

双重差分法

Difference-in-Differences

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 实验 experiments
- 2 准实验 Quasi-experiments
- 3 双重差分法 Difference-in-Differences /1
- 4 应用



回顾过去

- 使用观察性数据来进行因果推断需要解决选择偏差问题
- OLS 策略：使用控制变量。当选择偏差是因为可观察的干扰因素 confounder 导致时没有问题
- 面板数据 Panel data: 使用固定效应来控制不可观察的单位和时间固定效应
- 我们已经深挖回归一段时间了，但还是无法和真的实验相媲美
- 但有时候观察性数据可能由一个近似真实随机的过程产生
 - 准实验 quasi-experiment \Rightarrow 研究设计
 - 自然实验 natural experiment \Rightarrow 研究背景



差分估计量 The differences estimator

- 设想一个**实验**，涉及二元处理 binary treatment $X \in \{0, 1\}$.
- 然后我们可以使用 OLS 估计 X 对 Y 的影响：

$$Y_i = \alpha + \beta X_i + e_i$$

- $\hat{\beta}^{OLS}$ 则是差分估计量 **differences estimator**, \Leftarrow

$$E(Y|X=1) = \alpha + \beta$$

$$E(Y|X=0) = \alpha$$

- 随机分配**可以解决 OVB
- X 可以取很多值（不再是二元）的，在这种情况下系数 β 的解读为“处理强度” **treatment intensity** 的影响。在这种情况下我们的式子 functional form 很重要
 - 剂量



实验中的内部效度 internal validity 问题

- 非随机分配（错误的随机标准 criterion）
- 非随机磨损 attrition
- 霍桑效应 Hawthorne effect
- 处理分组 (intention-to-treat) \neq 处理 (treatment): 处理组只有一部分人选择遵照处理方案 comply with the treatment protocol.
 - 处理组不读报纸 (Gerber, Karlan, and Bergan 2009)



实验的外部效度 external validity 问题

- 样本可能不具有人口的代表性 (not representative)
 - 比如说征召志愿者做实验
- 实验可能不代表真实世界
 - 比如说实验的长短
- 一般均衡效应 (general equilibrium effects): 实验设定中外生 exogenous 的要素在现实生活中更大的方案中可能是内生 endogenous 的。



控制变量还是有用的 /1

要测量一个因果效应我们需要满足**条件平均独立假设** conditional mean independence assumption: 误差项 error term 和自变量 X 不相关, 至少在控制变量后不相关。

$$E(e_i | x_i, w_{1i}, \dots, w_{ki}) = f(w_1, \dots, w_{ki})$$

$$Y_i = \alpha + \beta X_i + \gamma W_i + e_i$$

$$E(Y | X = 1, W_i) = \alpha + \beta + \gamma W_i$$

$$E(Y | X = 0, W_i) = \alpha + \gamma W_i$$

$$\Rightarrow E(Y | X = 1, W_i) - E(Y | X = 0, W_i) = \beta$$



控制变量还是有用的 /2

- 在实验中加入控制变量的原因：
 - 接受实验分组的几率可能跟某些处理前的特征有关： X 是随机分配的，但是还是取决于 conditional on W .
 - 有些人加入不了实验？
 - 检查随机性：如果处理是随机分配的那么 β 不应该在加入 W 之后改变
 - 加入 W 可以提高模型的效率 efficiency (se 降低).
 - 但是注意：被处理影响的协变量 (covariates)/控制变量不应该加进模型 (bad controls)
 - Post-treatment bias



本节课内容

- 1 实验 experiments
- 2 准实验 Quasi-experiments
- 3 双重差分法 Difference-in-Differences /1
- 4 应用



自然 Natural 实验（或者准实验）

- 在真正的实验中，研究者能够决定实验处理（随机分配）
- 在自然实验（或准实验）中，研究者使用**自然的外生变化**来模拟一个随机的实验环境（但不能掌握随机过程）
- 通过一个特定的研究设计，而不是仅仅靠控制变量来解决内生性问题
- 使用自然实验往往需要对研究课题的机构 institutional 和历史背景有充分了解
 - 想要使你使用观察性数据的“准实验”可信，你需要了解很多的知识和细节。
- 总结：不是所有使用“自然实验”的研究都可信



