第四章　面板数据

罗姗姗

March, 2023

目录

例子 1 (劳动经济学) I

研究背景

最低工资是政府调控劳动力市场的一种手段. 然而，提高最低工资可能会对雇主造成经济负担，从而导致雇佣机会的减少. 这种影响在经济学中被称为“就业歧视效应”，但在实际情况中并不总是发生. 因此，有必要研究最低工资对就业的影响.

例 1 (Card & Krueger, 1994)

目标：评估新泽西州最低工资变化前后，新泽西州快餐店的平均就业人数.

- 在 1992 年 4 月 1 日，新泽西州将州最低工资从 4. 15 美元每小时提高到 5. 05 美元每小时
- 在同一时期，宾夕法尼亚的最低工资一直维持在 4. 25 美元每小时.
- 在 1992 年 2 月份，宾夕法尼亚州快餐店就业人数高于新泽西州，但是在 11 月份却下降了.

例子 1 (劳动经济学) II

研究结果

研究发现，尽管新泽西州提高了最低工资标准，但就业人数并没有减少. 相反，在新泽西州的快餐店，就业人数略有上升，而在宾夕法尼亚州则有下降趋势. 这一结果表明，提高最低工资标准不一定会导致就业歧视效应. 然而，这个结果也不能推广到其他行业或地区，因为不同的行业和地区有其独特的就业市场和工资结构.

双重差分 I

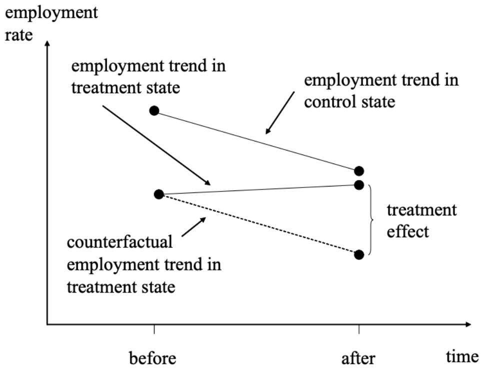

Figure 5.2.1: Causal effects in the differences-in-differences model

提高最低工资使新泽西州的就业趋势偏离了原有的趋势.

双重差分 II

- 双重差分法适用于面板数据或重复观测的横截面数据：两个或更多组别
- 每个组别里的个体有至少有两个观测时间点
- 在某些组别的某些时间点，个体接受处理
- 我们以两个组别和两个时间点为例介绍双重差分法.
- $G = 1$ 表示处理组，$G = 0$ 表示对照组
- 协变量 X
- 用 $T = t, T = t + 1$ 分别表示两个时刻，我们假设 $T = t + 1$ 时刻施加处理
- 定义处理状态 $Z = G * I(T = t + 1)$，也即仅在处理组、$t + 1$ 时刻，$Z = 1$; T 时刻的潜在结果 $Y_T(z)$，其中 $z \in \{0, 1\}$
- 两个时刻的结果变量分别记为 Y_t, Y_{t+1};

双重差分 III

	处理前	处理后
对照组 $G=0$	$Y_t(0)$	$Y_{t+1}(0)$
处理组 $G=1$	$Y_t(0)$	$Y_{t+1}(1)$

- 感兴趣的因果参数：$\tau_{\mathrm{ATT}} = \mathbb{E}\left\{Y_{t+1}(1) - Y_{t+1}(0) \mid G=1\right\} \triangleq \mu_1 - \mu_0$，其中
  - $\mu_1 = \mathbb{E}\{Y_{t+1}(1) \mid G=1\} = \mathbb{E}(Y_{t+1} \mid G=1)$.
  - $\mu_0 = \mathbb{E}\{Y_{t+1}(0) \mid G=1\} = \mathbb{E}(Y_{t+1}(0) \mid G=1)$. (待识别)
- 在正式介绍 DID 估计量之前，我们介绍两个直观的估计量

Before-After Estimator I

	处理前	处理后
对照组 $G=0$	$Y_t(0)$	$Y_{t+1}(0)$
处理组 $G=1$	$Y_t(0)$	$Y_{t+1}(1)$

- 我们考虑以下估计量：

$$\hat{\tau}^{BA} = \bar{Y}_{1,t+1} - \bar{Y}_{1,t}$$

这里，$\bar{Y}_{1,t}$ 是处理组个体在 t 时刻的样本均值结果.

