta + \theta_s + \lambda_t + \varepsilon_{st}$$

- T_{st}: 如果 s 是处理组且 t 为处理后，那么 $T_{st} = 1$，等价于 $Treat_s \times Post_t$
- θ_s: 组固定效应，等价于 $Treat_s$
- λ_t: 年固定效应，等价于 $Post_t$

此时模型不包含截距项，因为固定效应可以完全吸收截距。

一个常见的数据集的存储形式如下所示：

		θ_1	θ_2	λ_1	λ_2
t_1	1	1	0	1	0
t_2	1	1	0	0	1
t_3	1	1	0	0	0
t_1	2	0	1	1	0
t_2	2	0	1	0	1

		θ_1	θ_2	λ_1	λ_2
t_3	2	0	1	0	0
t_1	3	0	0	1	0
t_2	3	0	0	0	1
t_3	3	0	0	0	0

做实证分析时，尽量**将固定效应控制到尽量低的层级**。

3.1.2 拓展到多处理期

$$Y_{st} = \beta(Treat_s \times Post_{st}) + \theta_s + \lambda_t + \varepsilon_{st}$$

- $Post_{st}$: 当 s 组在 t 年被干预时初次取值为 1，干预之后的时期取值也为 1；不同组的干预开始年份可以不同。

任何时候做事件研究（event study）都需要纯对照组，否则容易存在伪平行趋势。

3.1.3 拓展到不同处理强度

$$Y_{st} = \beta(Intensity_s \times Post_t) + \theta_s + \lambda_t + \varepsilon_{st}$$

- $Intensity_s$: 连续变量，每组的取值可能不同
  - 需要注意的是，**处理强度可能是内生的**。这时就需要为处理强度找一个工具变量，它只需要和政策不相关，和处理强度相关即可；由于政策在一定程度上是外生的，因此 DID 中的工具变量更为有效。

3.1.4 DID 处理遗漏变量问题

DID 的优势在于可以控制很多干扰因素，即很好解决遗漏变量问题。DID 能**控制不随时间变化的、随时间共同变化的遗漏变量**，这正是 DID 有效的主要原因。

	处理前	处理后	变化
处理组	$Y_{11} = \theta_1 + \lambda_1 + \varepsilon_{11}$	$Y_{12} = \beta + \theta_1 + \lambda_2 + \varepsilon_{12}$	$\Delta Y_T = (Y_{12} - Y_{11})$
控制组	$Y_{21} = \theta_2 + \lambda_1 + \varepsilon_{21}$	$Y_{22} = \theta_2 + \lambda_2 + \varepsilon_{22}$	$\Delta Y_C = (Y_{22} - Y_{11})$
DID			$DID = \Delta Y_T - \Delta Y_C$

DID 更多的优势还体现在：

- 处理的灵活性
  - 允许不同的开始时间
  - 允许处理组停止处理
  - 处理强度可以是连续变量
- 时间维度不是必须的
  - 通常模型有时间和分组两个维度，但时间维度也可以被替换，比如
    * 产品种类 × 地区
    * 年龄（性别）× 地理
  - 关键是可以通过双重查分排除遗漏变量
- 假设
  - 如果没有处理，处理组和对照组的趋势相同

3.2 稳健性检验

3.2.1 控制随时间变化的相关变量

- 可以得到更小的标准误，对系数的估计更准确
  - 前提是**控制的不是中介变量**
- 更宽松的假设：条件独立
  - 原先需要外生和平行两个假设，“独立”是指解释变量和误差项无关
  - 做 DID 时不需要担心控制的变量可能为控制变量，更关注的是“外生”假设

3.2.2 允许影响随时间变化

在模型中加入滞后项

$$Y_{st} = T_{st}\beta_0 + T_{s,t-1}\beta_1 + T_{s,t-2}\beta_2 + \cdots + T_{s,t-k}\beta_k + \theta_s + \lambda_t + \varepsilon_{st}$$

该模型的系数反映的是**过去的政策对现在的影响**。

3.2.3 处理前的趋势

加入提前项

$$Y_{st} = T_{st}\beta_0 + T_{s,t+1}\beta_1 + T_{s,t+2}\beta_2 + \cdots + T_{s,t+k}\beta_k + \theta_s + \lambda_t + \varepsilon_{st}$$

