

断点回归

Regression Discontinuity

杨点溢

Department of Government
London School of Economics and Political Science

Spark 社科量化系列课程



本节课内容

- 1 热身运动
- 2 案例
- 3 尖锐 Sharp RDD
- 4 模糊 Fuzzy RDD
- 5 总结



入门介绍

- 世界充满了 Y 不连续地 discontinuously 取决于 x 的情况 (门槛 Threshold)。比如说:



入门介绍

- 世界充满了 Y 不连续地 discontinuously 取决于 x 的情况 (门槛 Threshold)。比如说:
 - 如果你的分数高于 50 分, 你就及格, 反之不及格



入门介绍

- 世界充满了 Y **不连续地** discontinuously 取决于 x 的情况 (门槛 Threshold)。比如说:
 - 如果你的分数高于 50 分, 你就及格, 反之不及格
 - 你满 18 岁可以投票, 低于就不可以



入门介绍

- 世界充满了 Y 不连续地 discontinuously 取决于 x 的情况 (门槛 Threshold)。比如说:
 - 如果你的分数高于 50 分, 你就及格, 反之不及格
 - 你满 18 岁可以投票, 低于就不可以
 - 意大利市长选举在人口 15000 下是多数 (plurality) 选举, 在 15000 上是多轮决选 (run-off)



入门介绍

- 世界充满了 Y 不连续地 discontinuously 取决于 x 的情况 (门槛 Threshold)。比如说:
 - 如果你的分数高于 50 分, 你就及格, 反之不及格
 - 你满 18 岁可以投票, 低于就不可以
 - 意大利市长选举在人口 15000 下是多数 (plurality) 选举, 在 15000 上是多轮决选 (run-off)
- 这些都是应用 “断点回归” (RDD) 的好机会
 - Regression Discontinuity Design



入门介绍

- 世界充满了 Y 不连续地 discontinuously 取决于 x 的情况 (门槛 Threshold)。比如说:
 - 如果你的分数高于 50 分, 你就及格, 反之不及格
 - 你满 18 岁可以投票, 低于就不可以
 - 意大利市长选举在人口 15000 下是多数 (plurality) 选举, 在 15000 上是多轮决选 (run-off)
- 这些都是应用 “断点回归” (RDD) 的好机会
 - Regression Discontinuity Design
- RDD 在政治科学格外有用: 断点很多



Sharp vs fuzzy (模糊)

- RDD 分为两个品种：
 - “尖锐” sharp design
 - 使用 OLS 回归
 - “模糊” fuzzy design
 - 使用 IV 回归
- 本节课主要讨论 sharp RDD，但是也会讲一讲 fuzzy RDD



- 主要特征：处理分组不连续地取决于一个可被观察的变量 x .



主要概念

- 主要特征：处理分组不连续地取决于一个可被观察的变量 x .
- 简单来讲，处理的分组取决于 x （又叫强制变量 Forcing variable）有没有到一个门槛 (critical value) x_0 ——即断点

$$D_i = \begin{cases} 1 & \text{if } x_i \geq x_0 \\ 0 & \text{if } x_i < x_0 \end{cases}$$



主要概念

- 主要特征：处理分组不连续地取决于一个可被观察的变量 x .
- 简单来讲，处理的分组取决于 x （又叫强制变量 Forcing variable）有没有到一个门槛 (critical value) x_0 ——即断点

$$D_i = \begin{cases} 1 & \text{if } x_i \geq x_0 \\ 0 & \text{if } x_i < x_0 \end{cases}$$

- 处理分组是 x 的**决定性** deterministic 函数：即不是随机分配



主要概念

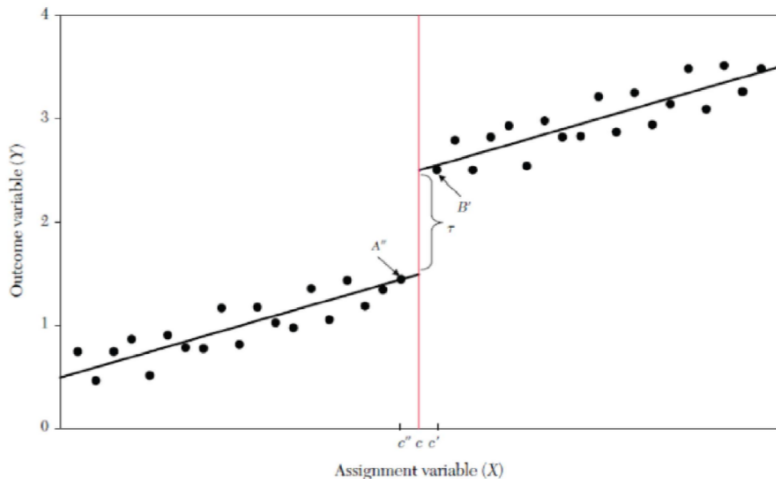
- 主要特征：处理分组不连续地取决于一个可被观察的变量 x 。
- 简单来讲，处理的分组取决于 x （又叫强制变量 Forcing variable）有没有到一个门槛（critical value） x_0 ——即断点

$$D_i = \begin{cases} 1 & \text{if } x_i \geq x_0 \\ 0 & \text{if } x_i < x_0 \end{cases}$$

