

# 面板数据因果推断

## 第6讲 合成控制法及其拓展

徐轶青

斯坦福大学

# 课程安排

- 合成控制法

- › 快速回顾
- › 替代算法

- 研究进展

- › 合成双重差分法 (Synthetic DID)
- › 潜因子模型 (The latent factor approach)

# 合成控制法 (SCM)

## SCM：基本思想

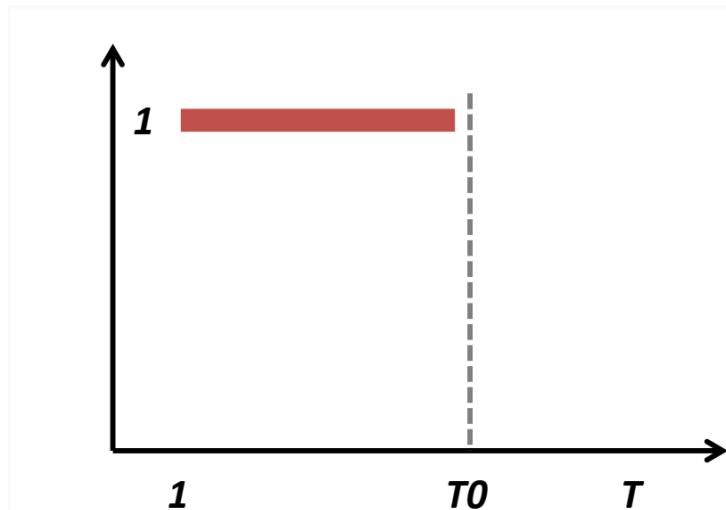
- 在时期  $1, 2, \dots, T$  存在  $J+1$  个观测单位；“1” 属于处理组，其他  $J$  个单位属于控制组

## SCM：基本思想

- 在时期  $1, 2, \dots, T$  存在  $J+1$  个观测单位；“1” 属于处理组，其他  $J$  个单位属于控制组
- 在时期  $T_0$  后，“1” 对应的部分受到干预

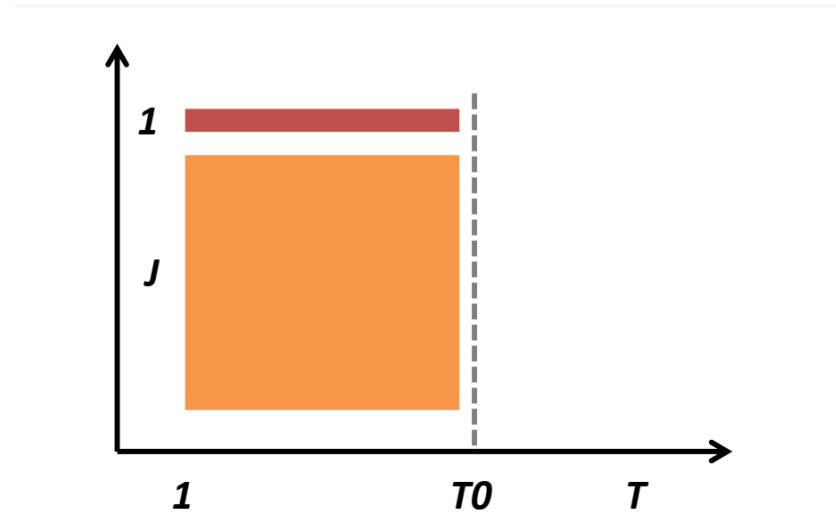
## SCM：基本思想

- 在时期  $1, 2, \dots, T$  存在  $J+1$  个观测单位；“1” 属于处理组，其他  $J$  个单位属于控制组
- 在时期  $T_0$  后，“1” 对应的部分受到干预



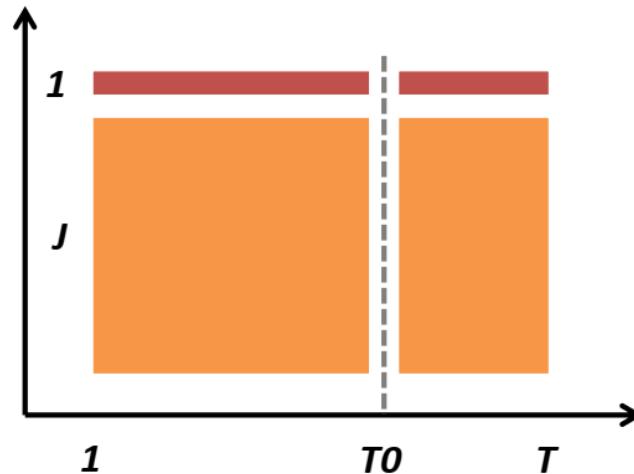
## SCM: 基本思想

- 在时期  $1, 2, \dots, T$  存在  $J+1$  个观测单位; “1” 属于处理组, 其他  $J$  个单位属于控制组
- 在时期  $T_0$  后, “1” 对应的部分受到干预



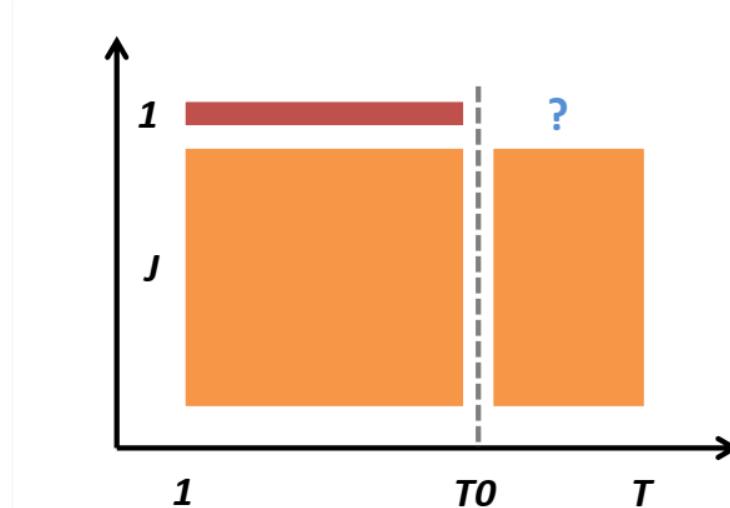
## SCM：基本思想

- 在时期  $1, 2, \dots, T$  存在  $J+1$  个观测单位；“1” 属于处理组，其他  $J$  个单位属于控制组
- 在时期  $T_0$  后，“1” 对应的部分受到干预
- 我们的目标是估计 “1” 这部分对应的干预效果



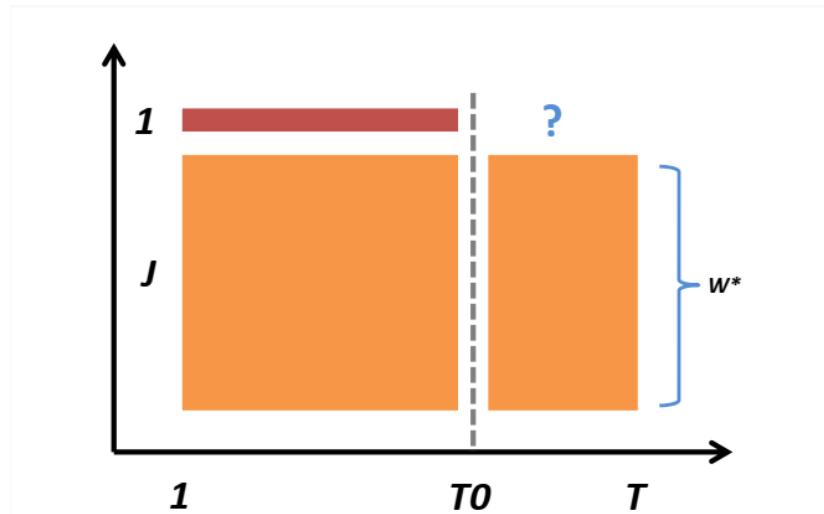
## SCM：基本思想

- 在时期  $1, 2, \dots, T$  存在  $J+1$  个观测单位；“1” 属于处理组，其他  $J$  个单位属于控制组
- 在时期  $T_0$  后，“1” 对应的部分受到干预
- 我们的目标是估计 “1” 这部分对应的干预效果



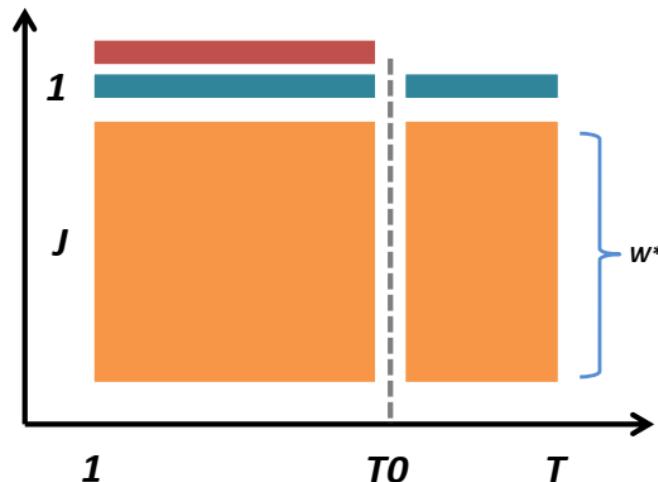
## SCM: 基本思想

- 在时期  $1, 2, \dots, T$  存在  $J+1$  个观测单位; “1” 属于处理组, 其他  $J$  个单位属于控制组
- 在时期  $T_0$  后, “1” 对应的部分受到干预
- 我们的目标是估计 “1” 这部分对应的干预效果