19 世纪伦敦的霍乱 Cholera

- 19 世纪时的英国伦敦地区经历了可怕的霍乱疫情，死亡逾千人。
- 直到 19 世纪中叶还有两种势均力敌的关于霍乱起源的理论
 - 主流理论认为霍乱通过空气传播（“瘴气” miasmas）。
 - 另一种理论认为霍乱是由一种主要靠水源传播的病菌导致的
- 不用解释就知道了找到对的理论对于解决疫情有多重要



John Snow's "on the mode of communication of cholera"

- John Snow 是一名医生，被视为现代传染病学 epidemiology 的开门鼻祖
- 他为第二种理论提供了很多证据
- 尽管他不知道，但他其实开创了双重差分法 Difference-in-differences estimator.



第一个自然实验 /1

- 霍乱疫情于 1854 年的伦敦爆发



第一个自然实验 /1

- 霍乱疫情于 1854 年的伦敦爆发
- 在当时，伦敦的一些区域同时由两个自来水公司供应
 - The Southwark & Vauxhall Water Company 的水来自于泰晤士河中被污水污染的水段
 - The Lambeth Water Company 的水来自于泰晤士河的上游 ⇒ 没有被污染



第一个自然实验 /2

- Snow 做了如下这些事情（考虑一下这些调查多费人力）



第一个自然实验 /2

- Snow 做了如下这些事情（考虑一下这些调查多费人力）
 - 获取了 1854 年霍乱疫情头 7 周的霍乱死亡名单



第一个自然实验 /2

- Snow 做了如下这些事情（考虑一下这些调查多费人力）
 - 获取了 1854 年霍乱疫情头 7 周的霍乱死亡名单
 - 查询了这些死亡所属的地区是由两个自来水公司中的哪个供应的



第一个自然实验 /2

- Snow 做了如下这些事情（考虑一下这些调查多费人力）
 - 获取了 1854 年霍乱疫情头 7 周的霍乱死亡名单
 - 查询了这些死亡所属的地区是由两个自来水公司中的哪个供应的
 - 对于每个死亡，他查询了死者本人的自来水公司



第一个自然实验 /2

- Snow 做了如下这些事情（考虑一下这些调查多费人力）
 - 获取了 1854 年霍乱疫情头 7 周的霍乱死亡名单
 - 查询了这些死亡所属的地区是由两个自来水公司中的哪个供应的
 - 对于每个死亡，他查询了死者本人的自来水公司
- 调查的结果在下一页



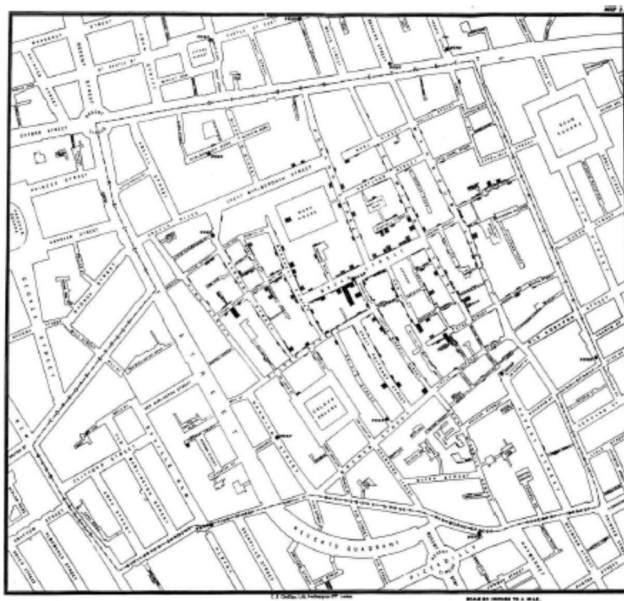
第一个自然实验 /3

Table IX of Snow (1855):

Water Company	Houses	Deaths	Deaths per 10000 Houses
Southwark & Vauxhall	40,046	1,263	315
Lambeth	26,107	98	37
Rest of London	256,423	1,422	59



第一个自然实验 /4



主要特征

- 在同时被两家自来水公司供应的地区中，邻居之间常常用不同的供应商
- 因此这些房子的居民都生活在一样的“瘴气” (miasmas) 中。
- 这些居民也往往不知道他们的水源是哪家公司
- 这种“如随机” (as if random assignment) 的水源分配使得这些社区中的居民在除了水源外的别的方面都非常相像
 - 自然实验!