- 处理分组是 x 的**决定性 deterministic**函数：即不是随机分配
- 但是在断点的**附近**，我们可以说是“如随机”。
 - 意会：在 x_0 **附近**的个体其实都差不多，达没达到门槛某种意义上是随机的



演示：正处理效应



本节课内容

- 1 热身运动
- 2 案例
- 3 尖锐 Sharp RDD
- 4 模糊 Fuzzy RDD
- 5 总结



奖学金对劳动力市场结果的影响 /1

- 我们想要评估评上奖学金本身对学生在职场上有没有正向影响
- 但是，奖学金的评价取决于成绩（或者说能力）
- 简单做一个劳动力市场结果和奖学金（作为虚拟变量）OLS 的回归肯定会有偏的。



奖学金对劳动力市场结果的影响 /2

- 断点回归设计可以解决这个问题——只要奖学金评选规则里有一些“门槛”
 - 比如说，只有学生的 GPA 高于一个门槛才会获得奖学金
- Sharp Design 的关键点在于对比“将高于”和“将低于”门槛的这部分人
 - 这两组人几乎拥有相同的成绩



奖学金对劳动力市场结果的影响 /3

- 设想成绩（或者其他和成绩相关的变量）对职场结果的影响是较为平滑的
 - 在这种情况下，两组学生应该在没有奖学金的情况下结果差不多
 - 一个 GPA 门槛附近的“飙升”可以被理解为奖学金的影响（假设没有其他门槛附近的质性变化）
- 一个重要的识别 identification 假设是成绩对于职场结果的影响是平滑的(smooth/continuous)
- 对于“断点” (discontinuity) 的研究设计，最重要的反而是连续性 (continuity).



在职 (incumbency) 优势 (Lee 2008)/1

- 政客在职的身份会使得他们在未来的选举当中更有优势吗？



在职 (incumbency) 优势 (Lee 2008)/1

- 政客在职的身份会使得他们在未来的选举当中更有优势吗？
- 简单的回归胜率（或者赢得的选票比例）和在职身份可以得到一个正且显著的系数
 - 在美国，大家熟知在职的众议院通常得以“守擂成功”
 - 使用权力、资源、存在感



在职 (incumbency) 优势 (Lee 2008)/1

- 政客在职的身份会使得他们在未来的选举当中更有优势吗？
- 简单的回归胜率（或者赢得的选票比例）和在职身份可以得到一个正且显著的系数
 - 在美国，大家熟知在职的众议院通常得以“守擂成功”
 - 使用权力、资源、存在感
- 但是简单回归得到的系数是有偏的
 - 是在职的优势还是更优秀的政治家本身更有可能选举和再选举成功呢？



在职 (incumbency) 优势 (Lee 2008)/2

- 在两党制种我们有一个天然的“门槛” /断点：只要拿到 50% 的选票就可以赢
 - 在 First-past-the-post 系统中你只要比第二名多 1 票就可以选中
- 我们可以把竞争激烈的选举结果当成随机分配“在职”身份的过程：对比“将将赢的”和“将将输掉”选举的人的未来选举成绩
- 候选人的选票份额应该是他优秀程度的连续函数（也包括符不符合选民胃口）。但是“在职”身份在 50% 这个门槛上有一个断点 \Rightarrow 在职优势在竞争极度激烈情况面下是“好似随机”的。



本节课内容

- 1 热身运动
- 2 案例
- 3 尖锐 Sharp RDD
- 4 模糊 Fuzzy RDD
- 5 总结



- 假设在没有处理 (Treatment) 的情况下, 结果变量关于 x 的条件平均值 (Y conditional on x) 为 $g_N(x)$



尖锐 Sharp RDD

- 假设在没有处理 (Treatment) 的情况下, 结果变量关于 x 的条件平均值 (Y conditional on x) 为 $g_N(x)$
- 假设有处理的情况下, 该值为 $g_t(x)$.



尖锐 Sharp RDD

- 假设在没有处理 (Treatment) 的情况下, 结果变量关于 x 的条件平均值 (Y conditional on x) 为 $g_N(x)$
- 假设有处理的情况下, 该值为 $g_T(x)$.
- 关于 x 的条件平均处理效应 (ATE conditional on x) 为 $g_T(x) - g_N(x)$



尖锐 Sharp RDD

- 假设在没有处理 (Treatment) 的情况下, 结果变量关于 x 的条件平均值 (Y conditional on x) 为 $g_N(x)$
- 假设有处理的情况下, 该值为 $g_T(x)$.
- 关于 x 的条件平均处理效应 (ATE conditional on x) 为 $g_T(x) - g_N(x)$
- 但是我们不能同时观察到同一个 x 值对应的 $g_T(x)$ 和 $g_N(x)$ (因果推断的根本问题: Missing data)



尖锐 Sharp RDD

- 假设在没有处理 (Treatment) 的情况下, 结果变量关于 x 的条件平均值 (Y conditional on x) 为 $g_N(x)$
- 假设有处理的情况下, 该值为 $g_T(x)$.
- 关于 x 的条件平均处理效应 (ATE conditional on x) 为 $g_T(x) - g_N(x)$
- 但是我们不能同时观察到同一个 x 值对应的 $g_T(x)$ 和 $g_N(x)$ (因果推断的根本问题: Missing data)
- 但是我们可以比对断点左右两边离得很近的结果
 - 对比 $g_T(x_0 + \delta) - g_N(x_0 - \delta)$