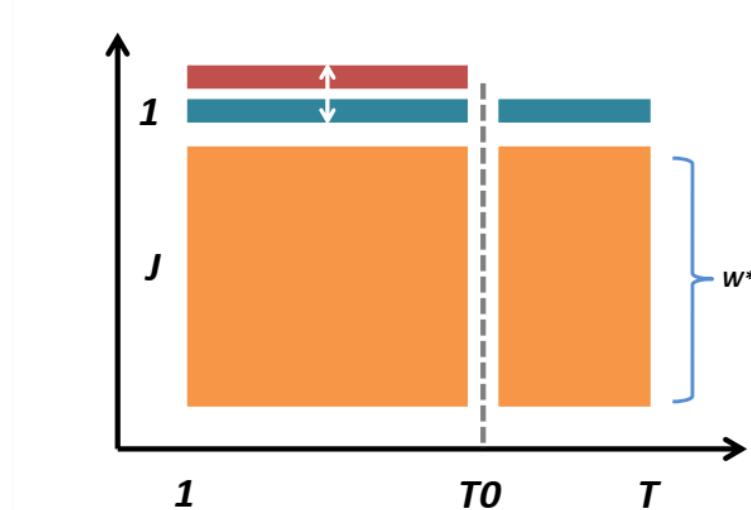
## SCM: 基本思想

- 在时期  $1, 2, \dots, T$  存在  $J+1$  个观测单位; “1” 属于处理组, 其他  $J$  个单位属于控制组
- 在时期  $T_0$  后, “1” 对应的部分受到干预
- 我们的目标是估计 “1” 这部分对应的干预效果



## SCM: 基本思想

- 在时期  $1, 2, \dots, T$  存在  $J+1$  个观测单位; “1” 属于处理组, 其他  $J$  个单位属于控制组
- 在时期  $T_0$  后, “1” 对应的部分受到干预
- 我们的目标是估计 “1” 这部分对应的干预效果



# SCM：深入了解

- Athey and Imbens (2016): “可以称为过去15年政策评估文献中最重要的创新。”

# SCM：深入了解

- Athey and Imbens (2016): “可以称为过去15年政策评估文献中最重要的创新。”
- 多方面创新的结合

# SCM：深入了解

- Athey and Imbens (2016): “可以称为过去15年政策评估文献中最重要的创新。”
- 多方面创新的结合
  - 利用处理前的结果变量

# SCM：深入了解

- Athey and Imbens (2016): “可以称为过去15年政策评估文献中最重要的创新。”
- 多方面创新的结合
  - 利用处理前的结果变量
  - 使用横截面的相关性而不是时间上的相关性

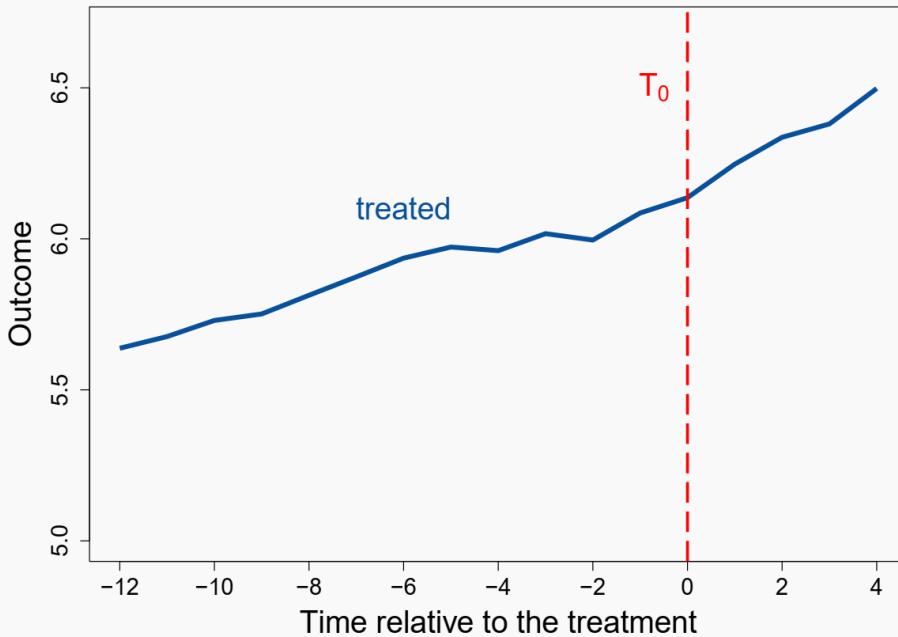
# SCM：深入了解

- Athey and Imbens (2016): “可以称为过去15年政策评估文献中最重要的创新。”
- 多方面创新的结合
  - 利用处理前的结果变量
  - 使用横截面的相关性而不是时间上的相关性
  - 使用选中控制组个体的凸组合来构造反事实

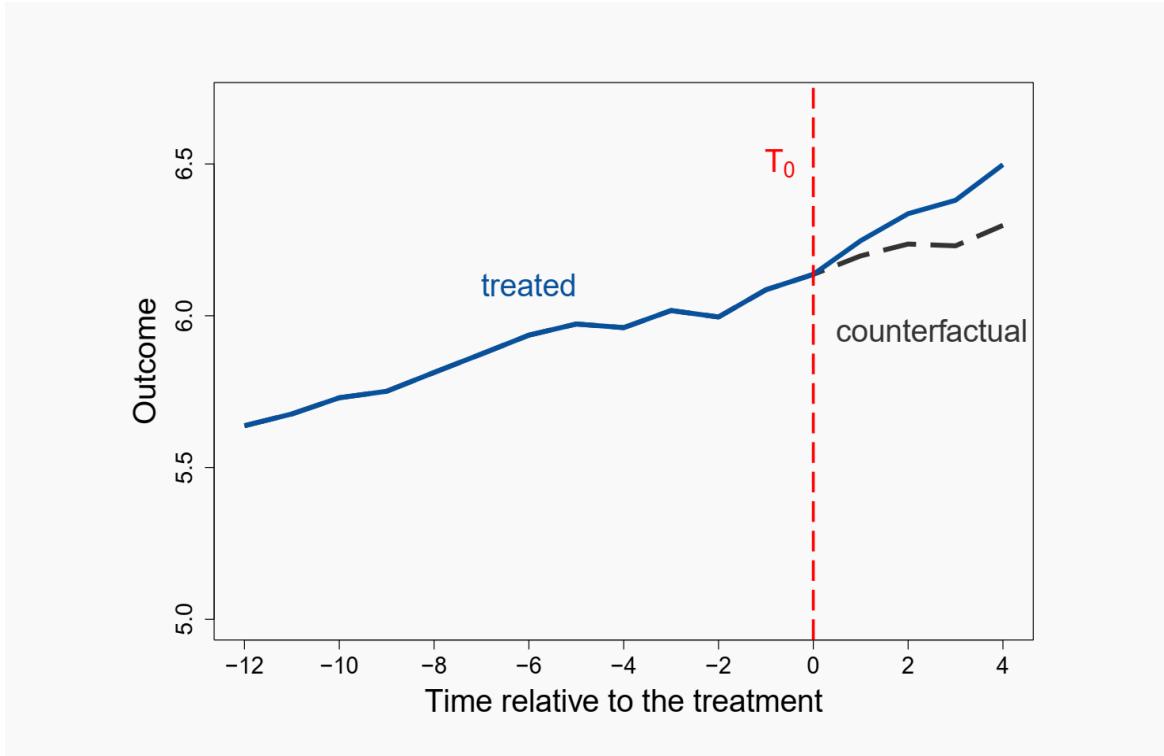
# SCM：深入了解

- Athey and Imbens (2016): “可以称为过去15年政策评估文献中最重要的创新。”
- 多方面创新的结合
  - 利用处理前的结果变量
  - 使用横截面的相关性而不是时间上的相关性
  - 使用选中控制组个体的凸组合来构造反事实
  - 保留干预前的时期以供检验