- 识别假设：$\hat{\tau}^{BA}$ 使用处理前 (在 t 时刻) 的结果作为在对照时 (在 $t+1$ 时刻) 的反事实结果的代理，即

$$\mathbb{E}\{Y_{t+1}(0) \mid G=1\} = \mathbb{E}\{Y_t(0) \mid G=1\}$$

Before-After Estimator II

	最低工资法案实施前	最低工资法案实施后
宾夕法尼亚州 $G = 0$	$Y_t(0)$	$Y_{t+1}(0)$
新泽西州 $G = 1$	$Y_t(0)$	$Y_{t+1}(1)$

- 例子：直接使用新泽西州前（t 时刻）后（$t + 1$ 时刻）的最低工资相减
- 问题：即使在没有接受任何处理的情况下，由于时间趋势、宏观因素、生命周期效应等原因，结果在 t 和 $t + 1$ 之间也可能发生变化...

Cross-Sectional Estimator I

	处理前	处理后
对照组 $G=0$	$Y_t(0)$	$Y_{t+1}(0)$
处理组 $G=1$	$Y_t(0)$	$Y_{t+1}(1)$

- 我们考虑以下估计量，将处理组的个体与同一时期的对照组个体进行比较：

$$\widehat{\tau}^{CS}=\bar{Y}_{1,t+1}-\bar{Y}_{0,t+1}$$

其中 $\bar{Y}_{1,t}\left(\bar{Y}_{0,t}\right)$ 表示处理组 (对照组) 在时间 t 时的样本平均结果.

- 识别假设：$\widehat{\tau}^{CS}$ 以对照组在 $t+1$ 时的结果作为在对照情况下 $t+1$ 时的反事实结果的代理，即

$$\mathbb{E}\left\{Y_{t+1}(0)\mid G=1\right\}=\mathbb{E}\left\{Y_{t+1}(0)\mid G=0\right\}$$

Cross-Sectional Estimator II

	最低工资法案实施前	最低工资法案实施后
宾夕法尼亚州 $G=0$	$Y_t(0)$	$Y_{t+1}(0)$
新泽西州 $G=1$	$Y_t(0)$	$Y_{t+1}(1)$

- 例子：直接使用宾夕法尼亚州的最低工资减去新泽西州的最低工资

- 问题：即使在没有接受任何处理的情况下，处理组和对照组的结果也可能因时间趋势、宏观因素、生命周期效应等而发生变化，这可能会导致结果偏误.

DID 方法 I

DID 解决方案：利用 2×2 设计和中心假设

Assumption 1 (“平行趋势”假设, parallel trends)

在没有接受处理的情况下，我们认为处理组和对照组将经历相同的趋势：

$$\mathbb{E}\left\{Y_{t+1}(0) - Y_t(0) \mid G=1\right\} = \mathbb{E}\left\{Y_{t+1}(0) - Y_t(0) \mid G=0\right\} \tag{1}$$

- 平行趋势假设(1)等价于“常数作用”(constant difference) 假设：在没有干预的情况下，处理组和对照组之间的差异在时间上是恒定的

$$\begin{aligned} & \mathbb{E}\left\{Y_{t+1}(0) \mid G=1\right\} - \mathbb{E}\left\{Y_{t+1}(0) \mid G=0\right\} && (2) \\ = & \mathbb{E}\left\{Y_t(0) \mid G=1\right\} - \mathbb{E}\left\{Y_t(0) \mid G=0\right\} && (3) \end{aligned}$$

DID 方法 II

- 平行趋势是 scale-dependent 的：对 Y 成立，但对 Y 的非线性单调变换可能不成立.