该模型的系数反映的是将来的政策对现在的影响。理论上，**将来的政策对现在没有影响**，故这可以用于检验平行趋势；**如果回归的结果显示将来的政策对现在有影响，说明平行趋势假设被违背。**

通常为了检验平行趋势，需要政策前多年的数据；**平行趋势检验还不仅仅看差异是否显著，还要看系数本身的大小。**

3.2.4 更多注意点

- Ashenfelter's Dip
  - 如果政策的宣传本身就有效，那么应该把政策的推行期往前提。实际上是一个自选择的过程。
- 标准误的处理
  - 遗漏变量，序列和空间的自相关都可能导致标准误较小。常用的方法是对标准误进行聚类。
  - Cluster Standard Errors
    * 当 cluster 小于 40 时，需要采用其它方法处理
    * 当 cluster 在 40 至 50 时，有些人仍会怀疑结果的可靠性
    * 当 cluster 大于 100 时，不需要特别处理
  - 其他标准误处理方法
    * Cluster Bootstrap
- Permutation Tests
  - 当估计出一个系数时，不知道系数是由于处理产生的，还是小的随机概率事件导致的。

3.3 三重差分

3.3.1 三重查分的原因

- 双重查分假设下，如果没有干预，实验组与对照组的时间趋势一样。然而，即使干预发生之前的两组时间序列一致，也不能保证干预发生后两组时间序列是一致的

  - 平行趋势很重要，但双重差分中只能检验事前的平行趋势；如果存在额外的备选组，那么可以用于检验事后的平行趋势。
  - 能做三重差分时，结果总会比双重差分的更为可靠。

- 有可能在干预发生的同时在实验组或者对照组中又发生了影响因变量的事件，则干预发生后两组的时间趋势是不一致的

3.3.2 三重差分的运用情景

基本的 DDD 模型（$T, B, Post$ 为虚拟变量）

$$
\begin{aligned}
y_{it} =& \beta_0 + \beta_1 T_i + \beta_2 B_i + \beta_3 Post_t \\
& + \beta_4 T_i \times B_i + \beta_5 B_i \times Post_t + \beta_6 T_i \times Post_t \\
& + \beta_7 T_i \times B_i \times Post_t + \varepsilon_{it}
\end{aligned}
$$

如果不添加其他控制变量，那么交叉项的系数通常更显著，甚至方向会扭转，这本质上是遗漏变量偏误。

三重差分使用得并不多，因为较为复杂且仍需要平行趋势假设。**一般在双重差分的平行趋势失败之后，才转而使用三重差分**。在汇报三种差分的文章中，一定会汇报双重查分的结果，并且比较双重查分和三重差分的结果。如果结果差异不大，说明结果较为可靠；**如果结果差异较大，通常认为三重差分的结果更可靠，因为控制了更多因素**。

三重差分估计量可以计算为两个 DID 的差分，但 DDD 并不需要两个平行趋势的假设。**只要两个估计量中的偏差相同，两个有偏差的 DID 估计量之间的差异就是无偏的**。实际上，减去第二个 DID 估计量的唯一目的是消除第一个估计量中的偏差。

3.3.3 用于识别影响机制

中介变量解释了两个变量之间的关系过程，而调节变量影响了这种关系的强度和方向。

假设研究地铁修建对于环境污染的影响，一个城市是否修建地铁标记为 $Treats_s$，修建地铁后标记为 $Post_t$，那么 DID 模型为

$$
Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 (Treat_s \times Post_t) + \varepsilon_{st}
$$

如果想要探究修建地铁是否是通过影响汽车的保有量（Car_{st}）来影响污染，则拓展到 DDD 模型

$$
\begin{aligned}
Y_{st} =& \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 Car_{st} \\
& + \beta_4 (Treat_s \times Post_t) + \beta_5 (Treat \times Car_{st}) + \beta_6 (Post_t \times Car_{st}) \\
& + \beta_7 (Treat_s \times Post_t \times Car_{st}) + \varepsilon_{st}
\end{aligned}
$$