尖锐 Sharp RDD

- 假设在没有处理 (Treatment) 的情况下, 结果变量关于 x 的条件平均值 (Y conditional on x) 为 $g_N(x)$
- 假设有处理的情况下, 该值为 $g_T(x)$.
- 关于 x 的条件平均处理效应 (ATE conditional on x) 为 $g_T(x) - g_N(x)$
- 但是我们不能同时观察到同一个 x 值对应的 $g_T(x)$ 和 $g_N(x)$ (因果推断的根本问题: Missing data)
- 但是我们可以比对断点左右两边离得很近的结果
 - 对比 $g_T(x_0 + \delta) - g_N(x_0 - \delta)$
- 当 $\delta \rightarrow 0$ 时, 上式趋近于 $g_T(x) - g_N(x)$
 - 即当 $x = x_0$ 时的处理效应



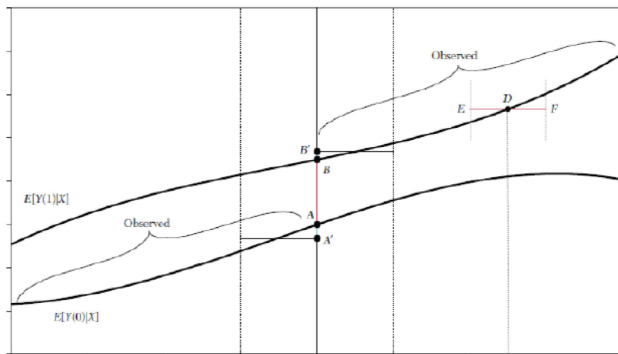
主要特征 /1

- RDD 估计量比较断点两边非常近的观察 (observations), 使用比较值之差 **difference in means** 来测量断点处的处理效应
- 但是简单地比较平均值是**有偏的**
 - 除非 $g_T(x)$ 和 $g_N(x)$ 的斜率是平的
 - 或者你有无限多的在离断点无限近的观察值
- RDD 不考虑离断点**远**的处理效应: 这是 RDD 的局限性之一
 - 局部平均处理效应 Local Average Treatment Effect



门槛附近的平均值比较 Comparison of means around the threshold

Nonparametric 非参数 estimates (simple comparison of means) are biased unless $g_T(x)$ and $g_N(x)$ have flat slopes



主要特征 /2

- 一个重要的假设是实际的 underlying 的 x 对结果的影响是连续的 **continuous** 所以有断点的唯一原因就是处理效应



主要特征 /2

- 一个重要的假设是实际的 underlying 的 x 对结果的影响是连续的 **continuous** 所以有断点的唯一原因就是处理效应
- 为了估计处理效应，你需要选择一个小的 (δ , 即窗口 window): 这会导致你的观察量比较小 \Rightarrow 一致 consistent 的估计值但是低精确度 (se 偏大)



主要特征 /2

- 一个重要的假设是实际的 underlying 的 x 对结果的影响是连续的 **continuous** 所以有断点的唯一原因就是处理效应
- 为了估计处理效应，你需要选择一个小的 (δ , 即窗口 window): 这会导致你的观察量比较小 \Rightarrow 一致 consistent 的估计值但是低精确度 (se 偏大)
- 增加你的样本 (观察) 数量你需要选择一个更大的 $\delta \Rightarrow$ 但是我们的处理效应估计值的一致性会有所欠缺 (处理组和控制组相似度更差)



主要特征 /2

- 一个重要的假设是实际的 underlying 的 x 对结果的影响是连续的 **continuous** 所以有断点的唯一原因就是处理效应
- 为了估计处理效应，你需要选择一个小的 (δ , 即窗口 window): 这会导致你的观察量比较小 \Rightarrow 一致 consistent 的估计值但是低精确度 (se 偏大)
- 增加你的样本 (观察) 数量你需要选择一个更大的 $\delta \Rightarrow$ 但是我们的处理效应估计值的一致性会有所欠缺 (处理组和控制组相似度更差)
- 这导致一个精确度 precision 和一致性 consistency 之间的取舍 trade-off. 通常来讲好的 RDD 需要很大的样本量



- 控制应该不被处理影响 \Rightarrow 不应该在断点附近有大的波动



主要特征 /3

- 控制应该不被处理影响 \Rightarrow 不应该在断点附近有大的波动
- Forcing variable (x) 在门槛 x_0 附近不应该有 “bunching” (聚束)
 - x 在 x_0 附近的频率密度 frequency density 不应该有断点



主要特征 /3

- 控制应该不被处理影响 \Rightarrow 不应该在断点附近有大的波动
- Forcing variable (x) 在门槛 x_0 附近不应该有 “bunching” (聚束)
 - x 在 x_0 附近的频率密度 frequency density 不应该有断点
- 研究者常常先画一个 y on x 的散点图来初步了解 functional form 的可能形式，以及窗口 δ 的选择



主要特征 /3

- 控制应该不被处理影响 \Rightarrow 不应该在断点附近有大的波动
- Forcing variable (x) 在门槛 x_0 附近不应该有 “bunching” (聚束)
 - x 在 x_0 附近的频率密度 frequency density 不应该有断点
- 研究者常常先画一个 y on x 的散点图来初步了解 functional form 的可能形式, 以及窗口 δ 的选择
- 无论如何, 使用不同形式来检验估计值的稳健性都是一个好主意



- 回到在职优势的例子 (Lee 2008)