使用差分估计量 Difference estimator

- 如果我们相信 Snow (1855) 的自然实验好似随机对照实验的话，我们可以通过以下模型估计水源污染的影响：

$$Deaths_i = \alpha + \beta Dirty_i + e_i \quad (1)$$

这里的 *Dirty* 是一个虚拟 dummy 变量 \Rightarrow 使用 Southwark & Vauxhall 公司的自来水为 1，反之为 0.



使用差分估计量 Difference estimator

- 如果我们相信 Snow (1855) 的自然实验好似随机对照实验的话，我们可以通过以下模型估计水源污染的影响：

$$Deaths_i = \alpha + \beta Dirty_i + e_i \quad (1)$$

这里的 *Dirty* 是一个虚拟 dummy 变量 \Rightarrow 使用 Southwark & Vauxhall 公司的自来水为 1，反之为 0.

- 如果可以的话，我们也可以在 (1) 中加上别的自变量/控制变量
 - 可以用来检查随机性（收入、年龄等等）
 - 额外的变量可以增加估计值的精确度 (precision)
 - 我们可能担心某些因素导致人们自我选择 self-select 某一家特定的自来水公司，同时影响感染几率（干扰因素）



第二个自然实验

- 其实 The Lambeth water company 曾经也使用泰晤士河被污染的水段，只是在 1852 年才开始使用干净的水源



第二个自然实验

- 其实 The Lambeth water company 曾经也使用泰晤士河被污染的水段，只是在 1852 年才开始使用干净的水源
- Snows 对比了 1854 年疫情和之前 1849 年的疫情 (i.e. 当 Lambeth 公司还是使用污染水源的时候)，他发现：
 - 只被 Southwark & Vauxhall 公司供应的区域在两个疫情中死亡数量相当
 - 部分/只被 Lambeth 公司供应的地区在 1854 年的疫情中死亡数量相较 1849 年低很多



第二个自然实验

- 其实 The Lambeth water company 曾经也使用泰晤士河被污染的水段，只是在 1852 年才开始使用干净的水源
- Snows 对比了 1854 年疫情和之前 1849 年的疫情 (i.e. 当 Lambeth 公司还是使用污染水源的时候)，他发现：
 - 只被 Southwark & Vauxhall 公司供应的区域在两个疫情中死亡数量相当
 - 部分/只被 Lambeth 公司供应的地区在 1854 年的疫情中死亡数量相较 1849 年低很多
- **双重差分法** Diff-in-Diffs!



本节课内容

- 1 实验 experiments
- 2 准实验 Quasi-experiments
- 3 双重差分法 Difference-in-Differences /1
- 4 应用



双重差分法 /1

- 关于在第二个自然实验中使用类似于第一个实验的简单回归模型 (1) 的问题是使用两个不同自来水公司的家庭/区域可能在 1848 年到 1853 年间有**系统性**systematic 的区别
- 思考如下可能性：
 - 使用 Lambeth 公司的家庭可能普遍更加富有？
 - 这些家庭可能相较使用 Southwark & Vauxhall 公司的家庭在没有干净水源的情况下也会有更低的死亡率（医疗条件更好）
- 为了解决这些问题，我们要使用**双重差分法**



双重差分法 Difference-in-Differences /2

- 思考 Snow 的数据是如何对比 1849 年和 1854 年疫情的
 - 为简单化，我们只考虑只用某一家自来水公司的区域
 - 把使用 Southwark & Vauxhall 公司的区域视为**控制组**control group.
 - 把使用 Lambeth 公司的区域视为**处理组**(treatment)
 - 讲 Lambeth 公司由污染水源转为干净水源的行为视为自然实验的“处理”