本质上不是三重差分，模型中不要求汽车的保有量在组间时平行的。

3.4 匹配 DID

3.4.1 基本逻辑与步骤

- 基本逻辑
  - 由于处理组和对照组并不是随机分配的，DID 可能不满足平行趋势假设
  - 找出处理组分配的规则，那么可以构造一个系数来表示每个样本进入对照组和处理组的概率
  - 当作者只保留等概率的处理组和对照组样本时，平行趋势容易得到满足
    * **这样的匹配只能基于可观测因素，但不可观测因素仍然可能是不平行的。**
- 具体步骤
  - 基于 Logit 模型估计各个因素对于进入处理组的概率的影响
  - 基于每个（显著的）变量的估计系数，为每个观测值构造一个得分（score）
  - 基于得分进行一对一（或其他类型：分层匹配、一对多、和匹配等）的匹配
  - 排除未匹配上的数据，只用匹配上的数据跑 DID 回归
  - 在匹配之后，要进行 balance tests

当处理前只有一年的数据时，无法进行平行趋势检验；这时只能进行平衡检验，即检验处理前两组的各个变量是否相似。做平衡检验时，值得关心的是**标准化的均值差异**，而不是绝对数值意义上的差异，

$$d = \frac{\bar{x}_{treatment} - \bar{x}_{control}}{\sqrt{\frac{s^2_{treatment} + s^2_{control}}{2}}}$$

1. 匹配往往不是最好的选择，如果初始平行趋势显著被违反，匹配后满足（特别是基于一对多的匹配），结果的可靠性仍然会受到质疑
2. 匹配应遵循 1:10 原则：一个变量对应 10 个处理样本；如果处理样本较少，可以用的匹配变量就少
3. DID 实践中如何选择匹配变量是关键
   1. 应**依据政策出台的细则来选定匹配变量，这些变量必须是与政策选择相关的**，可以是随时间变化的或者不随时间变化的变量（但其影响可能随时间变化）
   2. **如果该变量可能受到政策的影响，则匹配应该基于事件发生前的样本**

4 RDD

RDD 的优势在于其因果识别假设简单且容易满足（即**前定变量在断点处局部连续**，可完全被检验），通常认为 **RDD 的因果识别效果好于 DID**；其劣势在于只能估计**局部**平均处理效应（断点之外的有效性有待考证；同样地，**RCT 的局限性在于其效果只在样本内有效**，拓展性有限）。RDD 可以解决看似不可能的因果识别问题，例如评估在**全国同一时期**统一执行的政策（其缺乏地区间执行强度和执行时间的差异）的因果影响。

RDD 的基础回归很简单，但近年来该方法得到了大量拓展，包含许多前沿知识与技巧，例如非参数估计、最优带宽的选择、DID-RDD、Kink-RDD 等模型拓展。RDD 只要求处理组和对照组在断点附近是平行的，即使事前趋势不同，也可以用 DID 配合 RD 的方式进行分析。

4.1 RDD 简介

断点回归（RD）设计，作为在**随机分配处理不可用**的情况下的检验因果假设的方法。从 RD 设计中得出的因果推断可能比典型的“自然实验”（即指 DID）策略更为可信。

- 基本特征
  - **存在一个连续分布的变量（通常成为“运行变量”），该变量包含一个明确定义的截止值，该截止值影响数据中哪些样本被分配给处理组以及哪些样本被分配给控制组**。
  - RDD 需要验证其他变量在断点附近是否连续，如果连续，那么认为我们关心的因变量的跳跃正是由于断点引起的；当然，为了排除其他随机事件或者并行事件产生混淆，**选择的断点需要比较精细**。
  - 空间上的距离需要用二维坐标来定义，但这样会比较麻烦，比较简单的方式是对空间划区。
- 基本逻辑
  - **接受处理的概率不连续变化**：用于了解处理对得分处于或接近临界值的样本的结果的因果影响
  - 得分仅低于临界值的样本可以用作得分勉强高于它的样本的比较组（或控制组）。
- 关键假设
  - RDD 的唯一假设是**潜在结果在断点附近连续变化**：运行变量从略低于阈值到略高于阈值，潜在结果会连续变化。
    - 即**在断点处，除了处置，没有任何其他因素会影响结果的跳跃**。
    - 另一种理解是，在阈值附近，可以认为**将样本随机分配给对照组和处理组**，类似于随机实验中发生的情况。

经典的例子即为高考，高考分数线在一本线附近的考生被认为相近，可认为在分数上下随机分配，可以由此估计上一本的影响。如果单纯分析上一本对考生未来的影响，上一本是一个自选择结果，会有严重的内生性问题，无法准确识别因果。而上一本的因果效应只有 RDD 能很好地识别，并且结果很干净。同时，RDD 完全可以使用追溯的数据。