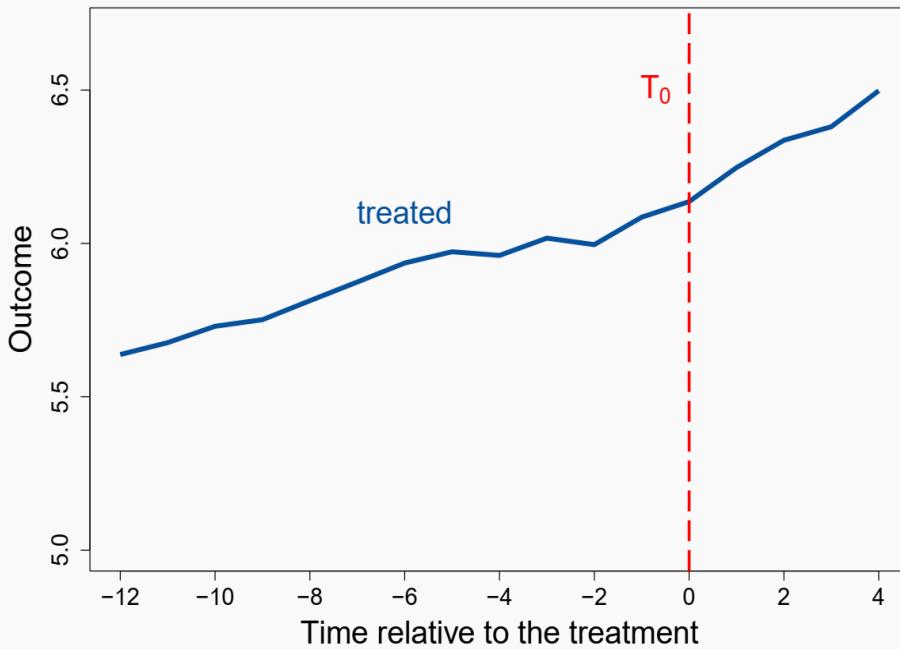
# 直觉解释



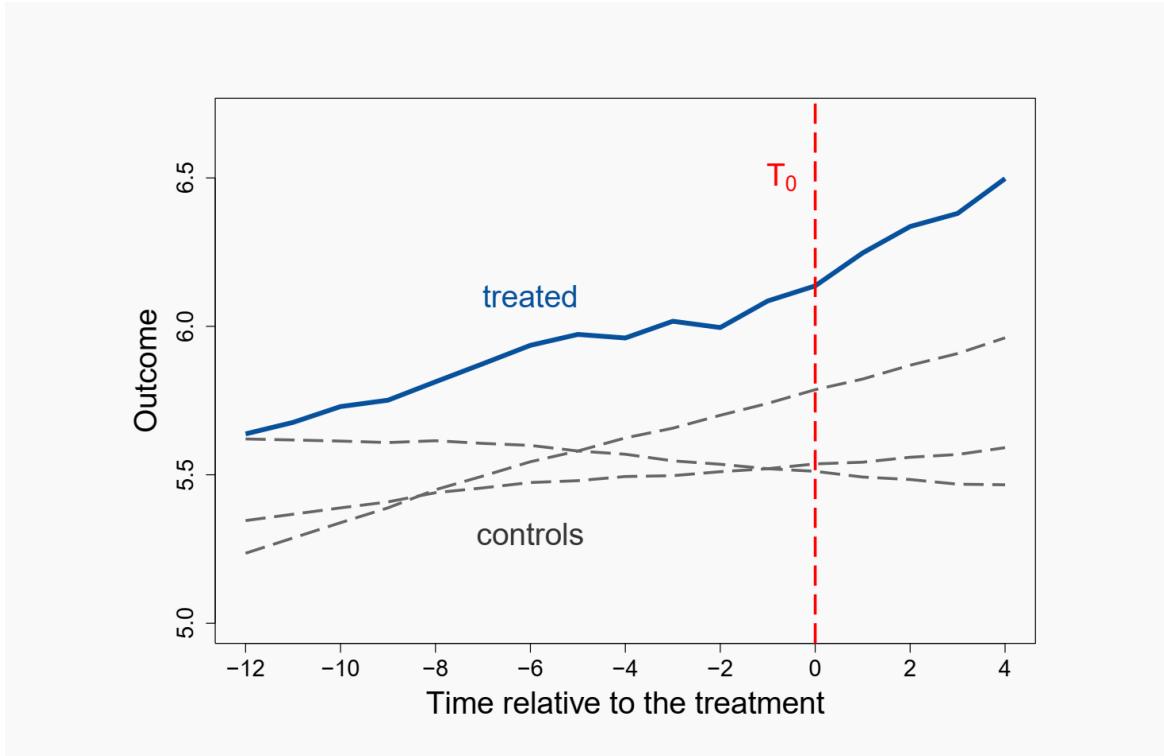
# 直觉解释



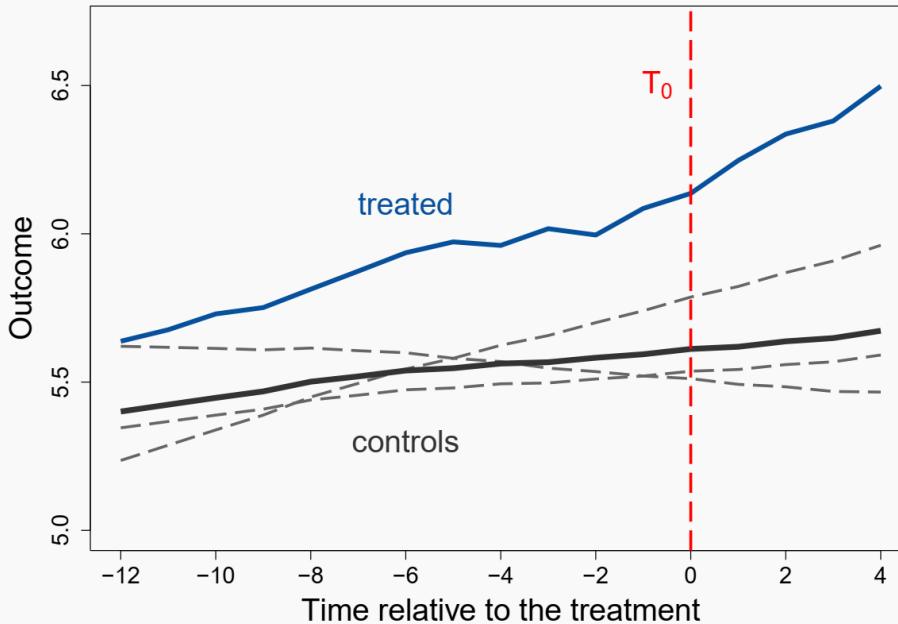
# 直觉解释



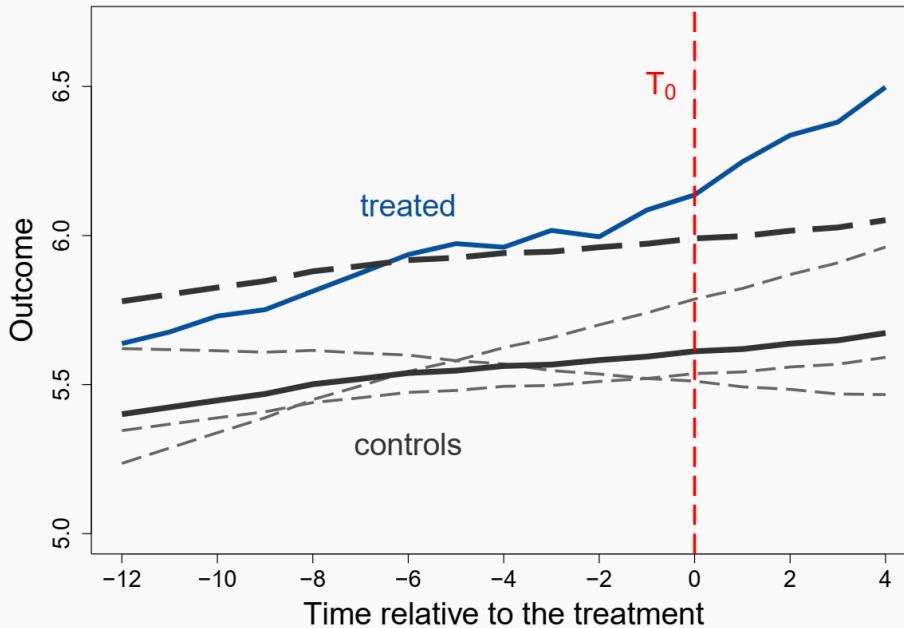
# 直觉解释



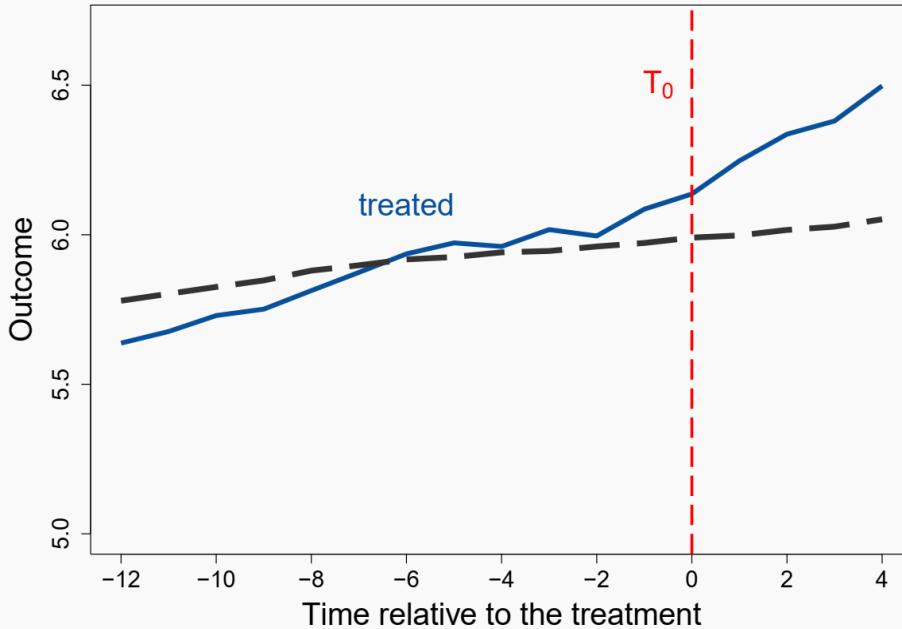
# 双重差分法 (DID)



