

# 香港回归让香港经济衰落了么？

## 基于惩罚降维合成控制法的证据

2020 级 统计学 盛焕新

2023 年 6 月 1 日

# 一、选题动机：兴趣使然

## 计量经济学 Project：研读萧政老师的论文

*Hsiao C, Steve Ching H, Ki Wan S. A panel data approach for program evaluation: measuring the benefits of political and economic integration of Hong Kong with mainland China[J]. Journal of Applied Econometrics, 2012, 27(5): 705-740.*

### 为什么我要继续做这项研究？

- 研究方法：合成控制法 (Synthetic Control Method)
- 研究问题：祖国统一与香港回归
- 研究影响：减少政治污名化
- 研究推进：更稳健的因果识别

# 一、选题动机：萧政教授



图：南加州大学、WISE 教授 萧政

## 二、方法背景：思想方法

传统的合成控制法 (Abadie et al., 2010)：

- 适用情况：找不到好的对照组甚至没有对照组
- 基本思想：没有对照组就人造出一个对照组
- 平行宇宙：用其他单位作为供体池加权合成虚拟对照组
- 因果识别：现实与虚拟对比发现政策效应

## 二、方法背景：基本假设

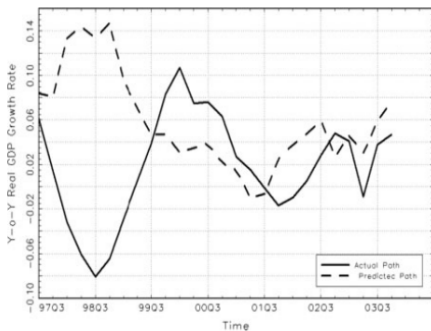
合成控制法五大假设 (Hollingsworth and Wing, 2020):

- 1、不同单位间没有相互干扰，treatment 只影响实验组，对其他单位几乎没有影响
- 2、因子模型： $Y_{st}^0 = a_t \alpha_s + \epsilon_{st}$ ，数据可以被某个无法被观察的因子模型解释
- 3、短期冲击： $\epsilon_{st}$  偶次幂期望有限， $\epsilon_{st}$  相互独立但不必同分布，且条件期望为 0
- 4、多重共线性：处理前公共因子不存在多重共线性
- 5、可解性：可以找到合成权重使得真实值能被合成

## 二、文献综述：实证研究现状

Hsiao C et al., 2012

- 回归合成控制法，认为没有显著影响
- 季度实际 GDP 增速数据 (1993-2003)，用信息准则确定供体
- 存在局限：使用台湾作为供体，供体池只有 24 个单位



图：萧政教授团队合成控制结果

## 二、文献综述：实证研究现状

Jing L et al., 2020

- 传统合成控制法，认为香港回归使香港人均 GDP 下降了 7-9%
- 1989-2008 年度购买力调整人均 GDP 数据，手动选取供体
- 存在局限：使用台湾作为供体，供体池单位更少

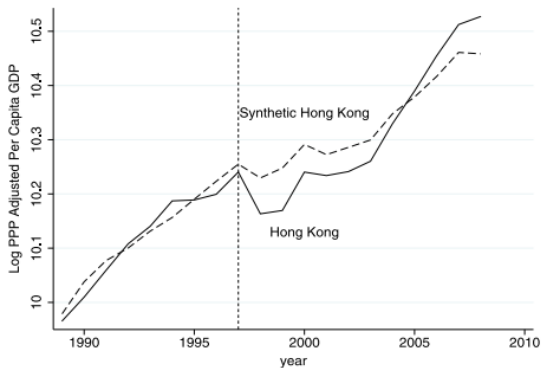


图: Jing Li 团队合成控制结果

香港回归让香港经济衰落了么？

## 二、文献综述：存在局限

### 1、使用台湾作为供体，违反假设一：

- 一国两制、祖国统一的首要目标，蒋政府很害怕
- 台湾与香港的各方面合作都非常密切，相互影响程度很大

### 2、供体太少且局限于亚太地区，时间较短：

- 世界国家地区众多，可能会有与香港相似的经济体
- 局限于亚太地区，这些地区必然会受到香港的影响
- 时间局限在主权移交附近，难以准确刻画长期影响

### 3、确立回归是 1984 年，以 1997 作为处理时间并不合理：

- 港英政府做了非常多的“改革”，“玫瑰园计划”掏空财政
- 近 80 万香港人移民，中上阶层居多，“居英权计划”掏空人才
- 政治稳定确立，吸引外来资本入驻香港获取金融服务
- 政治统一，人民团结，香港与内地经济联系超前紧密