使用 Sharp Design /1

- 回到在职优势的例子 (Lee 2008)
- 一个实行 sharp design 的方法是做一个选举结果和在职虚拟变量的回归，并聚焦于接近 50% 的选举结果。



- 回到在职优势的例子 (Lee 2008)
- 一个实行 sharp design 的方法是做一个选举结果和在职虚拟变量的回归，并聚焦于接近 50% 的选举结果。
- 在选取“接近”门槛的结果时，我们在面临一个取舍
 - 越严格的“接近”标准意味着两组的可比性越强（一致性 consistency）
 - 但同样意味着观察量更少，损失的是精确性（se 更大）



- 回到在职优势的例子 (Lee 2008)
- 一个实行 sharp design 的方法是做一个选举结果和在职虚拟变量的回归，并聚焦于接近 50% 的选举结果。
- 在选取“接近”门槛的结果时，我们在面临一个取舍
 - 越严格的“接近”标准意味着两组的可比性越强（一致性 consistency）
 - 但同样意味着观察量更少，损失的是精确性（se 更大）
- 单纯比对平均值是**有偏的**



使用 Sharp Design /2

- 另一种方法实行 RDD 的方法是在回归中控制强制变量 forcing variable(x)
 - 我们要用的方法



使用 Sharp Design /2

- 另一种方法实行 RDD 的方法是在回归中控制强制变量 forcing variable(x)
 - 我们要用的方法
- 记得选举胜利是选票份额 (vote share) 的**决定性函数**(deterministic function)



使用 Sharp Design /2

- 另一种方法实行 RDD 的方法是在回归中控制强制变量 forcing variable(x)
 - 我们要用的方法
- 记得选举胜利是选票份额 (vote share) 的**决定性函数**(deterministic function)
- 如果胜利完全取决于 x 那么我们可以通过控制 x (**local linear regression** 局部线性回归) 或者 $f(x)$ (polynomials 多项式) 的方式移出所有的干扰变量



使用 Sharp Design /2

- 另一种方法实行 RDD 的方法是在回归中控制强制变量 forcing variable(x)
 - 我们要用的方法
- 记得选举胜利是选票份额 (vote share) 的**决定性函数**(deterministic function)
- 如果胜利完全取决于 x 那么我们可以通过控制 x (**local linear regression** 局部线性回归) 或者 $f(x)$ (polynomials 多项式) 的方式移出所有的干扰变量
- 在这种情况下我们可以回归 $t+1$ 时选举胜利的几率 (Y_{t+1}) 和在职状态, 同时控制 $f(x_t)$:

$$Y_{t+1} = \beta_0 + \beta_1 Incumbency_{t+1} + f(x_t) + u_{t+1}$$



使用 Sharp Design /2

- 另一种方法实行 RDD 的方法是在回归中控制强制变量 forcing variable(x)
 - 我们要用的方法
- 记得选举胜利是选票份额 (vote share) 的**决定性函数**(deterministic function)
- 如果胜利完全取决于 x 那么我们可以通过控制 x (**local linear regression** 局部线性回归) 或者 $f(x)$ (polynomials 多项式) 的方式移出所有的干扰变量
- 在这种情况下我们可以回归 $t+1$ 时选举胜利的几率 (Y_{t+1}) 和在职状态, 同时控制 $f(x_t)$:

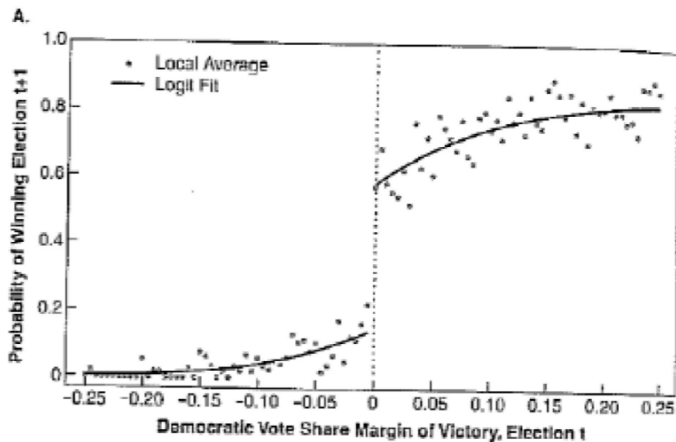
$$Y_{t+1} = \beta_0 + \beta_1 Incumbency_{t+1} + f(x_t) + u_{t+1}$$

- 你需要检查你的结果是否受多项式 polynomial 的形式和窗口 δ 的选择影响
 - 稳健性



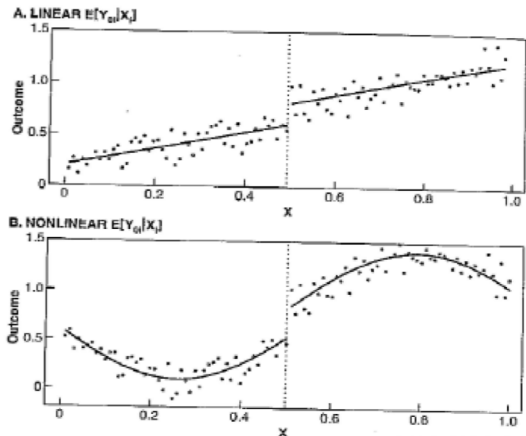
使用 Sharp Design /3

- 参数 β_1 测量在断点处 $f(x)$ 的“跳跃”：



使用 Sharp Design /4

- 函数 $f(x)$ 有好几种形式
 - 通常为多项式 (polynomials) 或者交互项 (interactions)