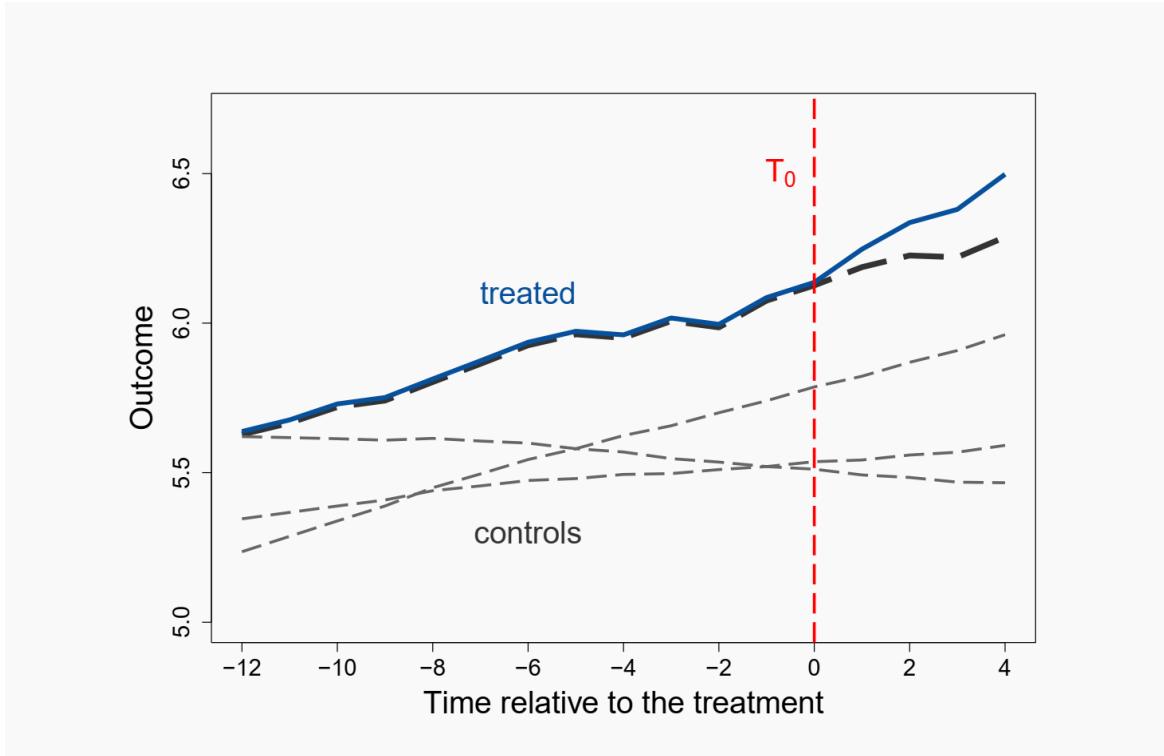
# 双重差分法 (DID)



# 双重差分法 (DID)



# SCM (和各种拓展)



# 理论推导

$$Y_{it} = \tau_{it} D_{it} + \theta_t' Z_i + \xi_t + \lambda_i' f_t + \varepsilon_{it}$$

## 理论推导

$$Y_{it} = \tau_{it} D_{it} + \theta_t' Z_i + \xi_t + \lambda_i' f_t + \varepsilon_{it}$$

or

$$\begin{cases} Y_{it}(0) = \theta_t' Z_i + \xi_t + \lambda_i' f_t + \varepsilon_{it} \\ Y_{it}(1) = Y_{it}^0 + \tau_{it} \end{cases}$$

# 理论推导

$$Y_{it} = \tau_{it} D_{it} + \theta_t' Z_i + \xi_t + \lambda_i' f_t + \varepsilon_{it}$$

or

$$\begin{cases} Y_{it}(0) = \theta_t' Z_i + \xi_t + \lambda_i' f_t + \varepsilon_{it} \\ Y_{it}(1) = Y_{it}^0 + \tau_{it} \end{cases}$$

- 假设存在 R 个随时间变化的因子  $f_t$
- 每个单位（如国家、参与者）依据因子载荷  $\lambda_i$  选取这些因子的某个固定线性组合
- 对于处理组和控制组，这些“混杂变量”在干预前的特征都得到呈现，因而我们可以使用这种信息来“平衡”混淆因素
- 我们将在稍后讨论基于模型的方法

# 理论推导

$$\begin{cases} Y_{it}(0) = \theta'_t Z_i + \xi_t + \lambda'_i f_t + \varepsilon_{it} \\ Y_{it}(1) = Y_{it}^0 + \tau_{it} \end{cases}$$

- 令  $W = (w_2, \dots, w_{J+1})'$  对于  $w_j \geq 0$ , 并且  $w_2 + \dots + w_{J+1} = 1$  .
- 令  $\bar{Y}_i^{K_1}, \dots, \bar{Y}_i^{K_M}$  干预前结果变量的 M 个线性组合 ( $M > R$ )
- 假设我们可以选择  $W^*$  满足:

$$Z_1 = \sum_{j=2}^{J+1} w_j^* Z_j, \bar{Y}_1^k = \sum_{j=2}^{J+1} w_j^* \bar{Y}_j^k, k \in \{K_1, \dots, K_M\}$$

- 当  $T_0$  足够大时,  $\tau_{1t}$  的一个近似无偏估计量可以表示为:

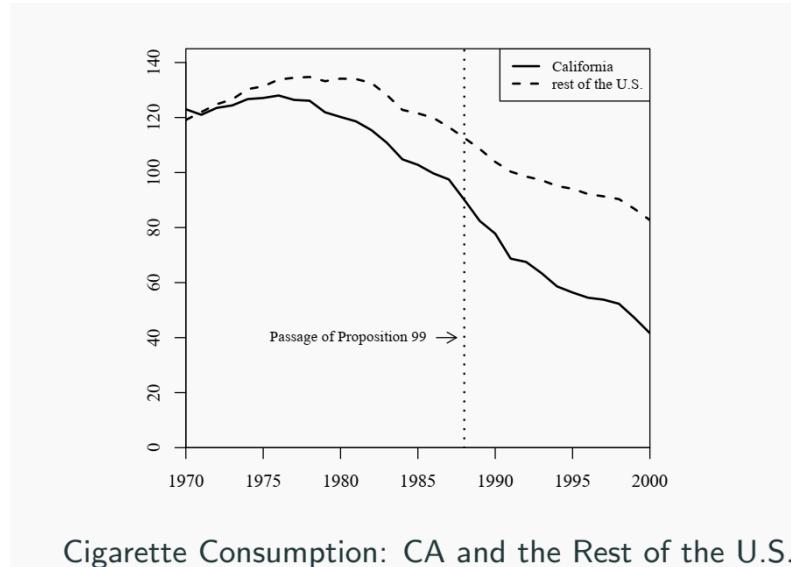
$$\hat{\tau}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}, t \in \{T_0 + 1, \dots, T\}$$

# 操作方法

- 令  $X_1 = (Z_1, \bar{Y}_i^{K_1}, \dots, \bar{Y}_i^{K_M})'$  为处理组干预前特征的一个  $(k \times 1)$  向量,  
 $X_0$ , 则是控制组的  $(k \times J)$  阶矩阵
- 向量  $W^*$  的选择目标就是在权重约束下最小化  $\|X_1 - X_0 W\|$ 
  - 我们考虑  $\|X_1 - X_0 W\|_V = \sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)}$  其中  $V$  是某个  $(k \times k)$  阶对称半正定矩阵
  - 可以有多种方式选择  $V$  (依据对  $X$  预测能力, 回归模型, MSPE最小化, 交叉验证等方面主观评价)