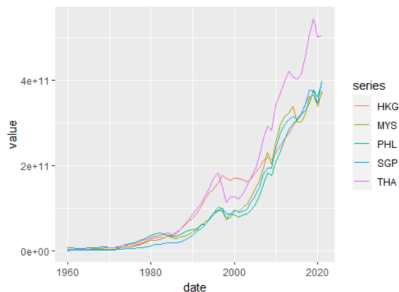


### 三、研究推进：研究数据

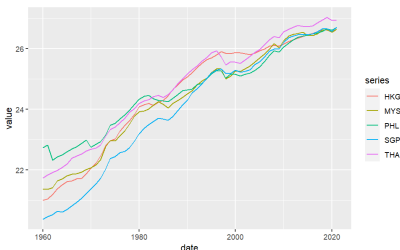
研究数据：

- 采用世界银行实际 GDP 年度数据（2015 年美元不变价）
- 时间范围：1960 年-2021 年，每个地区有 62 年的观测值
- 单位范围：填补缺失值后共 144 个国家或地区及地区集团
- 数据处理：对 GDP 数据做常用的取对数处理，转为线性

### 三、研究推进：研究数据



图：取对数前东南亚五个地区的GDP 变化



图：取对数后东南亚五个地区的GDP 变化

### 三、研究推进：研究数据

对于缺失值填补：

- 对于缺失值比例低于 10% 的地区进行填补
- 每个地区缺失值个数不超过 6 个，大多分布在首尾
- 这些地区中包含阿根廷、文莱、丹麦和新西兰等共计 22 个地区
- 考虑到数据相关性，选用随机森林算法进行填补
- 填补效果：相对准确，小的牺牲换得大的回报

### 三、研究推进：研究方法

研究方法：

- 面对高维数据，采取增加惩罚项来降维的方法
- 在 R 中使用 (Hollingsworth and Wing, 2020) 的 Lasso 回归合成控制
- 在 Python 中复现并将 Lasso 改为 ElasticNet 回归合成控制
- 分别考虑“1984 年中英联合公报宣布香港回归”和“1997 年香港正式回归”作为对香港的 Treatment Year

### 三、研究推进：研究方法

传统的合成控制法：

$$\hat{\omega}_{OLS} = \arg \min_{\omega} \left( \sum_{t=1}^{T_0} (y_{0t} - x_t \omega)^2 \right) \quad (1)$$

为什么要加惩罚项？信息不是越多预测就会越准吗？

- 导致过拟合：只在 Treatment 前预测得准，Treatment 后就不准了
- 过拟合原因：过度重视 Treatment 前的噪声干扰信息
- 防止过拟合：需要筛选出最优的供体来合成平行宇宙中的香港

### 三、研究推进：研究方法

惩罚降维的合成控制法：

$$\hat{\omega}_{Lasso} = \arg \min_{\omega} \left( \sum_{t=1}^{T_0} (y_{0t} - x_t \omega)^2 + \lambda |\omega|_1 \right) \quad (2)$$

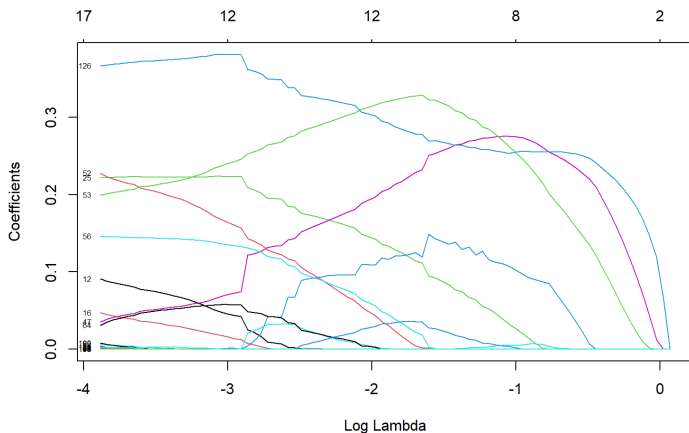
$$\hat{\omega}_{ElasticNet} = \arg \min_{\omega} \left( \sum_{t=1}^{T_0} (y_{0t} - x_t \omega)^2 + \lambda_1 |\omega|_1 + \lambda_2 |\omega|_2^2 \right) \quad (3)$$