- 我们不想把一个曲线 $f(x)$ 误解为一个断点，所以一定要考虑到曲线的形式

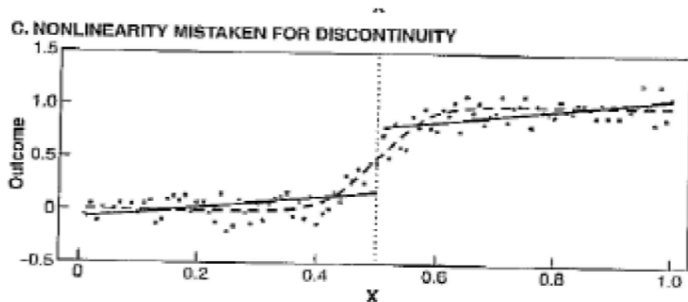
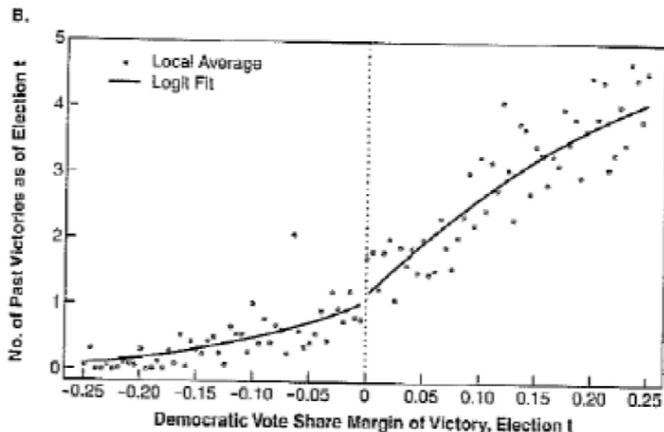


Figure 6.1.1 The sharp regression discontinuity design.



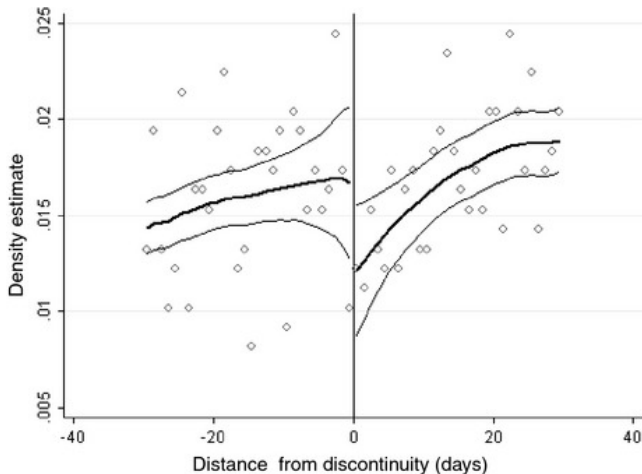
测试“准随机 (quasi-random)”分配 /1

- 在职优势：一个额外的测试为查看 $t-1, t-2 \dots$ (和其他预先决定 pred-determined 的变量) 不收到 x 在 t 时的影响.



测试“准随机 (quasi-random)”分配 /2

- 强制变量 (x) 在门槛 (x_0) 处不应该有断点或者**聚束 (bunching)**
 - 否则强制变量 (x) 可能被“篡改”过，“准随机”假设失效
 - 测试方法：检验 x 在门槛两边的**密度 density** (McCrary test)



应用：选民影响 affect 还是选出 elect 政策？[Lee, Moretti, and Butler 2004]

- 对于代议制民主有两种关于选举作用的观点：
 - 选举承诺是可信的（代表人物 Downs）：政策趋于中间立场
Convergence to middle ground policies
 - 选举承诺不可信：选民仅仅“选出”政策，而政客推行的是自身的偏好



应用：选民影响 affect 还是选出 elect 政策？[Lee, Moretti, and Butler 2004]

- 对于代议制民主有两种关于选举作用的观点：
 - 选举承诺是可信的（代表人物 Downs）：政策趋于中间立场
Convergence to middle ground policies
 - 选举承诺不可信：选民仅仅“选出”政策，而政客推行的是自身的偏好
- 识别 identification 问题：共和党代表只在亲共和党区域胜选；民主党代表只在亲民主党地区胜选



应用：选民影响 affect 还是选出 elect 政策？[Lee, Moretti, and Butler 2004]

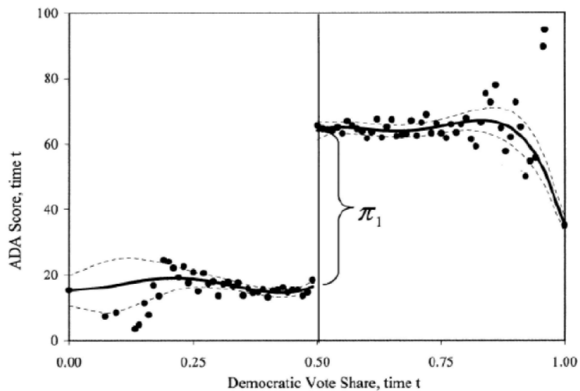
- 对于代议制民主有两种关于选举作用的观点：
 - 选举承诺是可信的（代表人物 Downs）：政策趋于中间立场
Convergence to middle ground policies
 - 选举承诺不可信：选民仅仅“选出”政策，而政客推行的是自身的偏好
- 识别 identification 问题：共和党代表只在亲共和党区域胜选；民主党代表只在亲民主党地区胜选
- 使用结果相似的选举作为样本可以“近似随机”