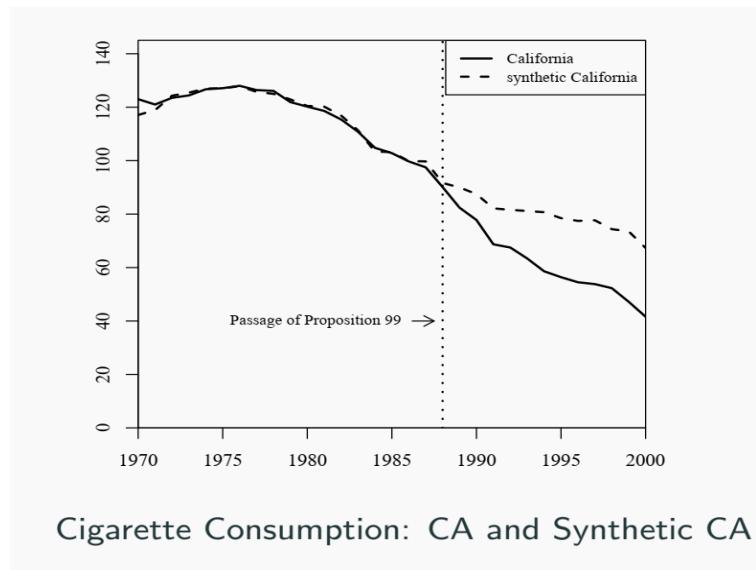
# 案例：基于烟草消费的99号修正案

- 1988年，加利福尼亚首先通过了全面的烟草管制法案（香烟消费税，媒体宣传等）
- 同时使用其他38个从未通过这类项目的州作为控制组



# 案例：基于烟草消费的99号修正案

- 1988年，加利福尼亚首先通过了全面的烟草管制法案（香烟消费税，媒体宣传等）
- 同时使用其他38个从未通过这类项目的州作为控制组

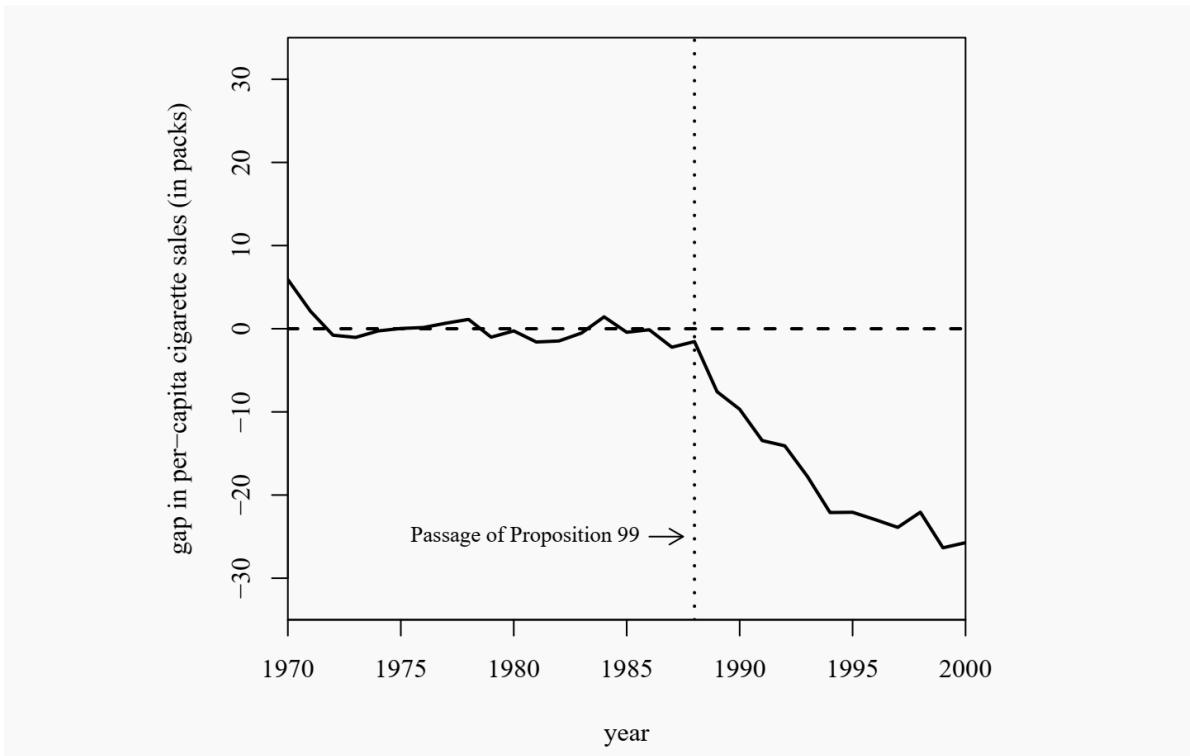


# 预测均值: 真实的 vs. 合成的加利福尼亚

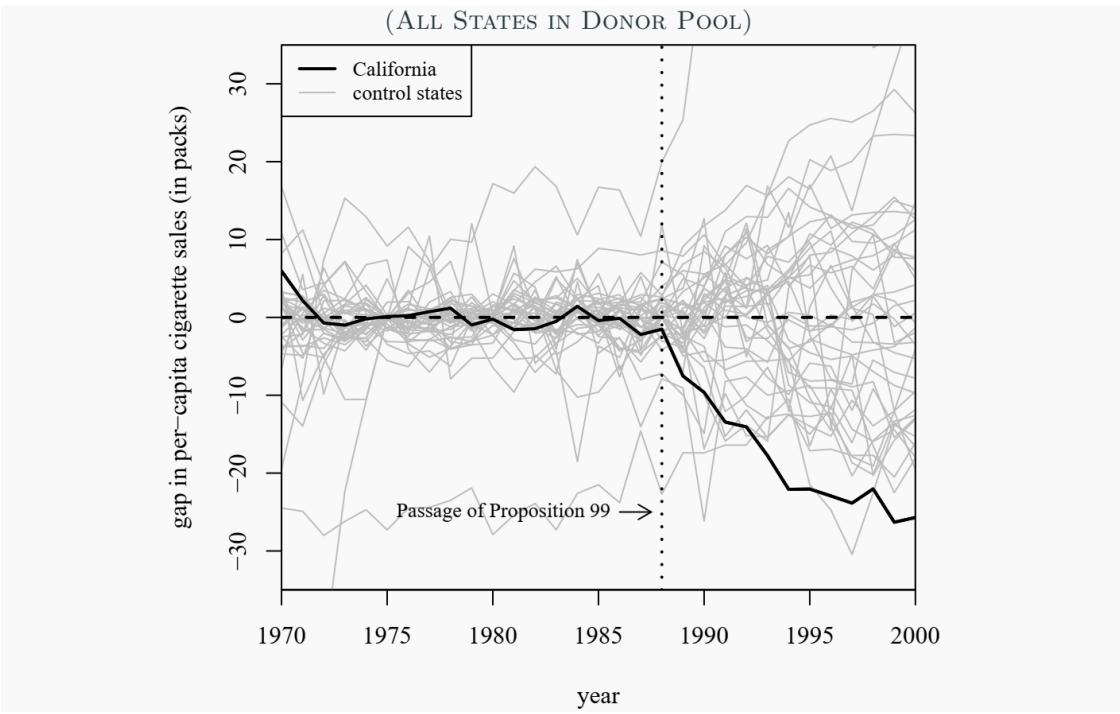
Variables	California		Average of
	Real	Synthetic	38 control states
Ln(GDP per capita)	10.08	9.86	9.86
Percent aged 15-24	17.40	17.40	17.29
Retail price	89.42	89.41	87.27
Beer consumption per capita	24.28	24.20	23.75
Cigarette sales per capita 1988	90.10	91.62	114.20
Cigarette sales per capita 1980	120.20	120.43	136.58
Cigarette sales per capita 1975	127.10	126.99	132.81

Note: All variables except lagged cigarette sales are averaged for the 1980-1988 period (beer consumption is averaged 1984-1988).

