

面板数据因果推断

第四讲 对异质性稳健的估计量

徐轶青

斯坦福大学

课程计划

- 第一讲：经典双向固定效应模型
- 第二讲：双重差分法：因果推断的框架
- 第三讲：双向固定效应模型的问题
- 第四讲：对异质性稳健的新估计量
- 第五讲：合成控制法及其拓展
- 第六讲：潜因子法与双重稳健法

第四讲大纲

- 对异质性稳健的估计量
 - 概览
 - 交互加权法
 - CSDID
 - 匹配法
 - 插补法
- 一项全面的复制研究
 - 样本和步骤
 - 典型案例一
 - 系统性评估
 - 典型案例三则
- 融会贯通

引子和符号

单元 (Unit) $i \in \{1, \dots, N\}$

时期 $t \in \{1, \dots, T\}$

处理组 $D_{it} \in \{0,1\}$

潜在结果 (Potential outcome) $Y_{it}(d), d \in \{0,1\}$

观测到的结果 (Observed outcome) $Y_{it} = D_{it}Y_{it}(1) + (1 - D_{it})Y_{it}(0)$

估计目标：

- 个体处理效应 (Individual treatment effects)

$$\tau_{it} = Y_{it}(1) - Y_{it}(0)$$

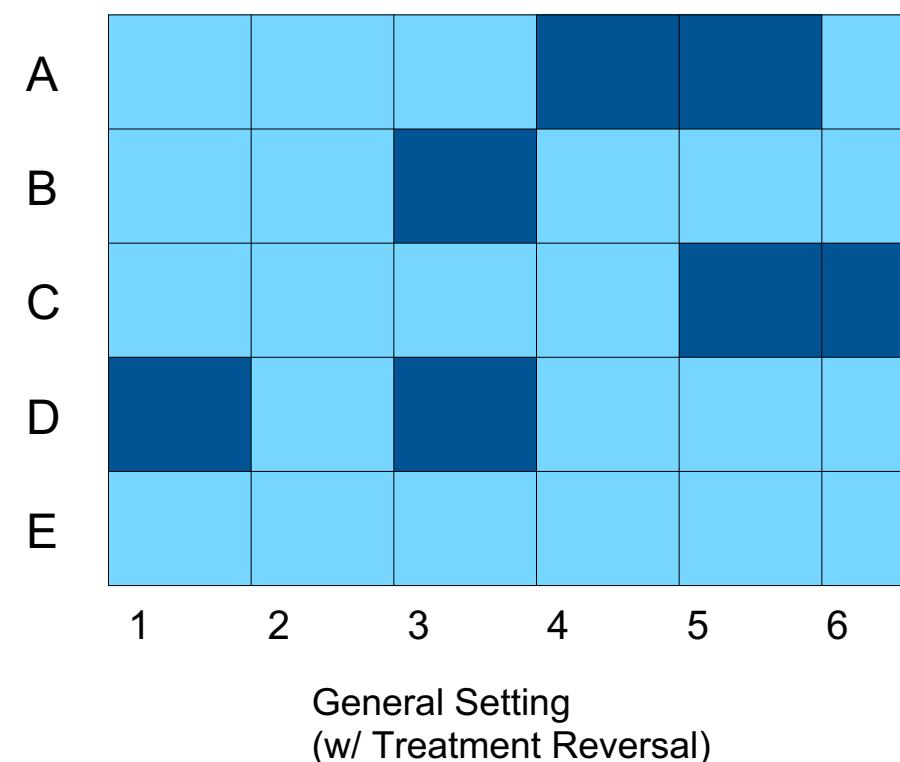
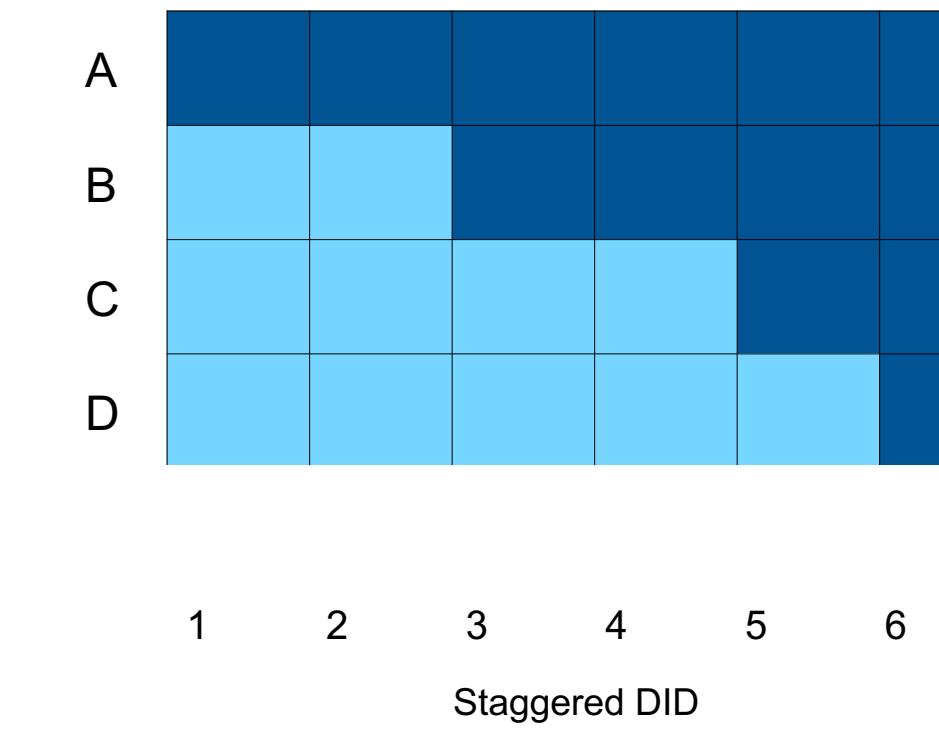
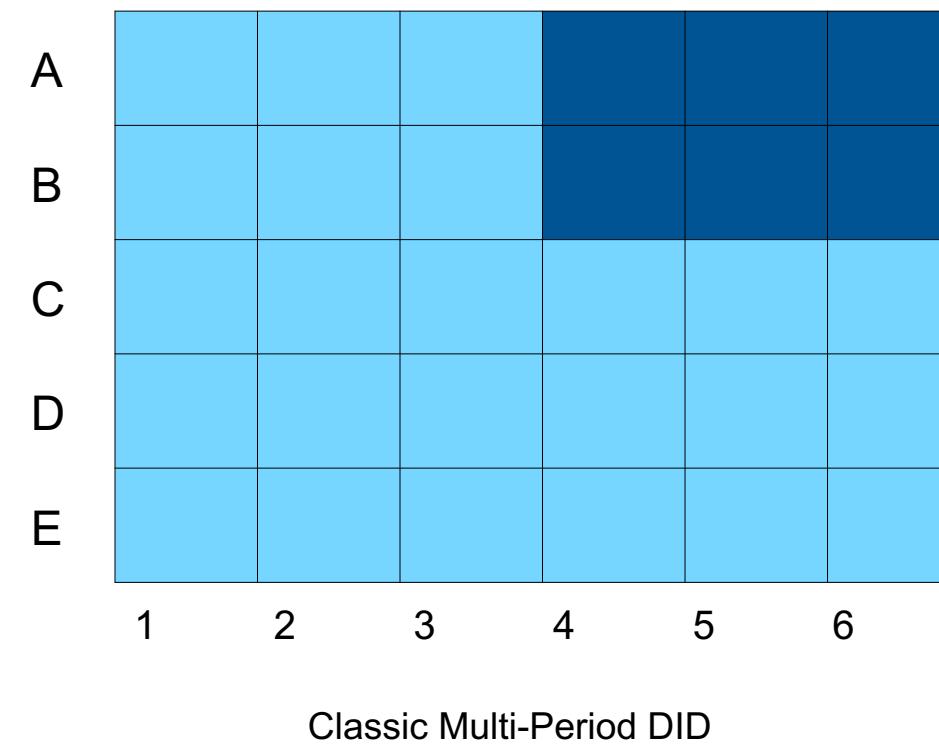
- (处理组的) 动态处理效应 (Dynamic treatment effects (for treated))

$$\tau_t = \mathbb{E}_i[\tau_{it} | D_{it} = 1]$$

- 部分 (处理组的) 平均处理效应 (average of treatment effects (for treated))