应用：选民影响 affect 还是选出 elect 政策？[Lee, Moretti, and Butler 2004]

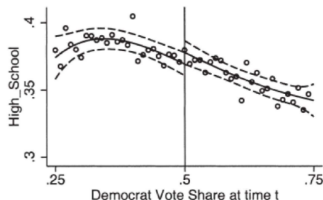
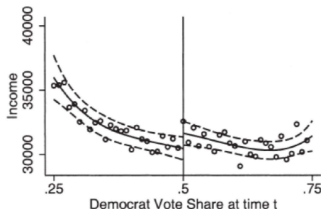
- 对于代议制民主有两种关于选举作用的观点：
 - 选举承诺是可信的（代表人物 Downs）：政策趋于中间立场
Convergence to middle ground policies
 - 选举承诺不可信：选民仅仅“选出”政策，而政客推行的是自身的偏好
- 识别 identification 问题：共和党代表只在亲共和党区域胜选；民主党代表只在亲民主党地区胜选
- 使用结果相似的选举作为样本可以“近似随机”
- 使用 ADA(Americans for Democratic Action) 统计参议院 roll call voting 在 1946-1955 年的分数 (0=conservative; 100=liberal) 作为结果变量





测试准实验分配假设

- 如果相近的选举结果 (close elections) 是近似随机分配的, 那么门槛两边的观察应该是在各种可被观察的层面 (observables) 上相似。(提示: RDD 理论上应该能处理不可被观察的特征 unobservables)
 - ⇒ 可被观察的特征在断点处不应该有跳跃 (jump)



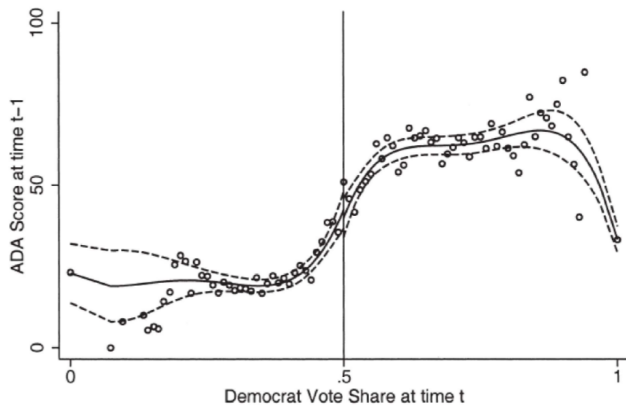


FIGURE V

应用：Reconsidering Duverger's law (Fujiwara 2011)

Duverger: 选举规则塑造党派系统 (party system)

1. 简单多数制 (SMDP/FPTP) 创造两党制
2. 比例制 (proportional) 和两轮多数制 (two-round majority) 创造多党制
 - 经验识别难题 (empirical identification challenge): 互为因果?



应用：Reconsidering Duverger's law (Fujiwara 2011)

Duverger: 选举规则塑造党派系统 (party system)

1. 简单多数制 (SMDP/FPTP) 创造两党制
2. 比例制 (proportional) 和两轮多数制 (two-round majority) 创造多党制
 - 经验识别难题 (empirical identification challenge): 互为因果?
 - 需要在控制别的变量不变的情况下选举规则出现变化



应用：Reconsidering Duverger's law (Fujiwara 2011)

Duverger: 选举规则塑造党派系统 (party system)

1. 简单多数制 (SMDP/FPTP) 创造两党制
2. 比例制 (proportional) 和两轮多数制 (two-round majority) 创造多党制
 - 经验识别难题 (empirical identification challenge): 互为因果？
 - 需要在控制别的变量不变的情况下选举规则出现变化
 - 在巴西，20 万人口以下的城市使用一轮多数制 (简单多数制)，以上使用两轮多数制 ⇒ 门槛



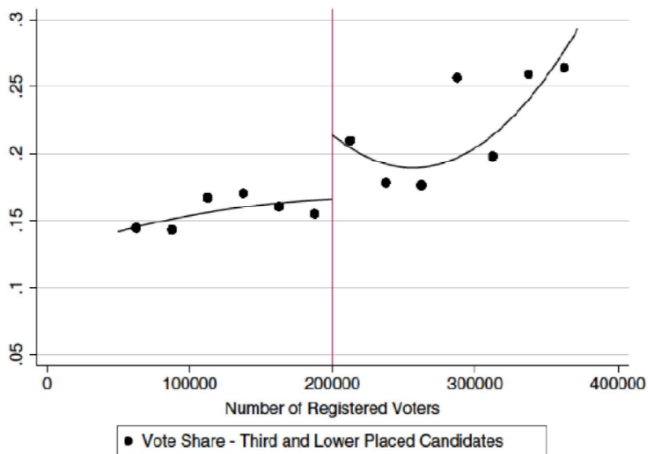
应用：Reconsidering Duverger's law (Fujiwara 2011)

Duverger: 选举规则塑造党派系统 (party system)