- RDD 的优势
  - 与其他非实验方法所需的假设相比，RD 设计只需要看似温和的假设
  - 可以特别**透明地测试识别假设**，并且可以通过数据的信息图明显支持论点
- RDD 的最大挑战
  - 个体可能能够战略性地和**精确**地改变他们的得分以分配到他们首选的处理条件。
    - 这可能会导致他们在截止点或附近的可观察和/或不可观察特征不发生连续变化，从而混淆因果结论。
    - 注意此处强调的是“**精确**”。
      - 这是**对内生性的极大放松**：可以是部分内生的，只要不是“纯内生”即可。相比之下，工具变量要求绝对外生。
      - “精确”控制的例子：失业补贴。至于是否归为“精确”，需要基于具体数据和案例具体思考。
    - 自选择：由于存在精确自选择，导致观测结果在截止值附近非常接近，从而带来 RDD 估计偏差
      - 现实中往往直接不能判断自选择的程度和类型，需要进行检验
    - **Donut-RD**: 丢弃与截止值接近的样本重新进行估计，从一部分排除断点附近自选择的样本可能带来的偏差
- RDD 的缺陷
  - 潜在的低外部效度：该因果识别是高度局部的
    - 然而，存在一系列技术方案解决这一问题

4.2 RDD 模型

4.2.1 局部平均处理效应

假设作者对处理变量 D（一个虚拟变量）对结果变量 Y 的影响感兴趣，定义平均值处理效应

$$\tau \equiv E[Y_1 - Y_0]$$

对受处理组的平均处理效应（average treatment effect on the treated, ATT）为

$$ATT \equiv E[Y_1 - Y_0|D = 1]$$

对对照组的平均处理效应（average treatment effect on the control, ATC）为

$$ATC \equiv E[Y_1 - Y_0|D = 0]$$

对照组的平均处理效应也称为未处理组的平均处理效应（average treatment effect on the untreated, ATU）。

如果自变量是 X，截断点为 c，则 RDD 估计定义为

$$\tau_{RDD} = E[Y_1 - Y_0|X = c = E[]$$

实际上，$X = c$ 是断点，对应的结果是无法观测的，因此只能用极限的定义来补充。在实践中，作者可以努力实现的是

$$\tau_{RDD} = \lim_{\Delta \to 0}\{E[Y_i|c < x_i < c + \Delta] - E[Y_i|c - \Delta < x_i < c]\}$$

这意味着当 $\Delta \to 0$ 时，潜在结果的连续性时作者能够构建缺失的凡是视情况。实际上，这个方程估计的是局部效应，因为它只适用于哪些处于阈值附近的人，这也是 RD 或得到一个成为局部平均处理效果（LATE）的量。

4.2.2 参数估计

估计（标准化）运行变量收敛到 0 时函数的右极限和左极限的差异:

$$Y_i = \alpha + \beta(X_i - c) + \tau D_i + v_i$$

- X: 运行变量；在空间中，标准化为 0
  - 实际在运行中，为使运行变量中心化的 c 可以省略，没有实际影响。
- D: **表征处理状态的二值变量**
- α: 在断点处且不接受处理的均值
- $\alpha + \tau$: 在断点处且接受处理的均值
- τ: **在断点处预期结果的差异**

以上模型假设 X 和 Y 的关系可以用相同的函数形式来描述，无论在阈值以上还是以下。实际上，可以放松这一假设而估计模型

$$Y_i = \alpha + \beta(X_i - c) + \tau D_i + \gamma D(X_i - c) + v_i$$

这样**允许截断值上下的两个样本具有不同的斜率**（处置前，β；处置后，$\beta+\gamma$）。当然，这个模型（即交叉项）并不是必须的，我们关心的是 τ。**如果遗漏了交叉项，即便会有遗漏变量偏误，但由于我们关心小范围的局部影响，因此断点处影响的估计不受影响**。

此外，这里我们假设 X 对 Y 的影响是线性的，实际上断点前后影响的形式可能有多种。更通用的形式将是

$$Y_i = \alpha + \beta f(X_i - c) + \tau D_i + v_i$$

然而，RD 在异质性分析中会遇到困难。如果要分析估计的效果是否因其他变量不同而不同，通常做法是将处理指标与一个或几个组指标进行交互。这个简单的逻辑在 RDD 中不成立。