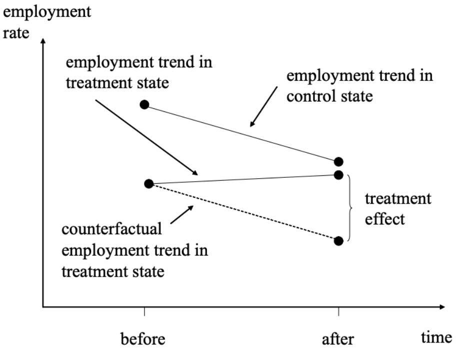

Figure 5.2.1: Causal effects in the differences-in-differences model

DID 方法 III

- 在平行趋势假设下，DID 可以非参数识别：

$$\tilde{\mu}_{0,\mathrm{DID}} = \mathbb{E}\left(Y_{it} \mid G_i = 1\right) + \mathbb{E}\left(Y_{i,t+1} \mid G_i = 0\right) - \mathbb{E}\left(Y_{it} \mid G_i = 0\right)$$

- DID 估计量计算了处理组和对照组在结果变量上的变化差异.

- 等价地，DID 估计量也可以表示为处理组和对照组前后估计量的差异：

$$\begin{aligned}\hat{\tau}_{\mathrm{DID}} &= \left(\bar{Y}_{1,t+1} - \bar{Y}_{1,t}\right) - \left(\bar{Y}_{0,t+1} - \bar{Y}_{0,t}\right) \\ &= \hat{\tau}_1^{BA} - \hat{\tau}_0^{BA}\end{aligned}$$

其中，$\bar{Y}_{g,t}$ 是时刻 t 处理组/对照组的平均结果变量.

Before-After estimator 的视角 I

- 在 SUTVA 假设和 Consistency 假设下，我们有

$$Y_t = Y_t(0), \quad Y_{t+1} = (1-G)Y_{t+1}(0) + GY_{t+1}(1),$$

因此，

$$\begin{aligned}
\mathbb{E}\left(\widehat{\tau}_0^{BA}\right) &= \mathbb{E}\left(\bar{Y}_{0,t+1} - \bar{Y}_{0,t}\right) \\
&= \mathbb{E}\left\{Y_{t+1}(0) - Y_t(0) \mid G=0\right\} \quad \text{(SUTVA)} \\
&= \mathbb{E}\left\{Y_{t+1}(0) - Y_t(0) \mid G=1\right\} \quad \text{(parallel trends)} \\
&= \mathbb{E}\left\{Y_{t+1}(0) \mid G=1\right\} - \mathbb{E}\left\{Y_t \mid G=1\right\} \ \text{(SUTVA)}
\end{aligned}$$

Before-After estimator 的视角 II

- 类似地，我们有

$$\mathbb{E}\left(\widehat{\tau}_1^{BA}\right)=\mathbb{E}\left\{Y_{t+1}(1)\mid G=1\right\}-\mathbb{E}\left\{Y_t\mid G=1\right\}$$

- 上述两项相减，我们有

$$\begin{aligned}\mathbb{E}\left(\widehat{\tau}_{\mathrm{DID}}\right)&=\mathbb{E}\left(\widehat{\tau}_1^{BA}-\widehat{\tau}_0^{BA}\right)\\&=\mathbb{E}\left\{Y_{t+1}(1)-Y_{t+1}(0)\mid G=1\right\}\\&=\tau_{\mathrm{ATT}}\end{aligned}$$

Cross-sectional estimator 的视角 I

- DID 估计的另一种写法是，将处理组和对照组在时间 t 和 $t+1$ 的横截面估计结果相减，即

$$\begin{aligned}\hat{\tau}_{\mathrm{DID}} &= \left(\bar{Y}_{1,t+1} - \bar{Y}_{0,t+1}\right) - \left(\bar{Y}_{1,t} - \bar{Y}_{0,t}\right) \\ &= \hat{\tau}_{t+1}^{CS} - \hat{\tau}_{t}^{CS}\end{aligned}$$

- 基于平行趋势假设 (式(2)中的“常数差异”解释)，可以证明上述 DID 估计量对 τ 是无偏的.