处理组的个体 P 的分布的 $\mathbb{E}_P[\tau_t]$

设定



双向固定效应

$$Y_{i,t} = \alpha_i + \xi_t + X'_{i,t}\beta + \sum_{\substack{l=-(a-1) \\ l \neq 0}}^b \delta_l^{TWFE} \cdot \mathbf{1}\{K_{i,t} = l\} + \delta_{b+}^{TWFE} \mathbf{1}\{K_{i,t} > b\} \cdot D_{i,t} + \epsilon_{i,t},$$

- 动态形式（事前和滞后期）
 - $K_{i,t}$ = 进入到处理组后的第几期 ($K_{i,t} = 1$ 表示干预后的第一期>)
 - $\hat{\delta}_l^{TWFE}$ ：从干预开始的第 l 期的动态效应 (Dynamic effect)
 - 通过动态平面图 (Dynamics plot) 去判断事前趋势 (pretrends)
 - 长期效应 $\hat{\delta}_{b+}^{TWFE}$
 - 在异质性处理效应 (HTE) 下，估计量可能无法进行因果解释，即使满足了强外生性 (strict exogeneity) 或平行趋势 (parallel trends)

异质性处理效应稳健估计量 (HTE-Robust Estimators)

- 放松“同质性处理效应”的假设
- 仍然要求平行趋势
 - 可以用来画动态平面图 (dynamics plots)

		DID的拓展 (将2x2 DID 作为基础)		插补法 (固定效应的结果模型)	
情景		交错	常规	常规	
估计目标 (Estimand)		CATT	ATT	ATT	