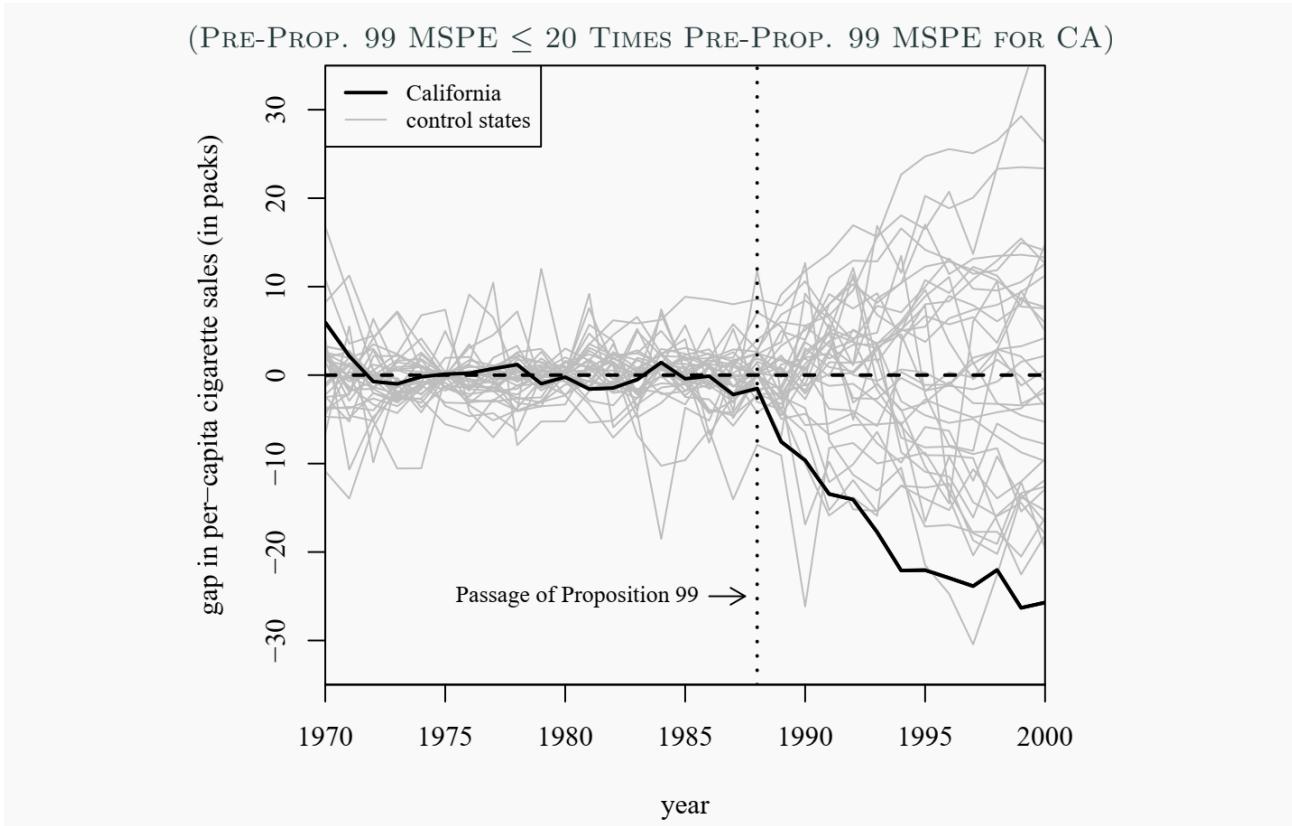
# 加州与“合成的”加州的香烟消费差异



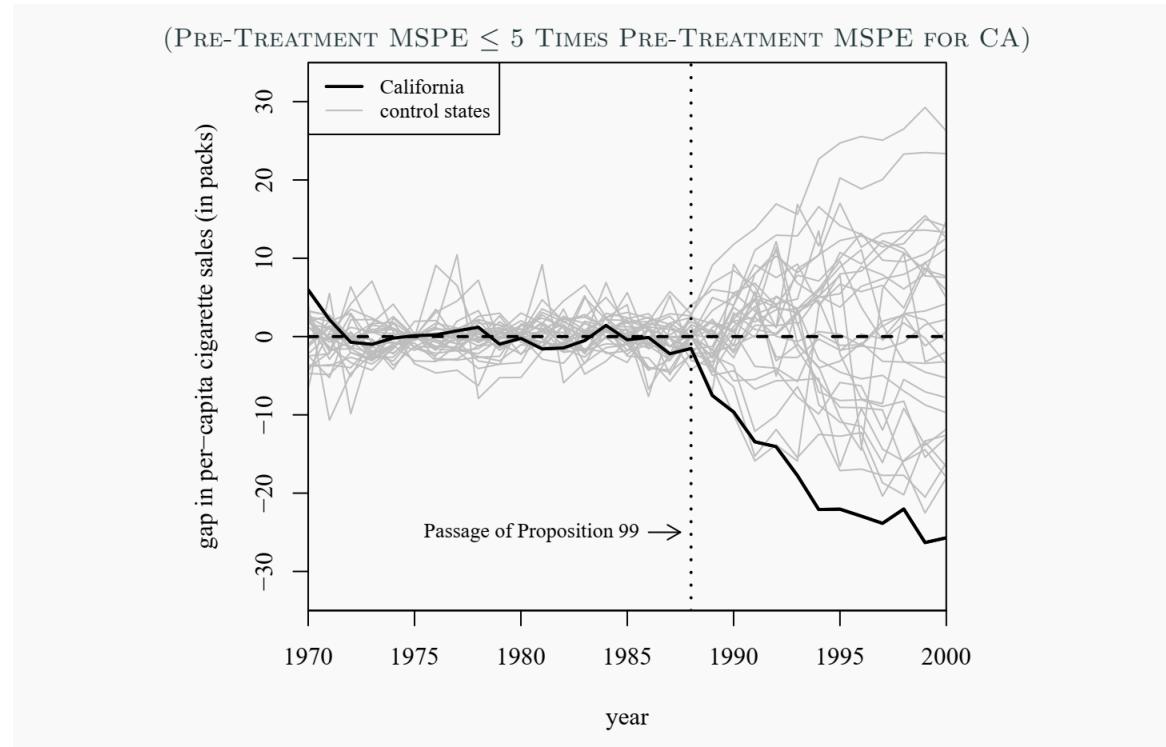
# 加州和 38 个控制州的香烟消费差异



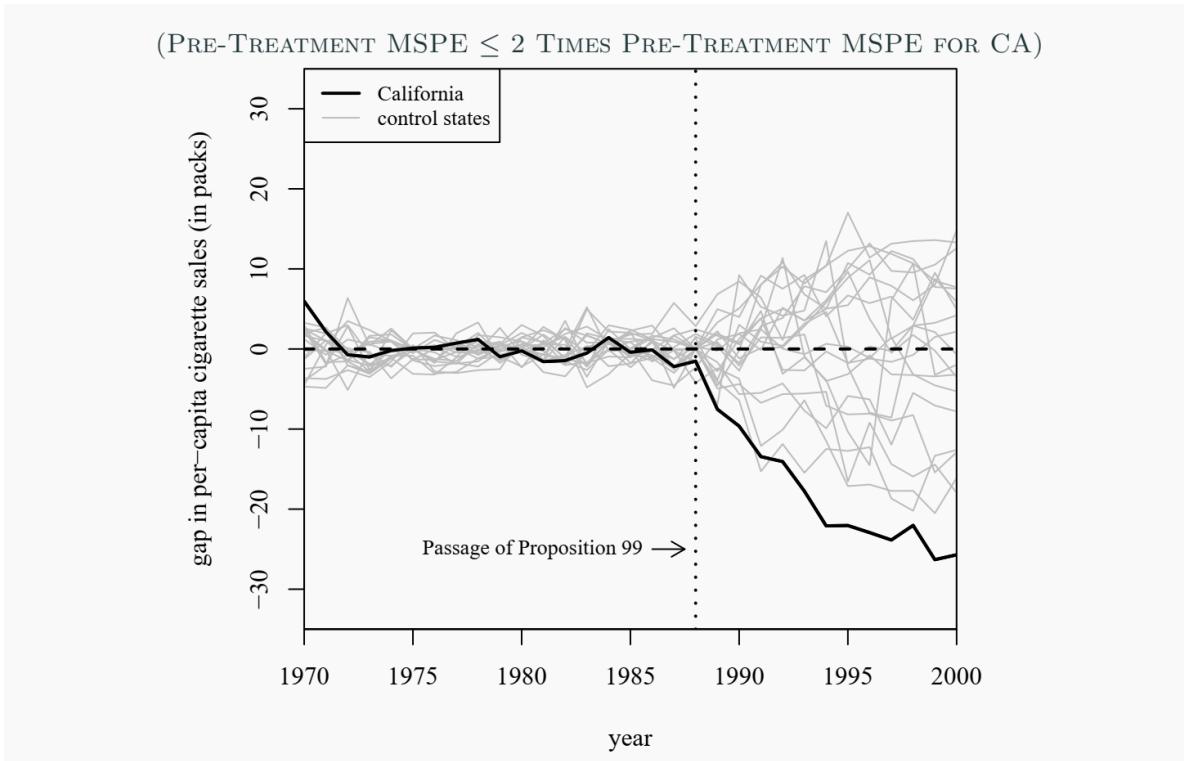
# 加州和 34 个控制州的香烟消费差异



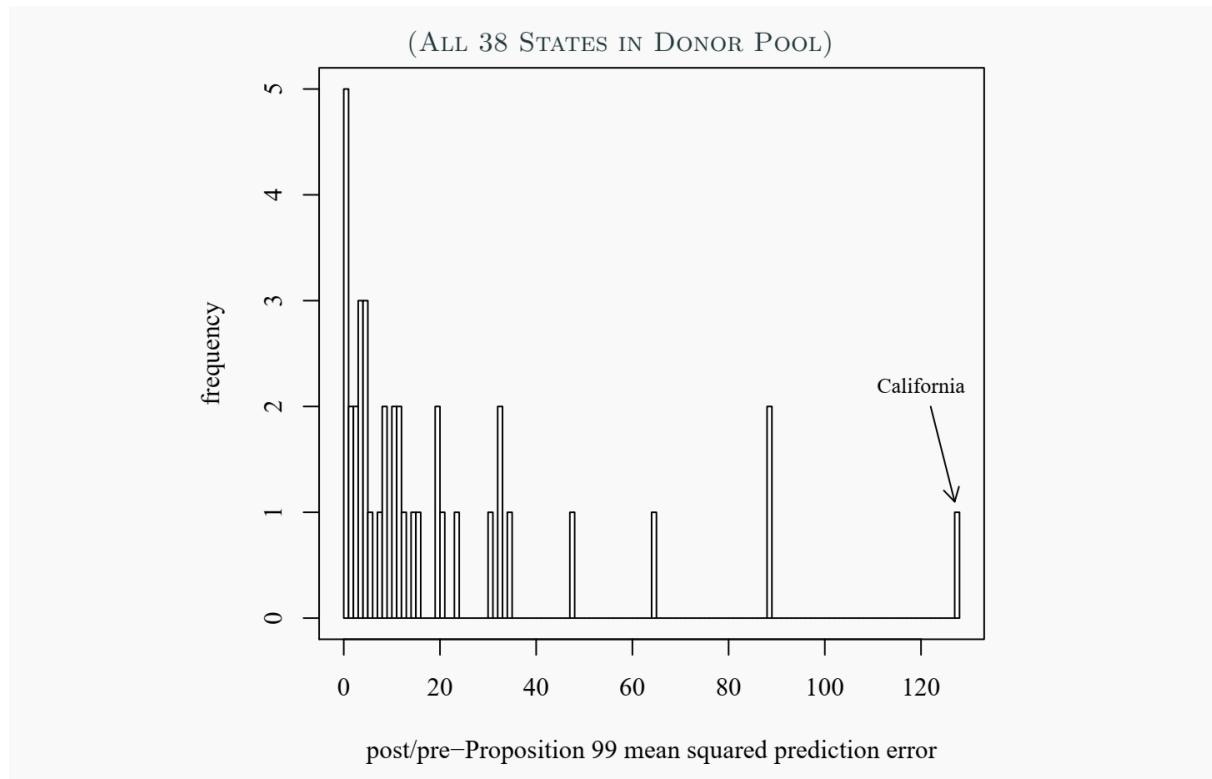
# 加州和 29 个控制州的香烟消费差异



# 加州和 19 个控制州的香烟消费差异



# 处理后 MSPE 与处理前 MSPE 的比值



# 局限性：

- 方法
  - 每次只能有一个处理组单位
  - 每次只能针对一个结果变量
  - 运行缓慢难以求解
  - 使用者有很大操作空间，如精心挑选的  $\bar{Y}_i^k$  导致过度拒绝 (Ferman et al. 2017)
- 推断
  - 置换检验和敏感性分析（例如 Hahn and Shi 2016; Firpo et al. 2017; Chernochukov 2017）
  - 使用非平稳数据会出现精度膨胀（Inflated precision）问题 (Cattaneo et al. 2019)
- 识别
  - 我的观点：本质上讲，基于严格外生性的方法（由  $\lambda_i$  引起的干扰）