双重差分法 Difference-in-Differences /3

- 那么双重差分估计量 diff-in-diff estimator 为:

$$\begin{aligned}\hat{\beta}_{DD} &= [\bar{y}_{ta} - \bar{y}_{tb}] - [\bar{y}_{ca} - \bar{y}_{cb}] \\ &= [\bar{y}_{ta} - \bar{y}_{ca}] - [\bar{y}_{tb} - \bar{y}_{cb}]\end{aligned}$$

- \bar{y}_{ta} 为处理组在 1854 年的平均死亡数
- \bar{y}_{tb} 为处理组在 1849 年的平均死亡数
- \bar{y}_{ca} 为控制组在 1854 年的平均死亡数
- \bar{y}_{cb} 为控制组在 1849 年的平均死亡数
- 本估计量比较处理组 Y 的变化和控制组的 Y 的变化
 - 计算两个差距之间的差距，故为双重差分法



双重差分法 Difference-in-Differences /4

以下表格可能更为直观：

Water supply	Cholera deaths, 1849, rate per 100,000	Cholera deaths, 1854, rate per 100,000	Difference in rates comparing 1854 to 1849, rate per 100,000
Southwark & Vauxhall Company only	1349	1466	118 (时间趋势)
Lambeth Company Only	847	193	-653 (时间趋势 + 处理效应)
Difference between Lambeth and Southwark & Vauxhall	-502 (单位差别)	-1273 (单位差别 + 处理效应)	-771 (处理效应)



双重差分法 Difference-in-Differences /5

- DID 通常用一下回归式估计：

$$Deaths_{it} = \alpha + \beta Treat_i + \gamma After_t + \delta(Treat_i \times After_t) + e_i \quad (2)$$

此时

- $Treat_i$ 是一个虚拟 dummy 变量，当第 i 个观察值在处理组 (Lambeth 公司) 时为 1，在控制组时为 0.
- $After_t$ 也是虚拟 dummy 变量，当第 i 个观察值为处理后时间段 (1854 年) 为 1，在处理前时间段为 0.
- $Treat_t \times After_i$ 是一个交互项 interaction term，其系数代表处理效应。

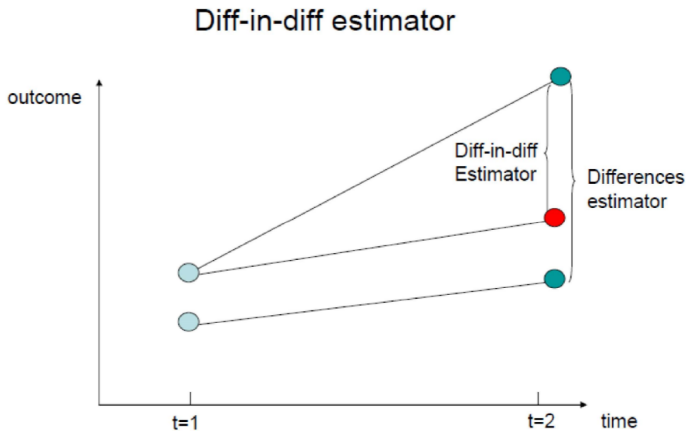


识别假设 Identifying assumption

- **平行趋势** Parallel trends assumption
 - 在没有处理影响的情况下，处理组和控制组的结果走向应该是**平行的**
 - 可以通过检验**处理前**的结果趋势来部分验证
 - interaction of treatment with pre-treatment periods 处理分组和处理前时间段的交互项
- **独立假设** Independence assumption
 - Treatment status 处理分组 is independent of 独立于 potential outcomes 潜在结果 and background attributes 背景特征。
 - “好似” 随机分配
- **排他假设** Excludability assumption
 - 我们可以排除别的因素的影响，从而可以把 Y 的变化归咎于处理本身
- **无干扰假设** Non-interference assumption
 - 每个对象不受别的都对象的处理/分组影响
 - No spillover effects