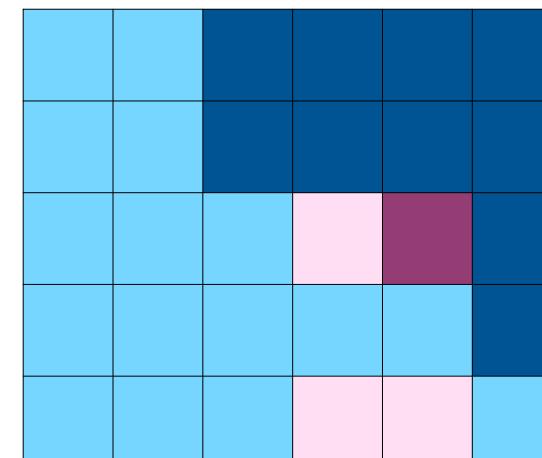
不同的估计量使用不同的对照组

DID的拓展

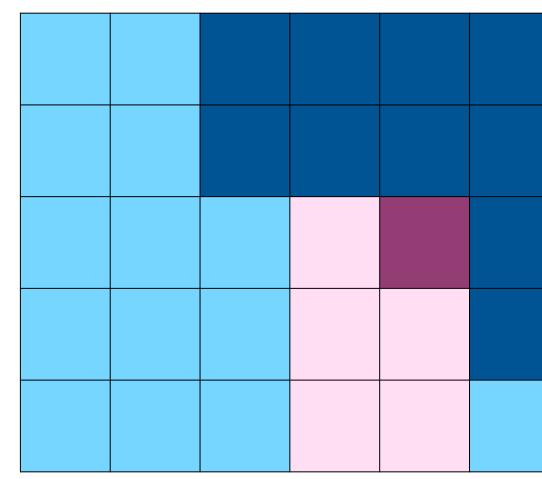
插补法

交错

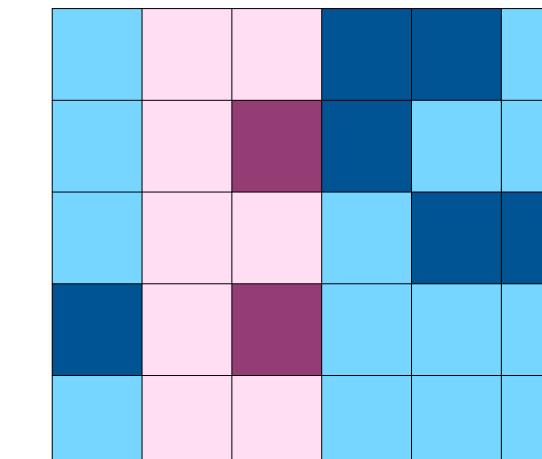
一般



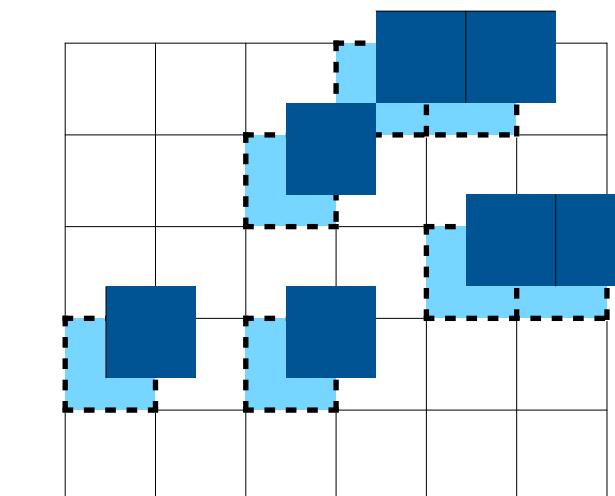
Interaction Weighted



CSDID



DID multiple/PanelMatch