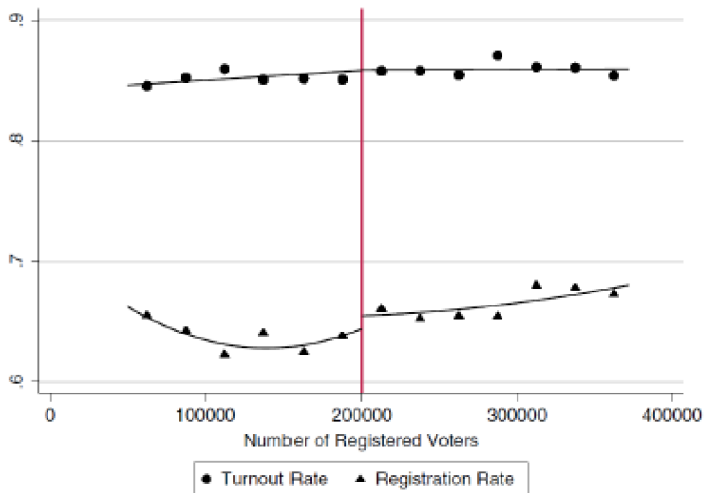
1. 简单多数制 (SMDP/FPTP) 创造两党制
2. 比例制 (proportional) 和两轮多数制 (two-round majority) 创造多党制
 - 经验识别难题 (empirical identification challenge): 互为因果?
 - 需要在控制别的变量不变的情况下选举规则出现变化
 - 在巴西, 20 万人口以下的城市使用一轮多数制 (简单多数制), 以上使用两轮多数制 \Rightarrow 门槛
 - Fujiwara (2011) 使用了巴西体制这个特点来做了一个断点回归: 以第三名及以后的候选人的总得票份额作为 y



A Regression Discontinuity Test of Strategic Voting and Duverger's Law



担忧：多数制系统压制投票率



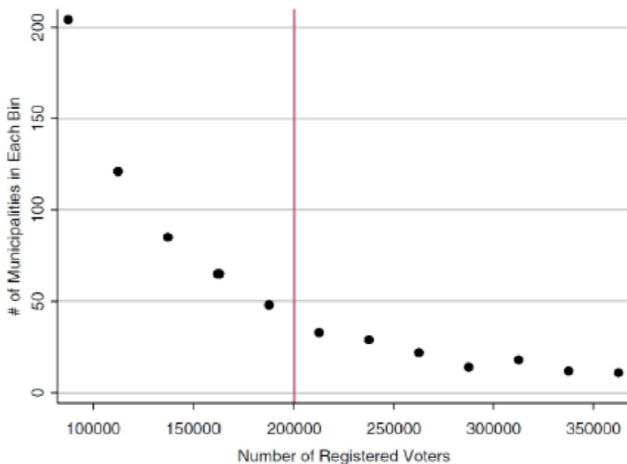


Figure 3. Distribution of electorate size.

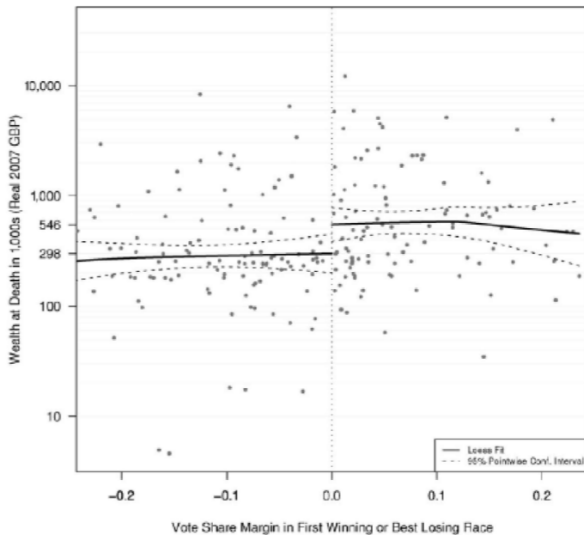


应用：MPs for sale? (Eggers and Hainmuller 2009)

- 成为议员会带来**经济好处**吗？
- 一些“轶事”表明议员会因为他们的身份带来的关系得到收入非常八错的工作
- 但是简单地比较 (naive comparison) 选上的议员和选举失败的候选人会有 OVB
- 这篇文章使用 RDD 比较了选举结果相近的成功议员/败选候选人去世时的财富
 - 遗产数据在英国是公开的



Conservative Candidates



本节课内容

- 1 热身运动
- 2 案例
- 3 尖锐 Sharp RDD
- 4 模糊 Fuzzy RDD
- 5 总结



- 利用了处理发生的几率probability of treatment(conditional on x) 中的断点

$$P(D_i = 1|x_i) = \begin{cases} g_1(x_i) & \text{if } x_i \geq x_0 \\ g_0(x_i) & \text{if } x_i < x_0 \end{cases}$$

$$E(D_i|x_i) = g_0(x_i) + [g_1(x_i) - g_0(x_i)] T_i$$

where $T_i = 1(x_i \geq x_0)$

- 用 T_i 作的工具变量 IV, 第一阶段 (First Stage) 为:

$$D_i = \alpha + \beta x_i + \gamma x_i^2 + \dots + \pi T_i + \epsilon_i$$



Maimonides rule (Angrist and Lavy 1999)