# 替代性算法

# 平衡、回归、双重差分与合成控制法

面板数据方法可以分为三个类别：

- 双重差分 (DID):  $\Delta Y^{Post} - \Delta Y^{Pre}$
- 匹配 (Matching): 保证干预前的结果变量和其他协变量的平衡性
- 合成控制法 (SCM): 对每一个干预个体，使用控制组的加权平均构造一个“合成的控制组”，选择控制组的条件是加权平均值与处理前的结果变量和协变量相匹配

$$\hat{Y}_{1,T}(0) = \mu + \sum_{i=1}^n \omega_i \cdot Y_{i,T}^{obs}$$

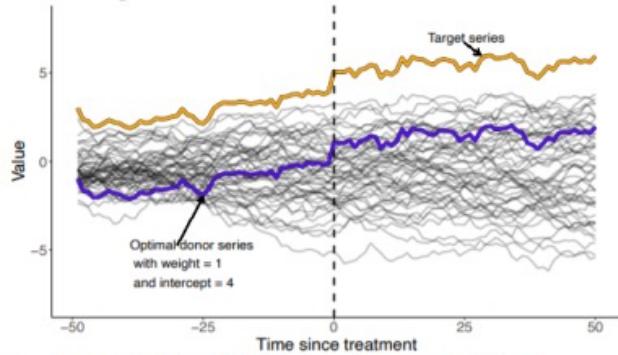
# 共同结构：四个约束

- 施加四个约束：
  1. 无截距项： $\mu = 0$ ，强于 DID 的平行趋势假设
  2. 权重之和为 1： $\sum_{i=1}^n \omega_i = 1$ ，常见于 DID 和 SCM
  3. 非负性： $\omega_i \geq 0 \forall i$ ，保证唯一性，但负权重可能会改善样本外预测
  4. 等权重：
- DID 施加假定 2-4  $\omega_i = \bar{\omega}, \forall i$
- SCM 施加假定 1-3：“凸包” (Convex Hull)

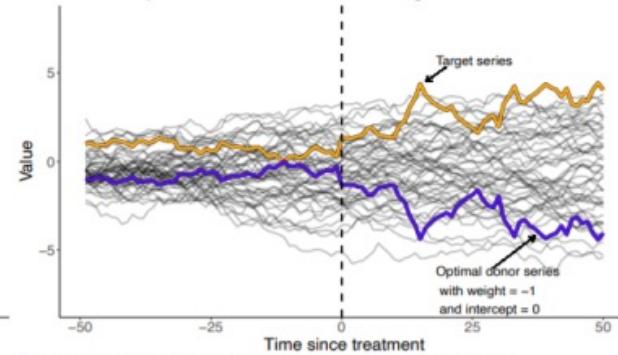
# 非凸或者负权重情形

Hollingsworth and Wing (2020)

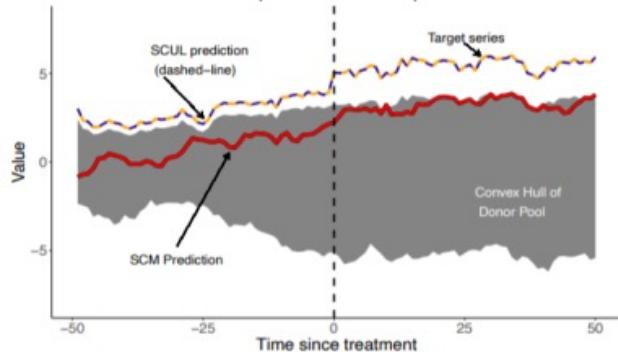
A Case 1: No convex combination of the donor pool can equal the target time series



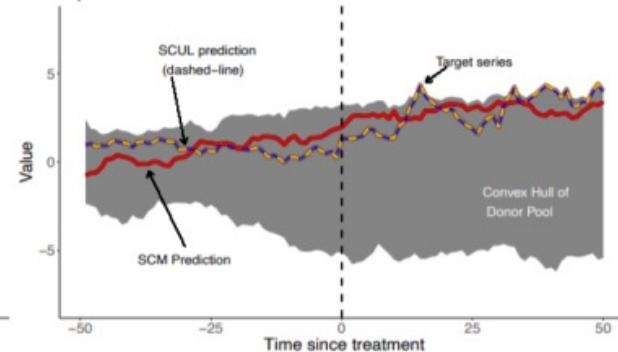
B Case 2: The best donor series for this time series is countercyclical and would need a weight of -1



C Case 1: Traditional SCM is bound by convex hull and cannot use an intercept to select the optimal donor series



D Case 2: Traditional SCM cannot give -1 weight to optimal donor series



# 替代的算法方法比较

- 回归法

$$(\hat{\mu}^{en}, \hat{\omega}^{en}) = \arg \min_{\mu, \omega} Q(\mu, \omega \mid \mathbf{Y}_{t, pre}, \mathbf{Y}_{c, pre}; \lambda, \alpha)$$

$$\begin{aligned} \text{where } Q(\mu, \omega \mid \mathbf{Y}_{t, pre}, \mathbf{Y}_{c, pre}; \lambda, \alpha) &= \|\mathbf{Y}_{t, pre} - \mu - \omega' \mathbf{Y}_{c, pre}\|_2^2 \\ &\quad + \lambda \left( \frac{1-\alpha}{2} \|\omega\|_2^2 + \alpha \|\omega\|_1 \right) \end{aligned}$$

- 平衡法

$$\min_{\mathbf{w}_C} L(\mathbf{w}_C)$$

$$\text{s.t. } \sum_{i \in \mathcal{T}} q_i \mathbf{Y}_{i, pre} = \sum_{j \in \mathcal{C}} w_j \mathbf{Y}_{j, pre}$$

- 潜因子模型

$$Y_{it}(0) = \theta_t' Z_i + \xi_t + \lambda_i' f_t + \varepsilon_{it}$$

# 研究进展

# 课程安排

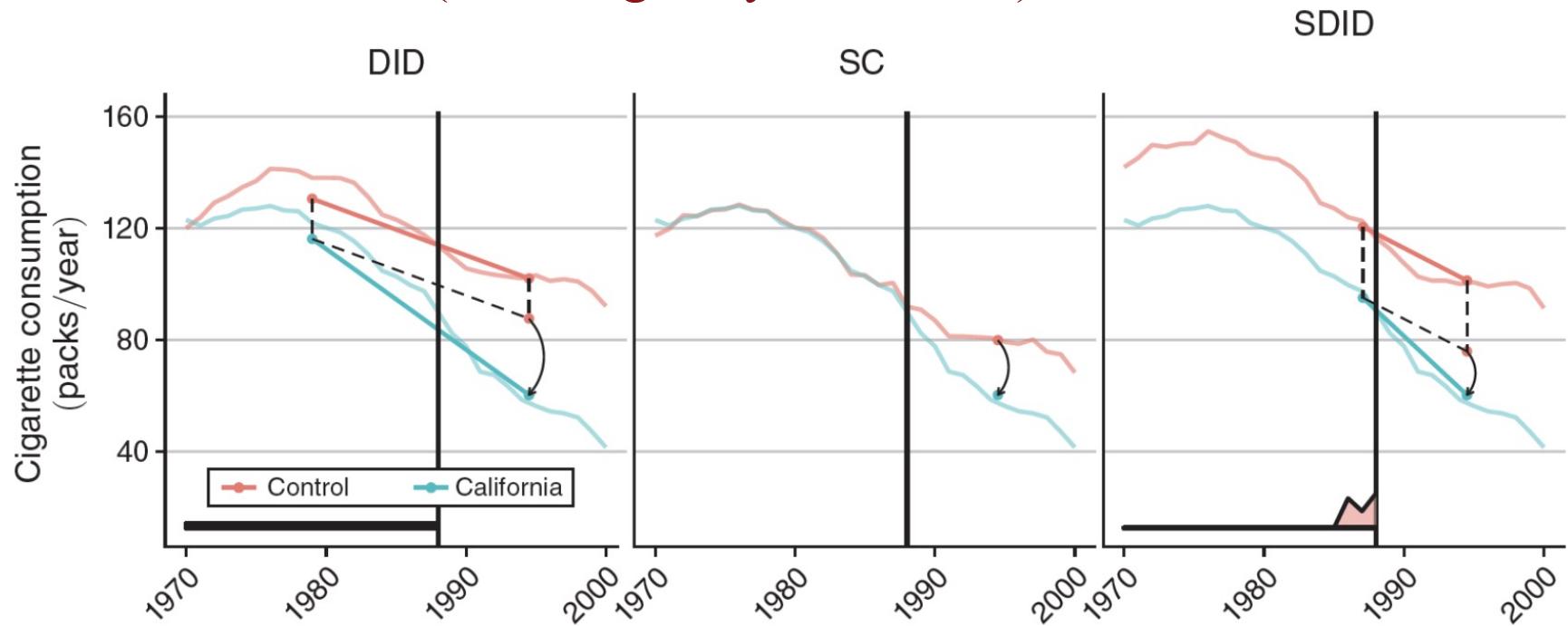
## 合成控制法

- › 快速回顾
- › 替代算法

## 研究进展

- › 合成双重差分法 (Synthetic DID)
- › 潜因子模型 (The latent factor approach)