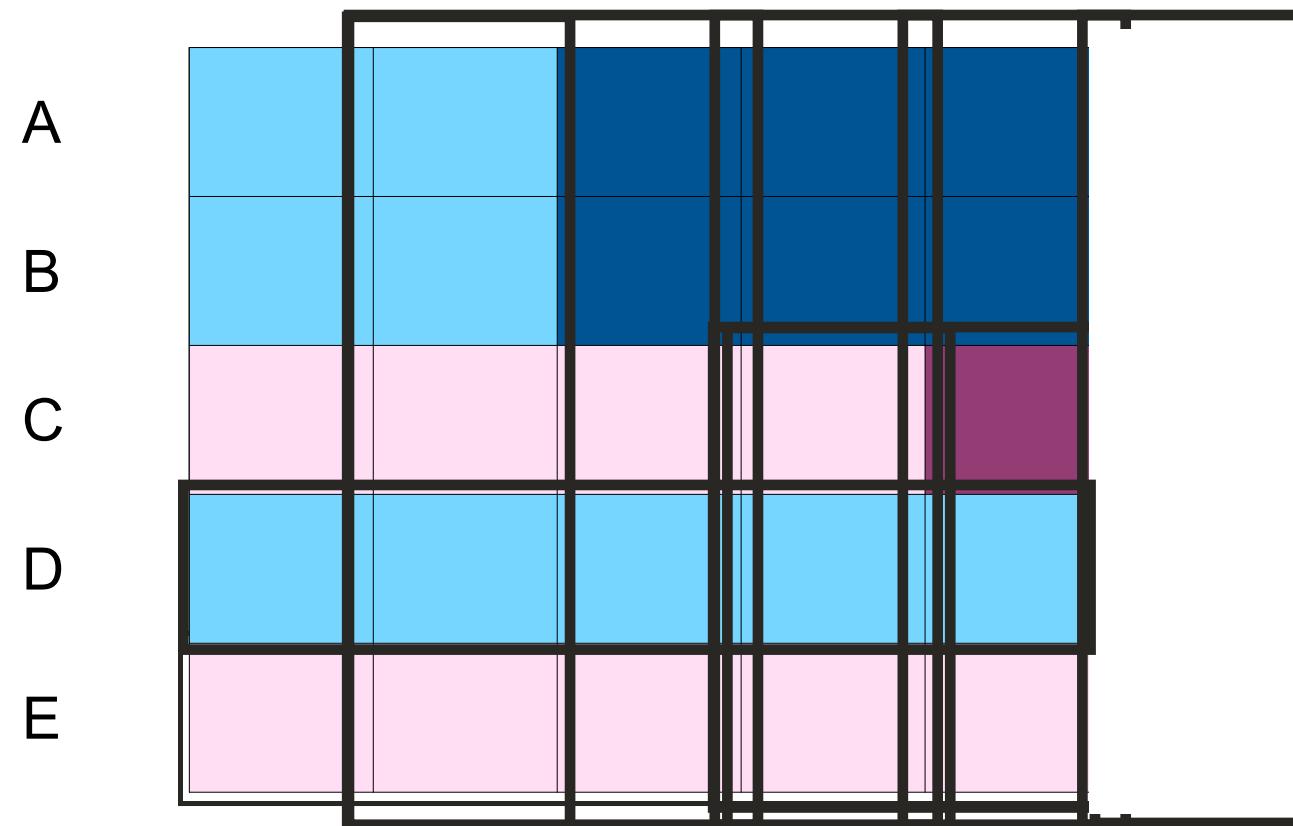


Imputation Method
e.g., FECT

Sun & Abraham (2021)

交互加权 (Interaction Weighted, IW)

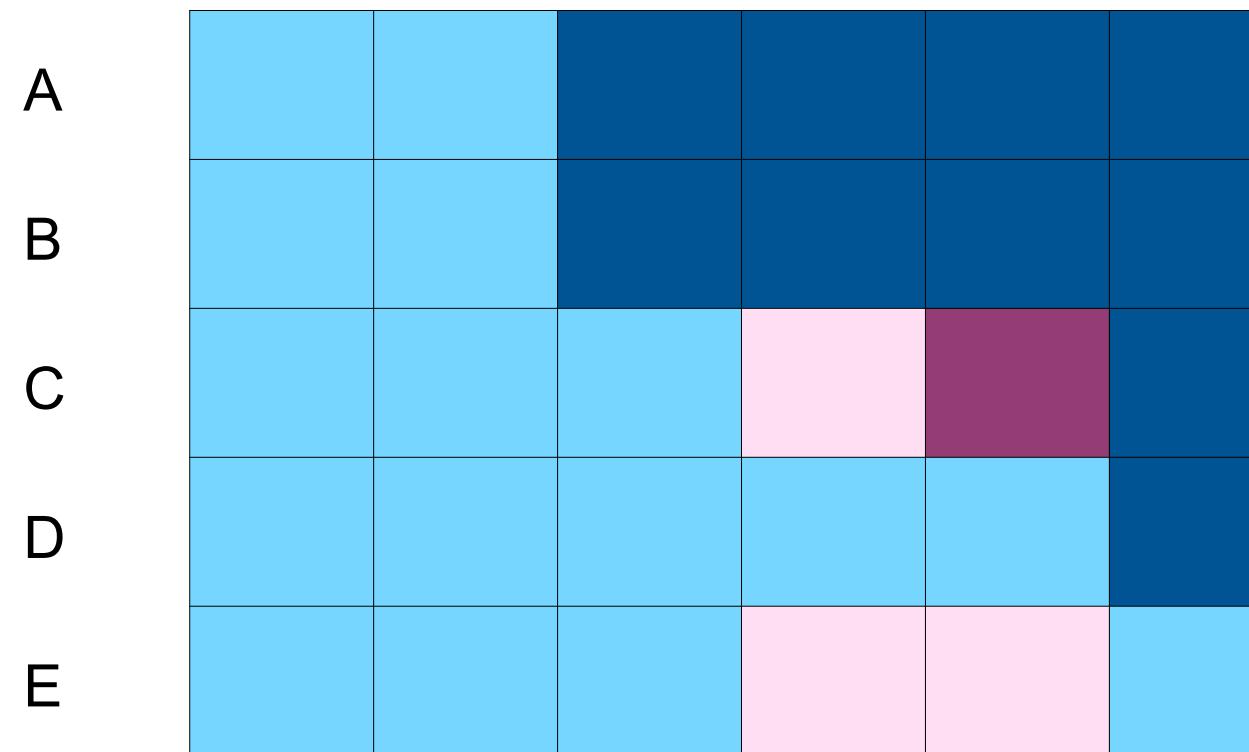
- 对照组: 非处理组 (never-treated) 或者最晚处理组 (last-treated)
- 对于每个队列 (cohort) g 和干预开始的第 l 期, 使用DID来估计CATT
- 基于队列规模 (cohort size) 加权的平均值