- DID 估计使用双重差异消除了单组前后估计量 (before-after estimator) 的偏差 (消除时间趋势).

Cross-sectional estimator 的视角 II

- DID 估计使用双重差异消除了单时间点横截面估计量 (cross-sectional estimator) 的偏差 (消除组别差异).

回归的视角 I

- 上文展示了 DID 估计量的非参数识别方法.

- 对于许多复杂情况 (多组、多时期、协变量)，一个参数化/建模视角是可取的.

- 实际上，在经济学文献中，DID 通常与固定效应的回归模型相关联，用于处理在未接受处理的情况下潜在结果 (例如，Angrist & Pischke，2009 年，第 5 章).

- 回归 DID 模型的核心是关于 $Y(0)$ 的可加固定效应 (additive fixed effects) 模型 (省略 X):

$$Y_{iT}(0) = \alpha + \gamma G_i + \delta_T + \epsilon_{iT} \tag{4}$$

其中 $\mathbb{E}(\epsilon_{iT} \mid G_i, T) = 0$；这里 γ 是组固定效应，δ_T 是时间效应.

回归的视角 II

- 我们记 $\alpha_i \equiv \alpha + \gamma G_i$，表示个体的时间固定效应.

- 模型(4)意味着平行趋势 (parallel trends)假设：对于 $g = 0, 1$

$$\mathbb{E}\{Y_{t+1}(0) \mid G = 0\} - \mathbb{E}\{Y_t(0) \mid G = 0\} = \delta_{t+1} - \delta_t,$$
$$\mathbb{E}\{Y_{t+1}(0) \mid G = 1\} - \mathbb{E}\{Y_t(0) \mid G = 1\} = \delta_{t+1} - \delta_t.$$

- 模型(4)意味着常数作用 (constant difference)假设：对于任何 T,

$$\mathbb{E}\{Y_T(0) \mid G = 1\} - \mathbb{E}\{Y_T(0) \mid G = 0\} = \gamma$$

回归的视角 III

- 在参数识别与估计以外，对于许多复杂情况 (多组、多周期、及协变量存在时)，模型参数化是可取的

- 除了模型(4)外，我们还假设 $Y_{iT}(1) = Y_{iT}(0) + \tau G_i \cdot 1\{T = t+1\}$.

- 上述假设一块意味着，观察到的结果变量遵循以下模型 (这里省略了 X):

$$Y_{iT} = \underbrace{\alpha + \gamma G_i}_{\alpha_i} + \delta_T + \underbrace{\tau \cdot Z_{iT}}_{\tau G_i \cdot 1\{T=t+1\}} + \epsilon_{iT} \tag{5}$$

其中 $\mathbb{E}(\epsilon_{iT} \mid G_i, T) = 0$.

回归的视角 IV

	政策实施前 $T=t$	政策实施后 $T=t+1$	Difference
处理组 $G=1$	$\alpha+\gamma+\delta_t$	$\alpha+\gamma+\delta_{t+1}$	$\delta_{t+1}-\delta_t$
对照组 $G=0$	$\alpha+\delta_t$	$\alpha+\delta_{t+1}+\tau$	$\delta_{t+1}-\delta_t+\tau$
Difference	$-\gamma$	$-\gamma+\tau$	τ

- 模型(5)中的系数 τ 是 ATT 的估计量.

$$Y_{iT}(1)=\alpha_i+\delta_T+\tau+\epsilon_{iT}$$

回归的视角 V

- 易证明双重差分模型(5)得到 τ_{ATT}：

$$\tau = \mathbb{E}\left\{\left(\bar{Y}_{1,t+1} - \bar{Y}_{1,t}\right) - \left(\bar{Y}_{0,t+1} - \bar{Y}_{0,t}\right)\right\} = \tau_{\mathrm{ATT}}$$

- 因此，上述回归模型符合非参数 DID 识别的思想.