图示：



添加协变量 covariates

- 有时候添加协变量是有益的：

$$\Delta Y_i = \alpha + \beta X_i + \gamma W_i + e_i$$

- 这里控制变量可以清除 $\hat{\beta}$ 中别的因素对 Y 的影响
- 在 X 是有条件地根据 W 随机分配 (randomly assigned conditional on W) 时格外有必要
- 也可以被视作一种随机度测试. 如果真的近似随机, $\hat{\beta}$ 不应该在加协变量之后有较大变化
- 如果 X 是随机分配的, 加入协变量 W 之后可以增加模型效率 efficiency (i.e. 减少 se)



本节课内容

- 1 实验 experiments
- 2 准实验 Quasi-experiments
- 3 双重差分法 Difference-in-Differences /1
- 4 应用

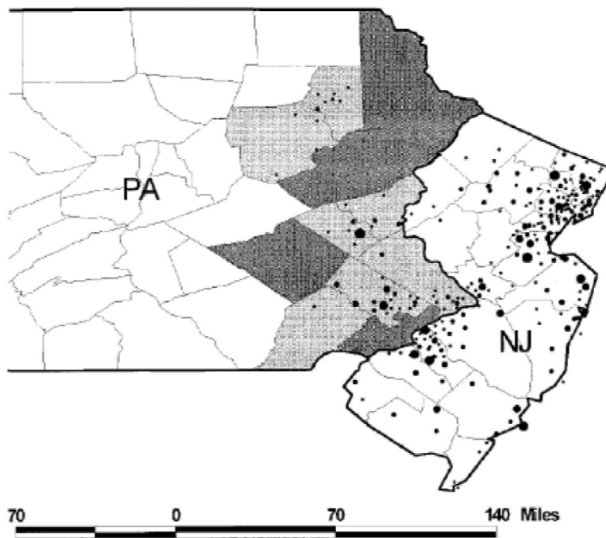


案例 1：最低工资 (Card and Krueger 1994)

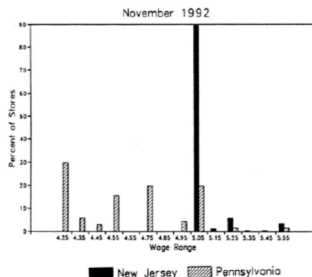
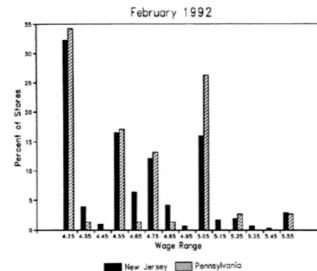
- 最低工资对就业的影响是什么样的
 - 完美市场假设：导致失业
 - 不完美市场假设：增加就业
- Card and Krueger (1994) 是 DID 现代意义上的开山之作，使用了如下自然实验
 - 1992 年 4 月，新泽西 New Jersey 州将全州最低时薪从 4.25 涨到了 5.05 美元
 - 对比新泽西（处理组）和临近的宾夕法尼亚州（控制组）就业的变化
 - 考察了 410 家快餐店在 1992 年二三月和 1992 年十一/十二月的变化
 - 数据包括工资、就业、饭店信息



饭店的位置 (Card and Krueger 1994)



政策前后的薪资水平 (Card and Krueger 1994)



结果 (Card and Krueger 1994)

Variable	PA	NJ	Difference NJ-PA
FTE employment before	23.33 (1.35)	20.44 (0.51)	-2.89 (1.44)
FTE employment after	21.17 (0.94)	21.03 (0.52)	-0.14 (1.07)
Change in mean FTE employment	-2.16 (1.25)	0.59 (0.54)	2.76 (1.36)



案例 1：最低工资 (Card and Krueger 1994)

- 结果的潜在问题：
 - 最低工资的变化可能和其他政策变化相关（后续的研究表明没有这个问题）
 - 新泽西和宾州的经济周期不同？
 - 平行假设？
 - 测量误差
 - 过度相邻 \Rightarrow 溢出效应？Spillover effect
- 这些问题在本文/后续的文章里有过讨论
- 不管怎么样，没有发现合理的最低工资对就业的不利影响



平行趋势假设 Parallel trends assumption

- 识别策略的关键：展示处理前的结果趋势
- 比如说，如果有最低工资改革前 5 年的宾州和新泽西州数据
- 如果前五年的就业数据都是平行的
- 那么我们的识别策略就是可信的
- 但是“预测出”一个未来发生的政策非常难。特别是需要我们自己收集数据的时候。
 - 通常采用公开数据