惩罚降维的原理：

- Lasso 惩罚：减少使用的供体数量，防止过拟合
- Lasso 的局限：变量彼此不相关，造成过拟合
- 增设 Ridge 惩罚：控制相关变量，防止过拟合
- ElasticNet 惩罚：合二为一，共同作用

### 三、研究推进：研究方法

惩罚系数 $\lambda$  的选取影响  $\rightarrow$  权重：



图：Lasso 降维  $\lambda$  选取对权重的影响

### 三、研究推进：研究方法

惩罚系数 $\lambda$ 的选取影响  $\rightarrow$  合成：

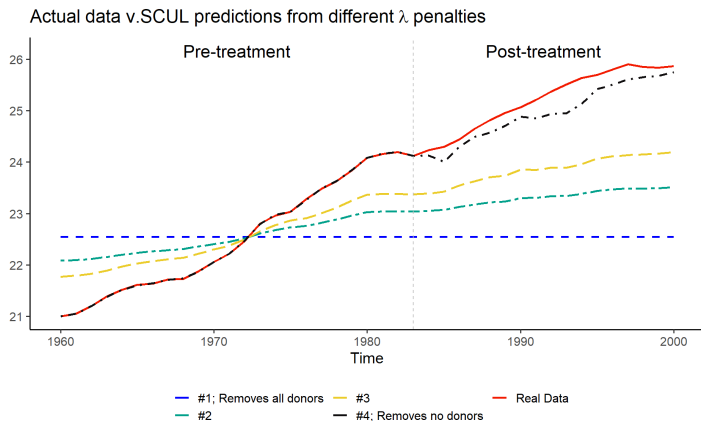
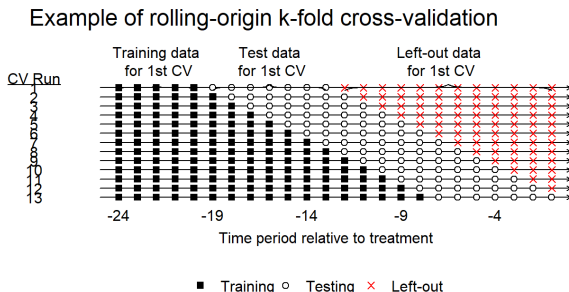


图: Lasso 降维  $\lambda$  选取对合成的影响



### 三、研究推进：研究方法

滚动 K 折交叉验证 → 确定 Lasso $\lambda$  的取值：



图：滚动 K 折交叉验证确定 Lasso $\lambda$  的取值

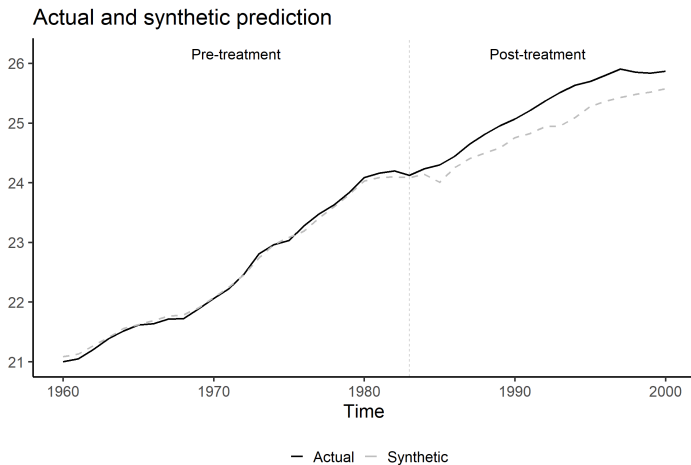
### 三、研究推进：研究方法

$$\hat{\omega}_{ElasticNet} = \arg \min_{\omega} \left( \sum_{t=1}^{T_0} (y_{0t} - x_t \omega)^2 + \lambda_1 |\omega|_1 + \lambda_2 |\omega|_2^2 \right) \quad (4)$$