因果解释 I

- 模型(5)转化为潜在结果 $Y(1)$ 的模型

$$Y_{iT}(1) = \alpha_i + \delta_T + \tau + \epsilon_{iT} \tag{6}$$

- 模型(6)(与模型(4)一起) 实际上假定所有个体的处理效应 τ 在所有 T 时刻上都是同质的 (homogeneous treatment effect).

$$\tau = Y_{iT}(1) - Y_{iT}(0) = \mathbb{E}\{Y_{iT}(1) - Y_{iT}(0)\}$$

- 因此，τ 是一个因果效应 τ 是 ATT 还是 ATE？这取决于对哪个人群强加模型(6).

- 尽管模型(6)适用于所有个体，但仅在处理组个体上应用即可识别 ATT，这是非参数识别下 DID 估计的估计量。

因果解释 II

- 实际上，由于在第二个 (处理) 时期仅处理组个体的 $Y(1)$ 是可观测的，无论模型假设如何，人们只能针对该子群体 (ATT) 做出因果推断。

- 在非参数识别中，固定效应模型(4)-(6)中的可加性假设 (即组特定和时间特定效应的可加性) 不是必需的。

	New Jersey	Pennsylvania	Difference
February	20.44	23.33	-2.89
November	21.03	21.17	-0.14
Change	0.59	-2.16	2.75

表: Card and Krueger (1994) example

引言 I

- 合成控制方法最初由 Abadie & Gardeazabal(2003)，Abadie, Diamond & Hainmuller(2010) 提出，旨在估计 aggregate interventions 的效果。
- “Aggregate interventions” 意思是指影响大个体 (如城市、地区或国家) 的干预措施，而不是影响单个个体或小个体的干预措施。例如，政府的全国性政策改变可以被认为是一种聚合干预，因为它们影响整个国家的人口和经济。这些干预措施的效果通常是通过对大个体的总体指标进行比较来衡量的，例如整个城市的平均失业率或 GDP 增长率。
- 许多有趣的事件或干预发生在 aggregate level 上，如城市、地区或国家。
- “Aggregate level” 意思是指在较高的统计层面上进行测量、分析或干预的范围。例如，一个国家的 GDP 是在国家级别上进行测量和分析的，而不是在单个城市或家庭层面上。在这种情况下，国家是聚合层面，而城市和家庭则是较低的个体层面。
- 即使在实验设置中，微观干预也可能是不可行的 (如公平性) 或无效的 (如干扰)。

引言 II

- 当分析的个体为少数几个聚合实体时，与任何单个比较个体相比，组合比较个体（“合成对照组”）往往能更好地再现接受处理的个体的特征。
- 在合成控制方法中，比较组是由所有潜在的比较组的加权平均值组成的，该加权平均值最能够反映接受处理个体的特征。

Setup I

- 假设我们观察到 $J+1$ 个个体在 $1, 2, \ldots, T$ 个时期中。
- 个体“1”在 $T_0+1, \ldots, T$ 期间接受干预。
- 剩下的 J 个个体未接受处理。
- 让 Y_{it}^N 表示在没有干预的情况下，个体 i 在时间 t 的结果。
- 让 Y_{it}^I 表示如果个体 i 在时期 T_0+1 到 T 中接受干预，那么在时间 t 个体 i 的结果将会是什么。
- 我们的目标是估计 ATT，即处理组因果作用，

$$\tau_{1t} = Y_{1t}^I - Y_{1t}^N = Y_{1t} - Y_{1t}^N$$

对于 $t > T_0$，其中 Y_{1t} 是个体 1 在时间 t 的结果。

Setup II

- Abadie, Diamond & Hainmuller(2010) 考虑以下因子模型 ($t = 1, \dots, T$):

$$Y_{it}(0) = \mu_i' \lambda_t + \delta_t + X_i' \beta + \epsilon_{it}, \tag{7}$$

其中 μ_i 是 r 维度的未观测混杂向量，λ_t 是相应的时间变化系数，δ_t 是未观测的时间混杂因素，X_i 是观测到的 p-维协变量.