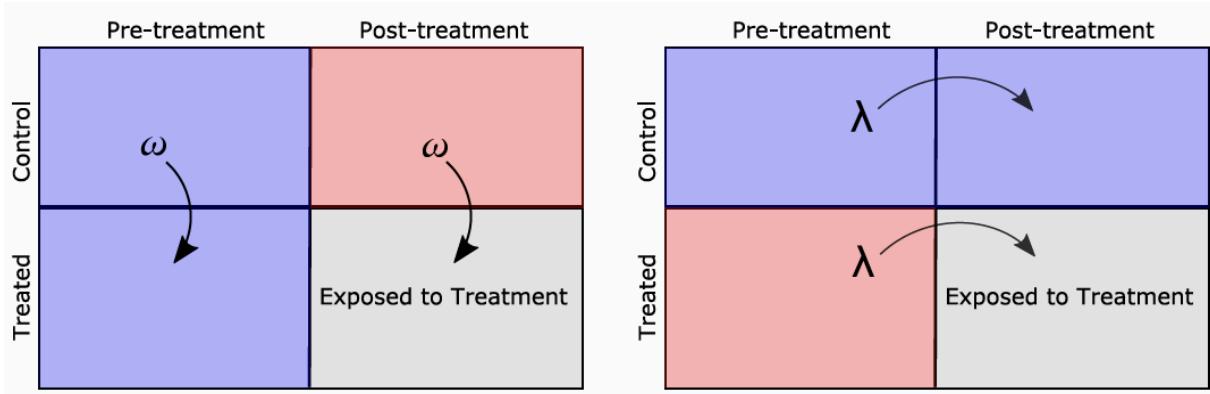
# 合成双重差分法 (Arkhangelsky et al 2021)



- 核心思想：不仅对处理组单元 (control unit) 进行加权，而且对处理前的时期 (pretreatment periods) 进行加权

# 单元和时间权重

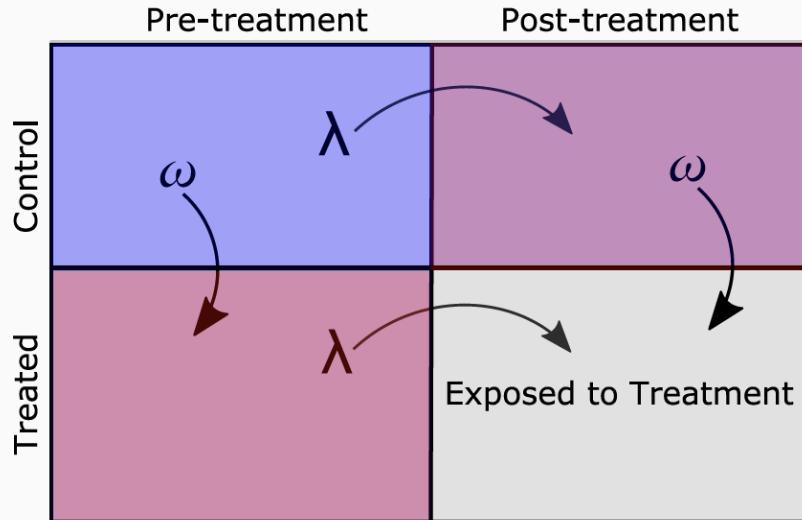
- 合成控制 (synthetic control)：
  - 利用处理前的数据, 对控制组单元进行加权平均 (average of controls) 用以预测加州
  - 假设该关系在处理后时期仍成立, 利用同样的加权来对如果未受处理的 (treatment-free) 加州数据进行插补
- 预测 (Forecasting)：
  - 利用控制组, 对时期进行加权平均 (average of periods) 用以预测到处理后时期
  - 假设该关系对处理组仍成立, 我们利用同样的加权来对如果未受处理的 (treatment-free) 加州数据进行插补



# 单元和时间权重

- 合成双重差分法 (SDID) :

- 通过双重差分法把合成控制和预测结合起来
- 仅要求一种关系成立即可，故双重稳健
- 通过合成控制得到“合成”加州，可以满足与加州的平行趋势



# 合成双重差分法

1. 利用处理前数据，通过估计单元权重  $\hat{\omega}$ ，从而加权得到合成的控制单元

$$\hat{\omega}_0 + \hat{\omega}^T Y_{co,pre} \approx Y_{\overline{tr},pre}.$$

2. 利用控制组数据，通过估计时期权重  $\hat{\lambda}$ ，从而加权得到合成的处理前时期

$$\hat{\lambda}_0 + Y_{co,pre} \hat{\lambda} \approx Y_{co,\overline{post}}.$$

3. 得到合成后的  $2 \times 2$  面板数据后，再在此基础上运用双重差分法

		Synthetic Pre-Treatment	Average Post-treatment
Synthetic Control	Synthetic Control	$\hat{\omega}^T Y_{co,pre} \hat{\lambda}$	$\hat{\omega}^T Y_{co,\overline{post}}$
	Average Treated	$Y_{\overline{tr},pre} \hat{\lambda}$	$Y_{\overline{tr},\overline{post}}$

# 估计权重

- 基于处理前数据，利用约束条件下的最小二乘法，估计单元权重从而加权得到合成的控制单元

$$\hat{\omega} = \underset{\omega_0, \omega}{\operatorname{argmin}} \| \omega_0 + \omega^T Y_{co,pre} - Y_{\overline{tr},pre} \|^2 + \zeta^2 T_{pre} \|\omega\|^2$$

$$\text{s.t. } \omega_i \geq 0, \quad \sum_{i=1}^{N_{co}} \omega_i = 1.$$

- 基于控制组数据，利用约束条件下的最小二乘法，估计时期权重从而加权得到合成的处理前时期

$$\hat{\lambda} = \underset{\lambda_0, \lambda}{\operatorname{argmin}} \| \lambda_0 + Y_{co,pre} \lambda - Y_{co,\overline{post}} \|^2$$

$$\text{s.t. } \lambda_t \geq 0, \quad \sum_{t=1}^{T_{pre}} \lambda_t = 1.$$

- 被证实在因子模型下具有优良性质（不直接估计）：等价于不过度拟合的 Oracle 估计器
- 通过 jackknife 法进行统计推断

# 参考文献

- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010). *Journal of the American Statistical Association*. June 1, 2010, 105(490): 493 {505.
- Athey, Susan, and Guido Imbens. 2016. "The State of Applied Econometrics - Causality and Policy Evaluation." arXiv [stat.ME]. arXiv. <http://arxiv.org/abs/1607.00699>.
- Hahn, Jinyong, and Ruoyao Shi. 2017. "Synthetic Control and Inference." *Econometrics* 5 (4): 52.
- Firpo, Sergio, and Vitor Possebom. 2018. \Synthetic Control Method: Inference, Sensitivity Analysis and Confidence Sets." *Journal of Causal Inference* 6 (2).
- Chernozhukov, Victor, Kaspar Wuthrich, and Yinchu Zhu. 2017. "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls." arXiv [econ.EM]. arXiv. <http://arxiv.org/abs/1712.09089>.
- Ferman, B., C. Pinto, and V. Possebom. 2020. "Cherry Picking with Synthetic Controls." *Journal of Policy Analysis and Management*.
- Cattaneo, Matias D., Yingjie Feng, and Rocio Titiunik. 2019. "Prediction Intervals for Synthetic Control Methods." arXiv [stat.ME]. arXiv. <http://arxiv.org/abs/1912.07120>.
- Doudchenko, Nikolay and Guido Imbens (2016). "Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis." Working Paper Stanford University.
- Hollingsworth, Alex, and Coady Wing. 2020. "Tactics for Design and Inference in Synthetic Control Studies: An Applied Example Using High-Dimensional Data." <https://doi.org/10.2139/ssrn.3592088>.
- Hsiao, Cheng, H. Steve Ching and Shui Ki Wan (2012). "A Panel Data Approach for Program Evaluation: Measuring the Benefits of Political and Economic Integration of Hong Kong with Mainland China," *Journal of Applied Econometrics*, Vol. 27, Iss. 5, August 2012, pp. 705-740.
- Robbins, Michael W., Jessica Saunders, and Beau Kilmer. 2017. "A Framework for Synthetic Control Methods With High-Dimensional, Micro-Level Data: Evaluating a Neighborhood-Specific Crime Intervention." *Journal of the American Statistical Association* 112 (517): 109-26.