- 以色列学校的班级大小上限被规定为 40(Maimonides rule)
- 如果年级人数达到了 41 个，该班级将会被拆为两个班；达到 81 会被拆成三个班
- 但是 Maimonides rule 不能完美地预测班级大小（有些学校选择采取比 40 更低的上限）
- 但是在年级人数达到 40/80/120... 时班级大小还是会出现“断崖式”下跌 \Rightarrow 使用这个作为 IV



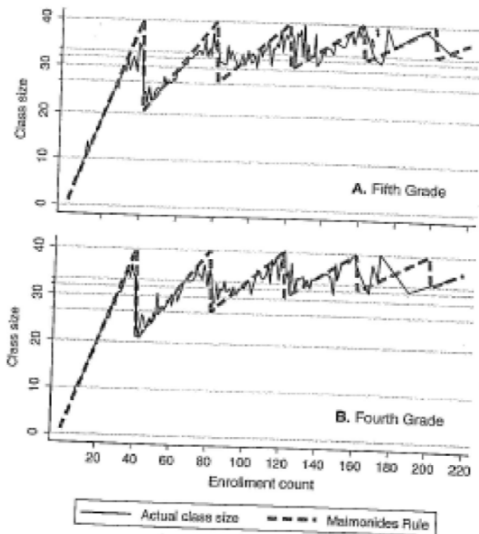


Figure 6.2.1 The fuzzy-RD first-stage for regression-discontinuity estimates of the effect of class size on test scores (from Angrist and Lavy, 1999).



TABLE 6.2.1
OLS and fuzzy RD estimates of the effect of class size on
fifth-grade math scores

| | OLS | | | 2SLS | | | | |
|---------------------------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------------|-----------------|-----------------|
| | | | | Full Sample | | Discontinuity Samples | | |
| | (1) | (2) | (3) | (4) | (5) | ±5 | ±3 | |
| Mean score | | 67.3 | | 67.3 | | 67.0 | 67.0 | |
| (SD) | | (9.6) | | (9.6) | | (10.2) | (10.6) | |
| Regressors | | | | | | | | |
| Class size | .322 (.039) | .076 (.036) | .019 (.044) | -.230 (.092) | -.261 (.113) | -.185 (.151) | -.443 (.236) | -.270 (.281) |
| Percent disadvantaged | | -.340 (.018) | -.332 (.018) | -.350 (.019) | -.350 (.019) | -.459 (.049) | -.435 (.049) | |
| Enrollment | | | .017 (.009) | .041 (.012) | .062 (.037) | | .079 (.036) | |
| Enrollment squared/100 | | | | | -.010 (.016) | | | |
| Segment 1 (enrollment 38-43) | | | | | | | | -12.6 (3.80) |
| Segment 2 (enrollment 78-83) | | | | | | | | -2.89 (2.41) |
| R ² | .048 | .249 | .252 | | | | | |
| Number of classes | | 2,018 | | 2,018 | | 471 | 302 | |

Notes: Adapted from Angrist and Lavy (1999). The table reports estimates of equation (6.2.6) in the text using class averages. Standard errors, reported in parentheses, are corrected for within-school correlation.



本节课内容

- 1 热身运动
- 2 案例
- 3 尖锐 Sharp RDD
- 4 模糊 Fuzzy RDD
- 5 总结



总结 Summary

优点




- 识别 (identification) 依赖于一个非常简单的假设：除了处理 (treatment) 外，其他的变量在断点处丝滑地 (continuously) 变化，从而在断点附近让处理的分配“如随机”
- 这个假设的符合程度很容易被判断出来（但是不能被直接检验）

缺点

- 只能估计断点处的影响（局部平均处理效应 LATE）
- 如果在断点附近有**博弈行为** (strategic behaviour) 那么结果会被污染
- 研究者面临着一致性和精确度的取舍
 - 窗口 δ 越小（断点周围距离限定越小），越一致，但越不精确



References I

-  Angrist, Joshua D. and Victor Lavy (1999). “Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement”. In: *The Quarterly Journal of Economics* 114.2, pp. 533–575. ISSN: 00335533, 15314650. URL: <http://www.jstor.org/stable/2587016> (visited on 09/02/2023).
-  Eggers, Andrew C. and Jens Hainmuller (2009). “MPs for Sale? Returns to Office in Postwar British Politics”. In: *American Political Science Review* 103.4, pp. 513–533. DOI: 10.1017/S0003055409990190.
-  Fujiwara, Thomas (Dec. 2011). “A Regression Discontinuity Test of Strategic Voting and Duverger’s Law”. In: *Quarterly Journal of Political Science* 6.3-4, pp. 197–233. DOI: 10.1561/100.00010037.



References II



Lee, David S. (Feb. 2008). “Randomized experiments from non-random selection in U.S. House elections”. In: *Journal of Econometrics*. The regression discontinuity design: Theory and applications 142.2, pp. 675–697. ISSN: 0304-4076. DOI: 10.1016/j.jeconom.2007.05.004. URL: <https://www.sciencedirect.com/science/article/pii/S0304407607001121> (visited on 08/28/2023).



Lee, David S., Enrico Moretti, and Matthew J. Butler (2004). “Do Voters Affect or Elect Policies? Evidence from the U. S. House”. In: *The Quarterly Journal of Economics* 119.3, pp. 807–859. URL: <https://EconPapers.repec.org/RePEc:oup:qjecon:v:119:y:2004:i:3:p:807-859..>