ElasticNet 惩罚系数的选取：

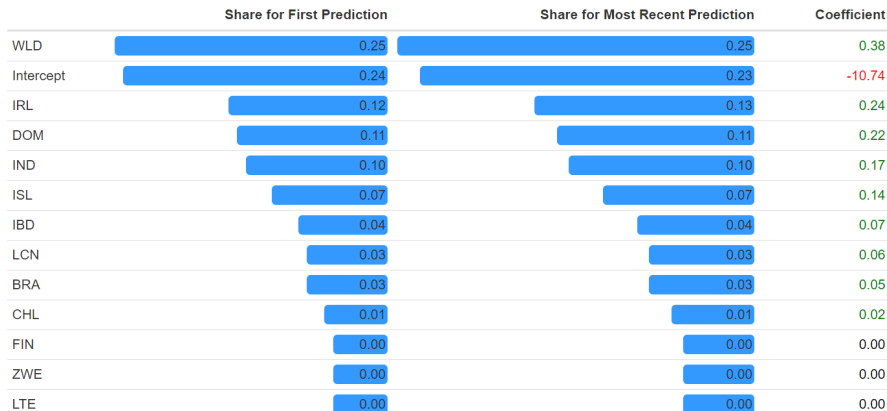
- 五折交叉验证进行随机参数搜索
- 限定参数范围：alpha[1e - 6, 1], l1\_ratio[0, 1]
- 搜索出使五次交叉验证总误差最小的参数
- 限于时间和能力，很多 ‘SCUL’ 中的内容还没实现...
- 最终搜索结果：'alpha' : 0.0070, 'l1\_ratio' : 0.5110