4.2.3 非参数估计

- 不适用多项式逼近非线性关系，而是通过专注于靠近截断点的区域来消除这种非线性关系的可能性
- 观察值通常按照它们与截断点 c 的距离进行加权，类似于核函数的方式
- 通常建议使用三角核，它会给靠近阈值的观察赋予更大的权重
  - 可以使用不同方法进行稳健性检验

决定文章质量的是选题和立意，以及因果识别的方法和有效性。各种稳健性检验只是丰富研究和增强文章的可靠程度，是相对次要的。

4.2.4 模糊 RDD

在模糊 RDD 中，**截断点只以概率方式指定处理状态，是否进入政策不是纯外生的**。幸运的是，断点回归不要求绝对的外生。**截断点 c 现在成为实际处理状态的工具变量**：越过它会改变接受处理的概率，但不再决定处理状态。

所谓的模糊 RDD 是将锐利 RDD 和 IV 逻辑结合起来进行估计，其估计方式如下

$$D_i = \alpha_1 + \beta_1(X_i - c) + \delta_1 Z_i + u_i$$
$$Y_i = \alpha_2 + \beta_2(X_i - c) + \delta_2 \hat{D}_i + v_i$$

其中 D_i 表示接受处理的处理状态，Z_i 是在截断点 c 上方的样本的虚拟变量。模糊 RDD 和锐利 RDD 的区别体现为，D_i 和 Z_i 的取值并不是一一对应的。

D_i 作为 Z_i 的工具变量还需要两个假设：

- 相关性：$\delta_1 \neq 0$
- 排除性：越过阈值影响结果的唯一方式是通过处理状态的概率变化。换句话说，其他可能影响结果的变量可以假定在阈值周围连续（即它们不应该跳跃）。

估计的效应是 $\dfrac{\delta_2}{\delta_1}$。

4.2.5 阈值的内生性

RDD 的最基本假设是**阈值是外生的**。为了探究阈值的外生性，有两种常见策略

1. 将结果变量替换为处理前的协变量，并查看这些**潜在混淆变量中是否已经存在不连续性**
2. 使用 Canay & Kamat 的检验方法

内生性检验和连续性检验在 RDD 中是相同的。如果发现阈值是内生的，则无法通过包括"控制"变量等方法解决。

4.3 断点差分法

4.3.1 基本逻辑

当使用**面板数据**时，断点差分法（DID-RD 设计）可用于评估政策，该方法同 DID 非常相似。

- 除了**连续性假设**之外，主要的额外假设是 **DID 的平行趋势假设**
  - 如果无论是否有处理，在断点附近都有跳跃，因果效应仍然可以识别，基于平行趋势假设估计。
    - 研究表明，**平行趋势假设在边界附近成立**，但在使用距离太远的数据时不成立。
  - **如果有多期数据，平行趋势假设可以得到检验**，即检验前期的跳跃是否相等。
  - 这样的好处是即使当年的数据不满足连续性假设，只要获取此前的更多期数据，那么通过 DID 的一阶差分，只要**一阶差分是连续的**，就可以识别断点的因果效应。这时，回归中针对的是一阶差分的结果。

DID-RD 在**地理边界的断点**经常使用，因为**行政区划**往往会导致断点附近的跳跃。做 DID-RD 的同时也要辅以 DID，**DID-RD 估计的是局部效应**，DID 估计的是平均效应。

4.3.2 多重处理和自选择偏差

对于标准的地理边界 RD 模型

$$y_i = f(D_i) + \tau(D_i)\mathbf{1}_{D_i \geq 0} + u_i$$

- D_i 是测量距离处理区域边界的运行变量，距离被归一化为零，正距离位于处理区域内；一般只会划定比较小的区域。
- $f(D)$ 总结了影响结果的特定于位置的特征（例如便利设施和劳动力市场）
- u_i 表示可能未观察到的个体特定特征，并影响结果变量
- $Y_i(0) = f(D_i) + u_i$ 确定在没有处理的情况下的结果变量
- $\tau(D_i)$ 是距离 D_i 处的平均处理效应。

这个模型的假设是：

- $f(D)$ 在截止点 $D = 0$ 处具有连续性：**运行变量对结果的影响在有和没有处理时在截止点连续**
- $E[u_i|D_i = D]$ 在截止点 $D = 0$ 具有连续性：**其他可能未观测到的个体特征变量对结果的影响在截止点连续**