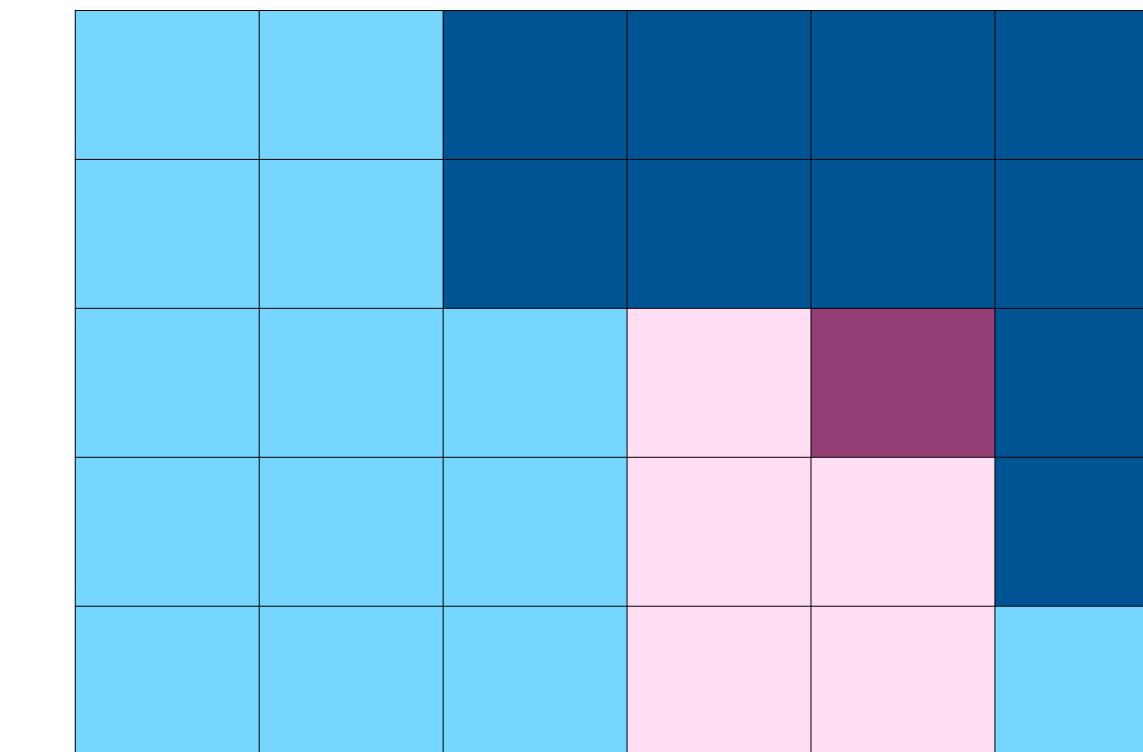
Callaway & Sant'Anna (2021)

- 对照组：一直未接收处理的个体（never-treated）或还未接受处理的个体（not-yet-treated）
- 使用DID去估计基础：CATT(g, l)

Sun & Abraham (2021)