## 四、实证结果：取对数 1984



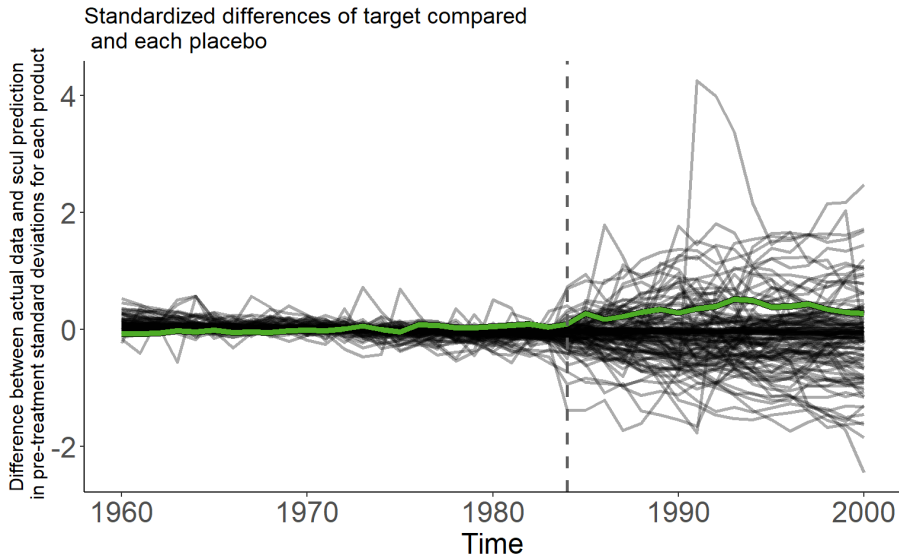
图：1984 宣布主权移交对香港 GDP 的影响

## 四、实证结果：取对数 1984



图：Lasso 取对数 1984：选取供体及其贡献

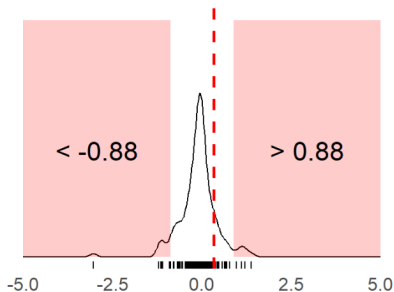
## 四、实证结果：取对数 1984



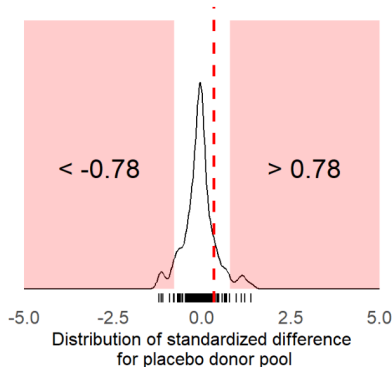
## 四、实证结果：取对数 1984

Placebo distribution compared to ATE estimate in pre-period standard deviations

No Cohen's D restriction



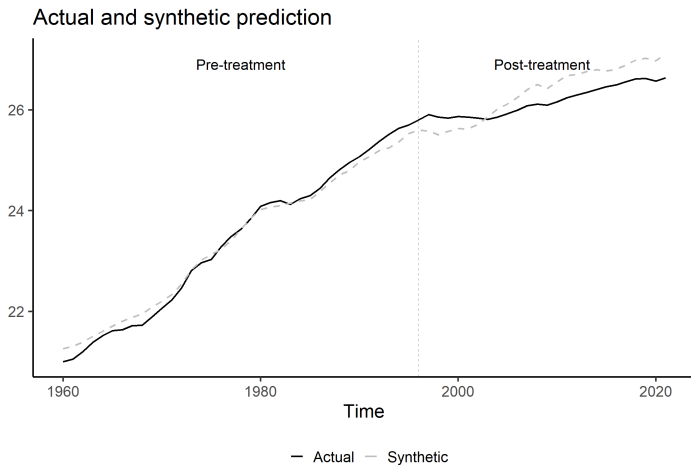
0.25 Cohen's D restriction



安慰剂检验：不限定 Cohen's D 统计量时，p 值为 0.39；限定 D 不高于 0.25 的安慰剂组进行对比时，p 值为 0.36。

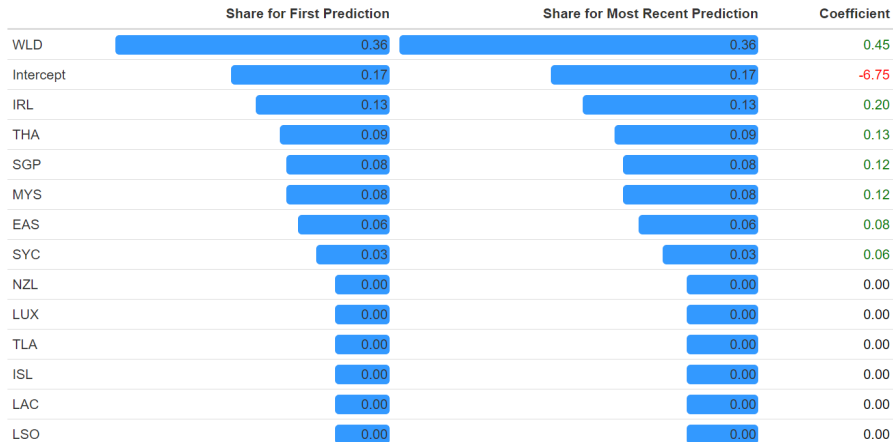
即：“1984 年宣布主权移交”促进了香港经济发展但不显著。

## 四、实证结果：取对数 1997



图：1997 主权正式移交对香港 GDP 的影响

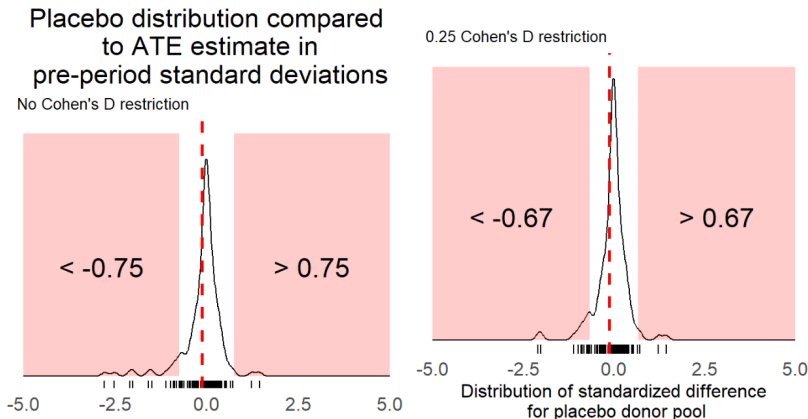
## 四、实证结果：取对数 1997



图：Lasso 取对数 1997：选取供体及其贡献