在地理不连续性的背景下，$f(D)$ 的不连续性可能来自**多个政策**在边界处发生变化（其他政策已经带来了不连续），而 $E[u_i|D_i = D]$ 的不连续性可能来自于**边界上的自选择**；这种情况下将无法控制，并会带来偏误。如果所有地区的自选择程度被认为相同的（比如人人选择最大化自身效用函数），那么政策的影响仍可以评估。实际研究中往往探讨政策对自选择的限制，那么部分的自选择还是可以容忍的。

4.3.3 断点差分模型

如果时间不变的自选择一直存在，那么预测断点附近本身就不连续，这样的情况可以通过一阶差分合适处理。当政策出台时的自选择很严重，那么政策的效果可能完全估计错误。当然，根据过去的指标确定政策，或者政策立即执行，就无法进行自选择。

假设在处理之前，$t = 0$ 时的边界的不连续性告诉作者其它政策变化和时间不变的自选择的影响，该模型变为

$$y_{it} = f_t(D_i) + \gamma(D_i)\mathbf{1}_{D_i \geq 0} + \tau(D_i)\mathbf{1}_{D_i \geq 0}\mathbf{1}_{t=1} + u_{it}$$

- $\gamma(D_i)$：没处理的处理组的影响；截止点处的时间不变的不连续性，可能是由于时间不变的自选择和/或边界上其他政策的影响
- $\tau(D_i)$：**政策的净影响；所关心的处理效应**

4.3.4 空间 RDD

在地理回归断点（GRD, Geographic Regression Discontinuity）设计中，地理或行政边界将样本划分成处理组和对照组。空间 RDD 面临的主要挑战有：

- **多重处理**：多种同时干预结果的政策
- **连续性假设可能不成立**：当不连续性是地理的时，代理人可能会非常**精确地围绕边界进行自选择**

因此需要相当多的实质知识才能可信地利用地理边界作为 RD 设计。

关于连续性的假设，有两种检验方法：

- 第一种方法可以使用上面概述的局部线性模型**将前定协变量作为结果进行处理**
  - 这些协变量在处理发生之前就确定，**理论上处理效应为零**，估计的效应应该不显著，目的是证明前定变量的连续性
  - **前定变量需要排除自选择的可能，不受政策影响**；最好是协变量，会影响被解释变量
    * 常用的有温度、环境污染程度、抢劫发生率等
- 第二种基于预处理协变量的常见虚拟假设检验类型是一系列针对**协变量的“平衡测试”**，即调查协变量的平均值（或分布的其他特征）是否在靠近截断处的处理和对照样本之间在统计上无法区分
  - 可以简单地比较围绕断点的等距离平均值
  - 但这种方法不如第一种方法直接

通常沿着地理边界高密度抽样，有的研究会全样本抽样。事实上，政策会对样本产生影响，因此调查到的结果可能是选择后的结果（比如耕地政策影响的两区，禁耕区人口更少，耕作区人口更多，留在禁耕区的人可能基于个人因素选择）。并非全样本抽样中的所有人都会服从调查。此外，跨地理边界还要考虑多重处理的影响，**一种方法是用历史数据证明没有断点；另一种方法是验证各个变量在地理边界附近连续**。全样本调查中的样本损失在 RDD 下的影响会被放大。

4.4 时间 RDD

采用时间断点分析（RDiT, Regression Discontinuity in Time）的数据通常有如下特点：

1. 政策的实施没有可用的横向差异，因此不可能使用差分-差分框架
   - 地区之间处理强度不同时，可以使用处理强度的双重差分模型。同时，为了保证处理强度是外生的，需要找到工具变量。

2. 环境空气质量数据的密度比较低
3. 存在长期混杂因素

比如考虑工人罢工（$strike_{it}$）对于交通拥堵程度的影响，在控制日期的效应并允许处理前后斜率差异后，估计的模型为

$$y_{it} = \alpha + \beta strike_{it} + \gamma_1 date_{it} + \gamma_2 date_{it} \times strike_{it} + \delta X_{it} + \varepsilon_{it}$$

$date_{it} \times strike_{it}$ 项的引入只是为了允许断点左右两边的斜率不同，β **估计的是断点处罢工的影响**。基础回归甚至可以略去 $date_{it} \times strike_{it}$。当然，这是基础回归，更进一步地可以考虑更高次项。