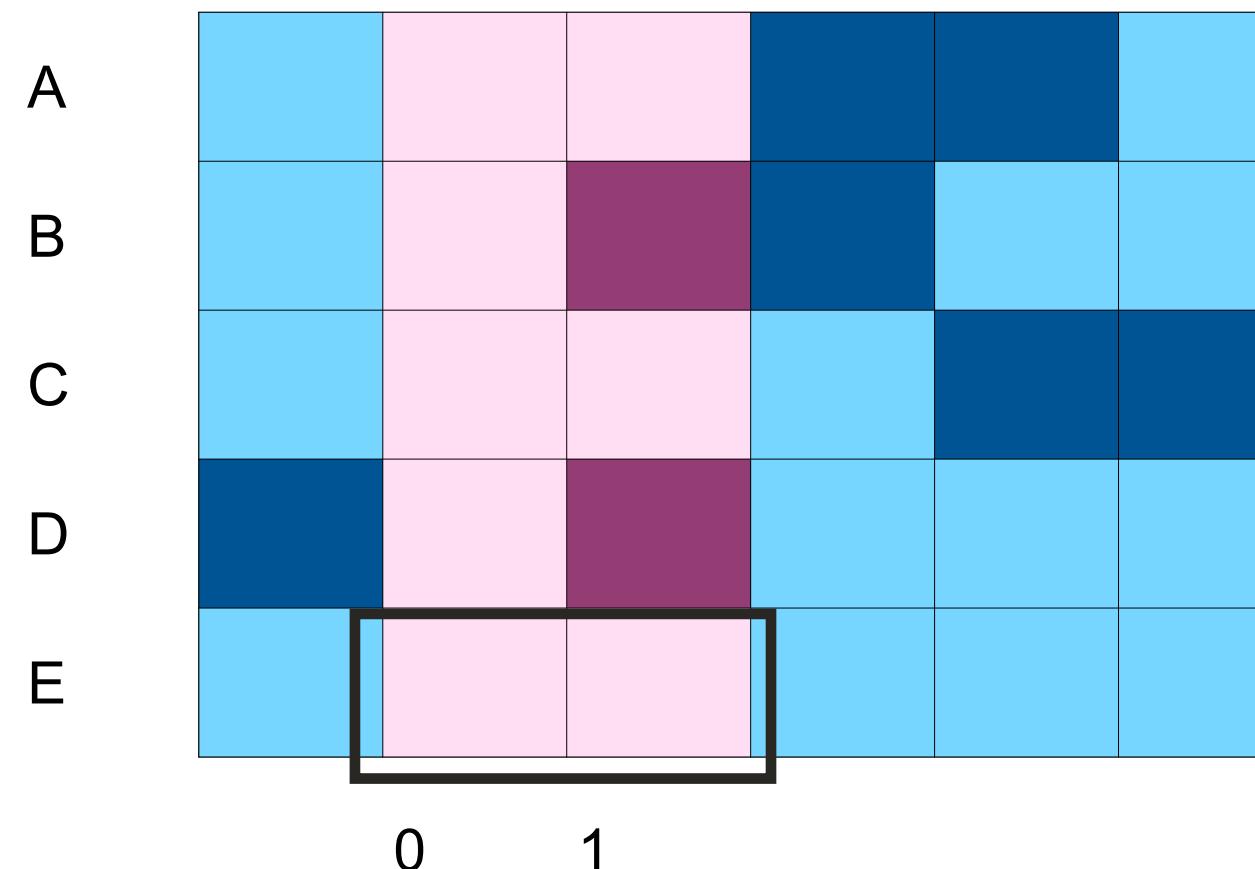


Callaway & Sant'Anna (2021)