安慰剂检验 Placebo tests

- 不难理解：把“处理时间”调制真实的时间之前，然后看看“处理影响”还是否存在
- 如果处理发生在时间 t ，那么双重差分法在比较时间 t 和 $t-1$ 的区别。安慰剂检验则是比较 $t-1$ 和 $t-2$ ， $t-2$ 和 $t-3$...
- 在伦敦霍乱例子里，Snow (1855) 比较了 1849 和 1854 疫情
 - 我们可以比较 1849 和再之前的疫情 (1832)
 - 1849 和 1832 年的疫情之间不应该有区别



案例 2：南非民主化和公共服务 (Kroth, Larcinese, and Wehner 2016)

- 研究问题：民主化是否会带来基本公共服务的提升？
- 比较不同国家间的研究很难逃脱内生性的影响
- 我们因此考虑了一个当代大规模的民主化进程
 - 1900 万群众，其中 84% 为南非种族隔离 apartheid 结束后的新选民
 - 使用了 1996 年和 2001 年的人口普查数据 + 夜晚灯光数据（城市为单位）



案例 2：背景 (Kroth, Larcinese, and Wehner 2016)

- 曼德拉的南非国大党 (ANC) 承诺给全国带来大规模电气化 electrification
- 历史上和宪法上，都是本地政府负责电网建设 (electricity reticulation)
- 电力供应本身在这个时间段不是个问题（1990 年空置发电量 reserve margin 为 55%）
- 但是分配（入户）是个问题：不是所有的城市的电网都很成熟



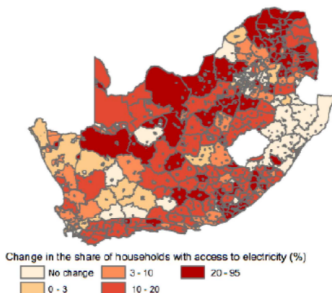
实证策略 Empirical strategy (Kroth, Larcinese, and Wehner 2016)/1

- 自变量 (“实验处理”)：新选民比例 (Enfranchised)：1996 年选举适龄人群中非白人比例
- 因变量： $\Delta Electricity$ 是 1996 年和 2001 年间使用电力作为主要照明源的家庭比例变化
- 次要因变量： $\delta Nightlight$ 是 1996 年到 2001 年间城市夜晚图像中亮灯的像素比例变化
- 注意：南非第一轮真正民主的本地政府选举是在 1995 年 11 月和 1996 年 6 月间，第二轮是在 2000 年 12 月。

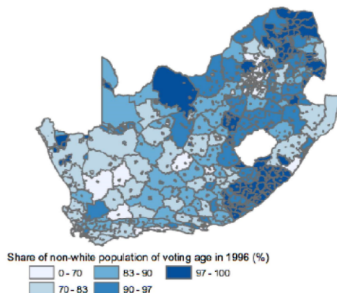


实证策略 Empirical strategy (Kroth, Larcinese, and Wehner 2016)/2

Δ Electricity
(HH access 1996-2001)



Enfranchised
(non-white 1996)



主要结果（人口普查数据）

Table 1: The impact of enfranchisement on electrification (census data)

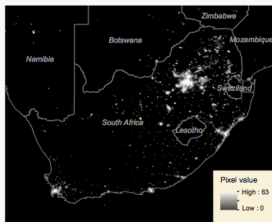
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Enfranchised	0.343*** (0.032)	0.280*** (0.032)	0.293*** (0.033)	0.232*** (0.041)	0.191*** (0.040)	0.277*** (0.040)	
Enfranchised black							0.327*** (0.041)
Enfranchised coloured							0.204*** (0.047)
Enfranchised Indian							-0.022 (0.059)
Constant	-19.058*** (2.534)	-11.056*** (3.001)	-6.992* (3.619)	-5.269 (4.399)	-2.204 (4.328)	-12.815*** (4.757)	-13.817*** (4.565)
Province fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	No	No	Yes	Yes	Yes	Yes	Yes
Population and socioeconomic controls (1996)	No	No	No	Yes	Yes	Yes	Yes
Households without electricity (1996)	No	No	No	No	Yes	Yes	Yes
Population and socioeconomic controls (1996-2001 diff.)	No	No	No	No	No	Yes	Yes
R-squared	0.111	0.161	0.173	0.192	0.245	0.307	0.325