- 注意未观测混杂变量是可分的 (unobserved confounding is separable).
- 且具有随时间变化的系数 (λ_t, time-varying effects)

Setup III

- 模型(7)推广了通常的双重差分固定效应模型 (DID)，其中 $\mu_i'\lambda_t$ 取代了式(4)中的 $\alpha_i = \alpha + \gamma G_i$，

$$Y_{it}(0) = \underbrace{\alpha + \gamma G_i}_{\alpha_i} + \delta_t + X_i'\beta + \epsilon_{it}, \qquad (4)$$

  也常在经济学中被称为交互固定效应模型 (interactive fixed effects model)，本质上是潜在因子模型 (latent factor model).

- 合成控制法 (Synthetic Control Method) 可用于估计因果效应.
- 设 $W = (w_2, \ldots, w_{J+1})'$，其中 $w_j \geq 0$，$\sum_{j=2}^{J+1} w_j = 1$。每一组 W 都代表一个潜在的合成控制组。

Setup IV

- 假设我们能够选择 W^*，使得处理组的处理前协变量和结果得以复制，即

$$\sum_{j=2}^{J+1} w_j^* X_j = X_1, \quad \sum_{j=2}^{J+1} w_j^* Y_{j1} = Y_{11}, \ldots, \quad \sum_{j=2}^{J+1} w_j^* Y_{jT_0} = Y_{1T_0}$$

即

$$\underbrace{\begin{pmatrix} X_{1,1} \\ \vdots \\ X_{1,p} \\ Y_{1,1} \\ \vdots \\ Y_{1,T_0} \end{pmatrix}}_{V_1} = w_2^* \underbrace{\begin{pmatrix} X_{2,1} \\ \vdots \\ X_{2,p} \\ Y_{2,1} \\ \vdots \\ Y_{2,T_0} \end{pmatrix}}_{V_2} + \ldots + w_{J+1}^* \underbrace{\begin{pmatrix} X_{J+1,1} \\ \vdots \\ X_{J+1,p} \\ Y_{J+1,1} \\ \vdots \\ Y_{J+1,T_0} \end{pmatrix}}_{V_{J+1}}$$

Setup V

理论上，最优向量应该满足：

$$(w_2^*, \ldots, w_{J+1}^*) = \underset{(w_2, \ldots, w_{J+1})}{\operatorname{argmin}} \|V_1 - w_2 V_2 - \ldots w_{J+1} V_{J+1}\|_2$$

In practice, these may hold only approximately.

- 实际的困难：在假设因子模型(7)成立时，并且合成控制的预处理理拟合是精确的，上述估计方法是否有效？
- 解决的关键办法：推导出因果作用误差的表达式。

解决方案

如果 $\sum_{t=1}^{T_0} \lambda_t' \lambda_t$ 是非奇异的，那么

$$Y_{1t}^N - \sum_{j=2}^{J+1} w_j^* Y_{jt} = \sum_{j=2}^{J+1} w_j^* \sum_{s=1}^{T_0} \lambda_t \left(\sum_{n=1}^{T_0} \lambda_n' \lambda_n \right)^{-1} \lambda_s' (\varepsilon_{js} - \varepsilon_{1s}) - \sum_{j=2}^{J+1} w_j^* (\varepsilon_{jt} - \varepsilon_{1t})$$

上述公式的验证细节可见 Abadie, Diamond & Hainmuller(2010) 附录 B.