## 四、实证结果：取对数 1997



安慰剂检验：不限定 Cohen's D 统计量时，p 值为 0.57；限定 D 不高于 0.25 的安慰剂组进行对比时，p 值为 0.56。

即：“1997 年正式主权移交”拖累了香港经济发展但很不显著。

## 四、实证结果：不取对数 1984

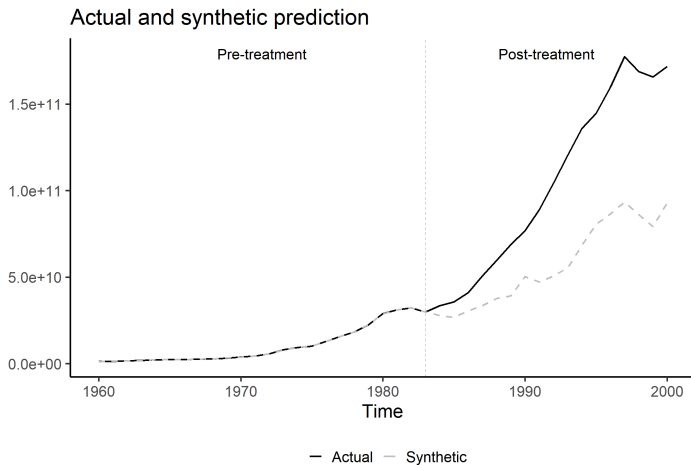
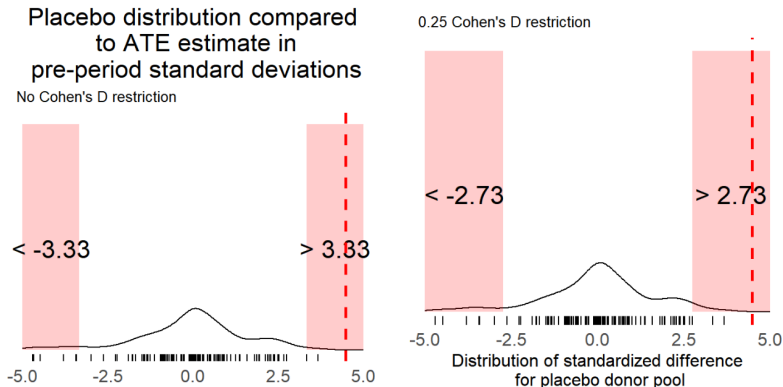


图: 1984 宣布主权移交对香港 GDP 的影响

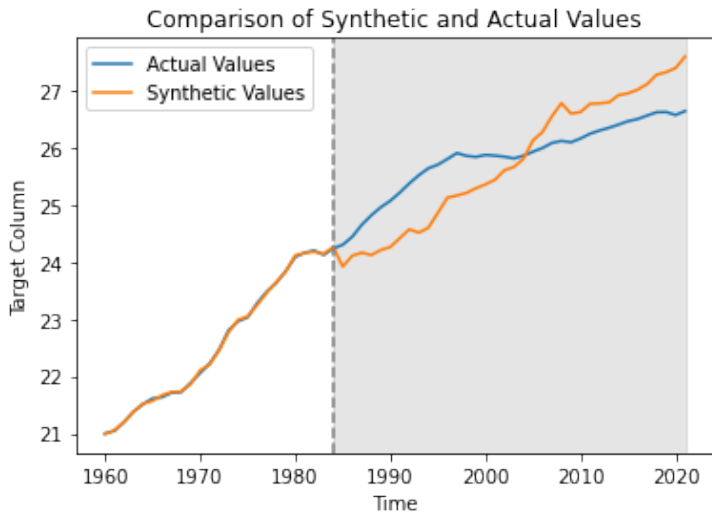
## 四、实证结果：不取对数 1984



安慰剂检验：不限定 Cohen's D 统计量时， $p$  值为 0.05；限定 D 不高于 0.25 的安慰剂组进行对比时， $p$  值为 0.04。

即：“1984 年宣布主权移交”促进了香港经济发展但非常显著。

## 四、实证结果：取对数 1984



图：1984 宣布主权移交对香港 GDP 的影响

## 五、研究结论及存在不足

从实际 GDP 的角度衡量香港经济发展，发现：

- 1984 年宣布主权移交促进了香港经济发展，不显著
- 1997 年正式主权移交拖累了香港经济发展，很不显著
- 总而言之：没有显著证据表明香港回归让香港经济衰落了

研究存在的不足：

- 1、难以考虑遗漏变量，依旧无法排除 1998 年亚太金融风暴对 1997 主权移交的因果识别影响
- 2、仍然受过拟合困扰，时间跨度大的预测难以保证准确
- 3、Lasso 惩罚是不够完美的，需要进一步完善 ElasticNet 的惩罚回归方法，得到更稳健的结果