Note: The dependent variable is the percentage share of households with electricity for lighting (difference 1996-2001) calculated from census data. Geographic controls are: (1) Distance from electricity grid; (2) Distance from main road; (3) Elevation; (4) Slope. Population controls are: (1) Population density; (2) Number of households. Socioeconomic controls are: (1) Share of population with no schooling; (2) Median income; (3) Share of labor force with low income (due to differences in the underlying variables in the 2001 census, this variable is only included as a 1996 level control and not as a 1996-2001 difference). Refer to the data appendix for full details. N = 799. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

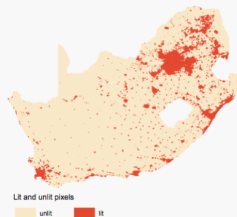


卫星图片数据

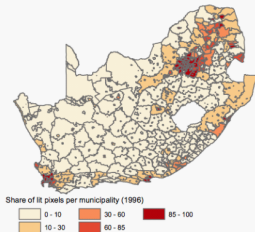
Map 1: Raw nighttime lights raster image



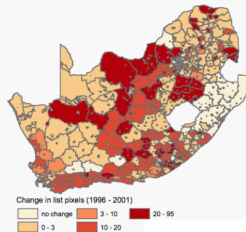
Map 2: Reclassified nighttime lights raster image



Map 3: Share of lit pixels per municipality (1996)



Map 4: Change in the share of lit pixels (1996-2001)



- 双重差分法有更复杂的形式
 - 我们不只两个时间段的数据
 - 不是所有处理个体都是在同一时间接受处理
 - Staggered 多期 DID
 - 有不同**级别**的处理，而不是统一的处理
- 同样使用的是双重差分的逻辑



主要结果（卫星图像数据）+ 安慰剂检验

Table 2: The impact of enfranchisement on electrification (satellite data, with placebo regressions)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>a. Dependent variable: ΔNightlight 1996-2001</i>							
Enfranchised	0.064*** (0.023)	0.063*** (0.024)	0.066*** (0.025)	0.092*** (0.034)	0.088** (0.035)	0.080** (0.037)	
Enfranchised black							0.109*** (0.038)
Enfranchised coloured							0.041 (0.044)
Enfranchised Indian							0.015 (0.049)
<i>b. Dependent variable: ΔNightlight 1992-1996</i>							
Enfranchised	0.068 (0.052)	0.029 (0.059)	0.037 (0.056)	0.047 (0.063)	0.059 (0.064)	-0.006 (0.070)	
Enfranchised black							0.007 (0.072)
Enfranchised coloured							-0.026 (0.078)
Enfranchised Indian							-0.078 (0.082)
Province fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	No	No	Yes	Yes	Yes	Yes	Yes
Population and socioeconomic controls (1996)	No	No	No	Yes	Yes	Yes	Yes
Households without electricity (1996)	No	No	No	No	Yes	Yes	Yes
Population and socioeconomic controls (1996-2001 diff.)	No	No	No	No	No	Yes	Yes

Note: The dependent variable is the percentage share of lit pixels (difference 1996-2001 and 1992-1996, respectively) calculated from satellite data. All regressions also include a constant. Refer to Table 1 for a description of control variables, and the data appendix for full details. N = 799. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.