- 在假设因子模型(7)和一些标准条件下，如果前期时期数 ($T \to \infty$) 相对于残差方差而言很大，那么可以表明 $Y_{1t}^N - \sum_{j=2}^{J+1} w_j^* Y_{jt} \approx 0$。

- 因此，τ_{1t} 的一个近似无偏估计量是

$$\hat{\tau}_{1t} = Y_{it} - \sum_{j=2}^{J+1} w_j^* Y_{jt}, \quad t = T_0 + 1, \ldots, T.$$

Recent developments

- 合成控制因其简单性和直观性而广受欢迎。
- 但是，合成控制处理所想解决的问题本质上非常具有挑战性：只有一个或最多几个处理个体–信息极度缺乏。
- 最近的发展：
  - 放宽合成控制的凸性约束 (Doudchenko& Imbens，2016).
  - 结合合成控制和 DID：synthetic DID (Arkhangelsky et al., 2018).
  - 矩阵填补 (Matrix completion) 方法 (Athey et al., 2017).
  - Augmented SC(Ben-Michael,Feller, & Rothstein, 2018).

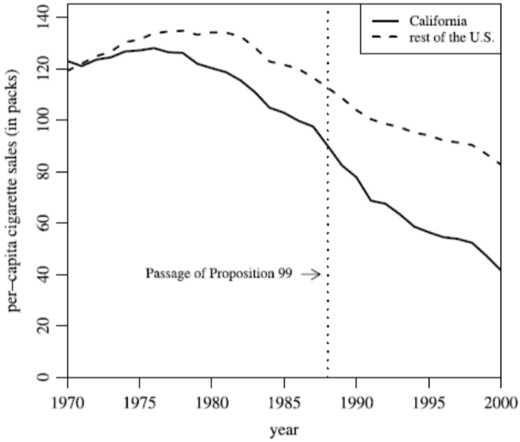

图: Trends in per-capita cigarette sales: California vs. the rest of the United States

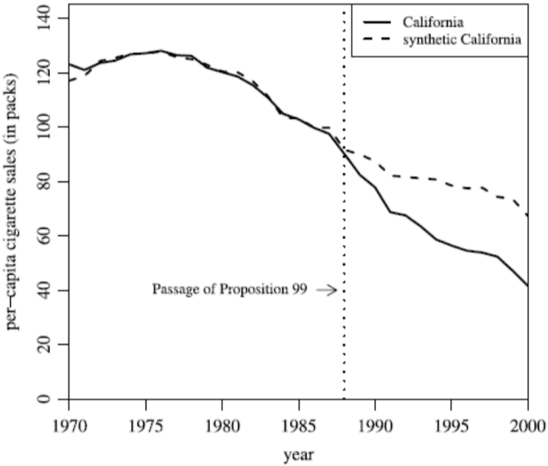

图: Trends in per-capita cigarette sales: California vs. the rest of the United States.

作业 I

1. 上机作业：利用 DID 方法分析数据集 rdat1，要求提供数据预处理及分析结论
   - 处理组：Basque 地区，编号为 rdat$regionno=17
   - 对照组：Basque 地区以外地区，编号为 rdat$regionno=c(2:16, 18)
   - 干预时间点：1969 年，即认为 1955-1969 所有地区都未施加干预，1970 以后施加干预
2. 上机作业：利用 DID 方法分析数据集 rdat2，要求提供数据预处理及分析结论
   - 处理组：Basque 地区，编号为 rdat$regionno=17
   - 对照组：Basque 地区以外地区，编号为 rdat$regionno=c(2:16, 18)
   - 干预时间点：1969 年，即认为 1955-1969 所有地区都未施加干预，1970 以后施加干预
   - 利用协变量信息
3. 上述作业任选其一，2 为附加题

作业 II

1. 证明题：Never-taker 人群的比例在可忽略性和单调性假设下可识别，即 $\mathrm{pr}(S_0 = 0, S_1 = 0 \mid X)$，要求写清楚假设和计算过程。
2. 证明题：Never-taker、Complier 人群的比例在可忽略性和单调性假设下可识别，即 $\mathrm{pr}(S_0 = 0, S_1 = 0 \mid X)$ 和 $\mathrm{pr}(S_0 = 0, S_1 = 1 \mid X)$，要求写清楚假设和计算过程。
3. 上述作业任选其一，2 为附加题