DID 和固定效应

基本 DID 形式:

$$Y_{it} = \alpha + \beta Treatment_i + \gamma After_t + \delta (Treatment_i \times After_t) + e_{it}$$

现在定义:

- $\alpha + \beta Treatment_i = \alpha_i$
- $Treatment_{it} = (Treatment_i \times After_t)$
- $Treatment$ 现在不代表处理组, 而是代表该个体在该时间段有没有受到处理, 然后可得

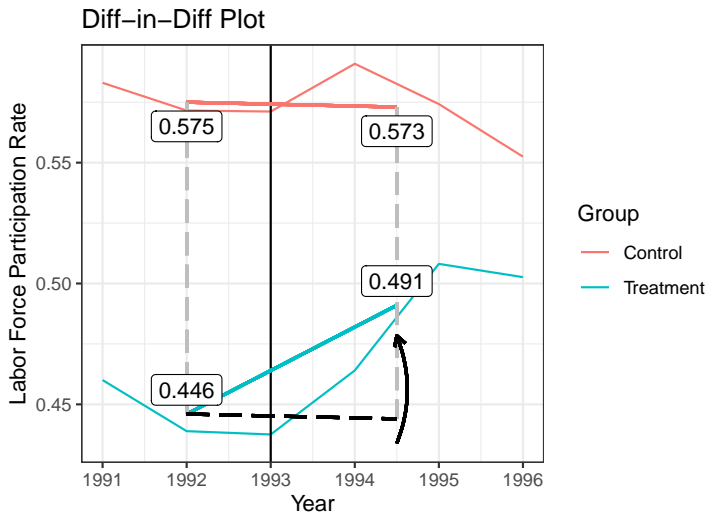
$$Y_{it} = \alpha_i + \gamma_t + \delta Treatment_{it} + e_{it}$$

- \Rightarrow 你可以使用 TWFE 来跑 DID 模型



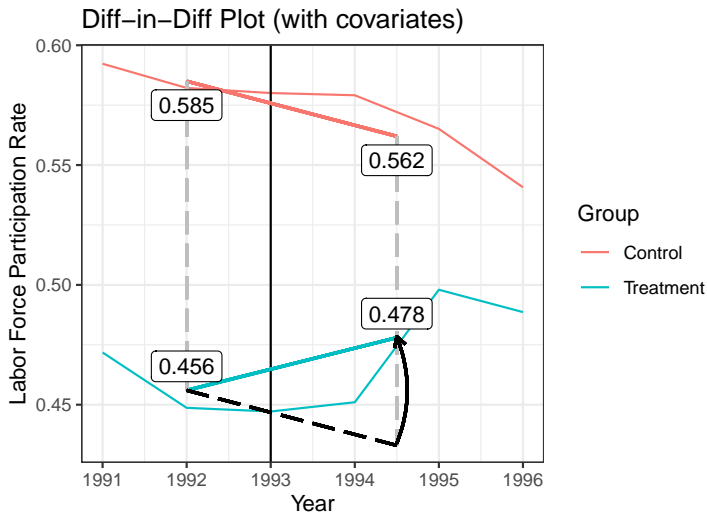
R 实操：代码原型来自Goulding 2011

课题：Effect of Earned Income Tax Credit on female labour force participation (Plot without covariates)



R 实操：代码原型来自Goulding 2011

课题：Effect of Earned Income Tax Credit on female labour force participation (Plot adjusted for covariates)



References I



Card, David and Alan Krueger (1994). “Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania”. In: *American Economic Review* 84.4, pp. 772–93. URL: <https://EconPapers.repec.org/RePEc:aea:aecrev:v:84:y:1994:i:4:p:772-93>.



Gerber, Alan S., Dean Karlan, and Daniel Bergan (Apr. 2009). “Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions”. en. In: *American Economic Journal: Applied Economics* 1.2, pp. 35–52. ISSN: 1945-7782. DOI: 10.1257/app.1.2.35. URL: <https://www.aeaweb.org/articles?id=10.1257/app.1.2.35> (visited on 07/29/2023).



References II



Goulding, Kevin (June 2011). *Differences-in-Differences estimation in R and Stata*. en. URL: <https://thetarzan.wordpress.com/2011/06/20/differences-in-differences-estimation-in-r-and-stata/> (visited on 08/10/2023).



Kroth, Verena, Valentino Larcinese, and Joachim Wehner (July 2016). "A Better Life for All? Democratization and Electrification in Post-Apartheid South Africa". In: *The Journal of Politics* 78.3, pp. 774–791. DOI: 10.1086/685451.



Snow, John (1855). *On the mode of communication of cholera*. eng. London : John Churchill. (Visited on 08/09/2023).