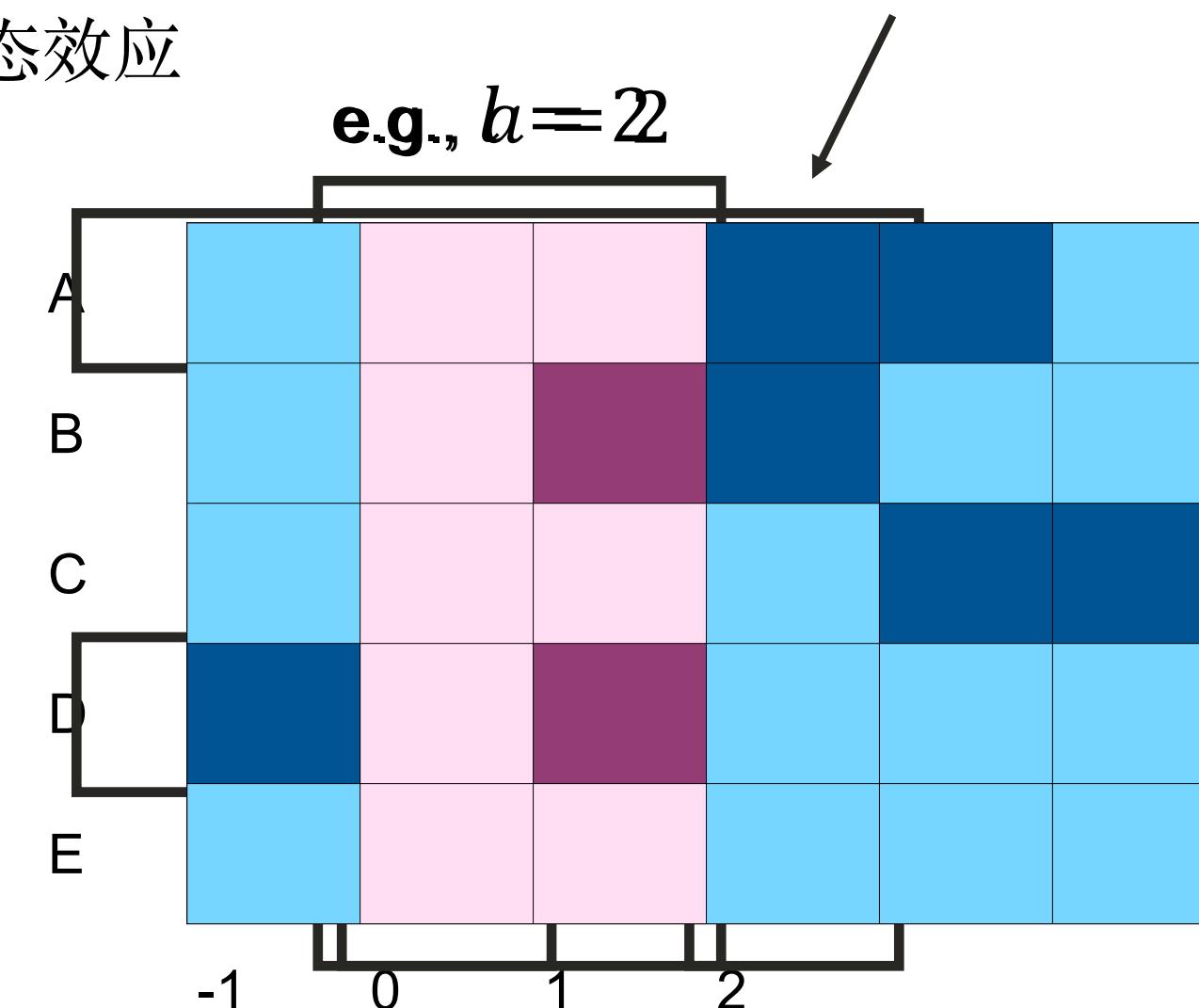
de Chaisemartin and D'Haultfœuille (2020)

- 没有队列 (cohort) ; 进入/退出者 (Switcher) 的效应
- 通过事前的共同干预状态来匹配处理组和控制组
 - 进入/退出者(i, t): $D_{it} = 1, D_{it-1} = 0$
 - 状态稳定者 (i, t): $D_{it} = 0, D_{it-1} = 0$
- DID_M : 使用DID来估计进入/退出 (Switch) 发生当期的效应



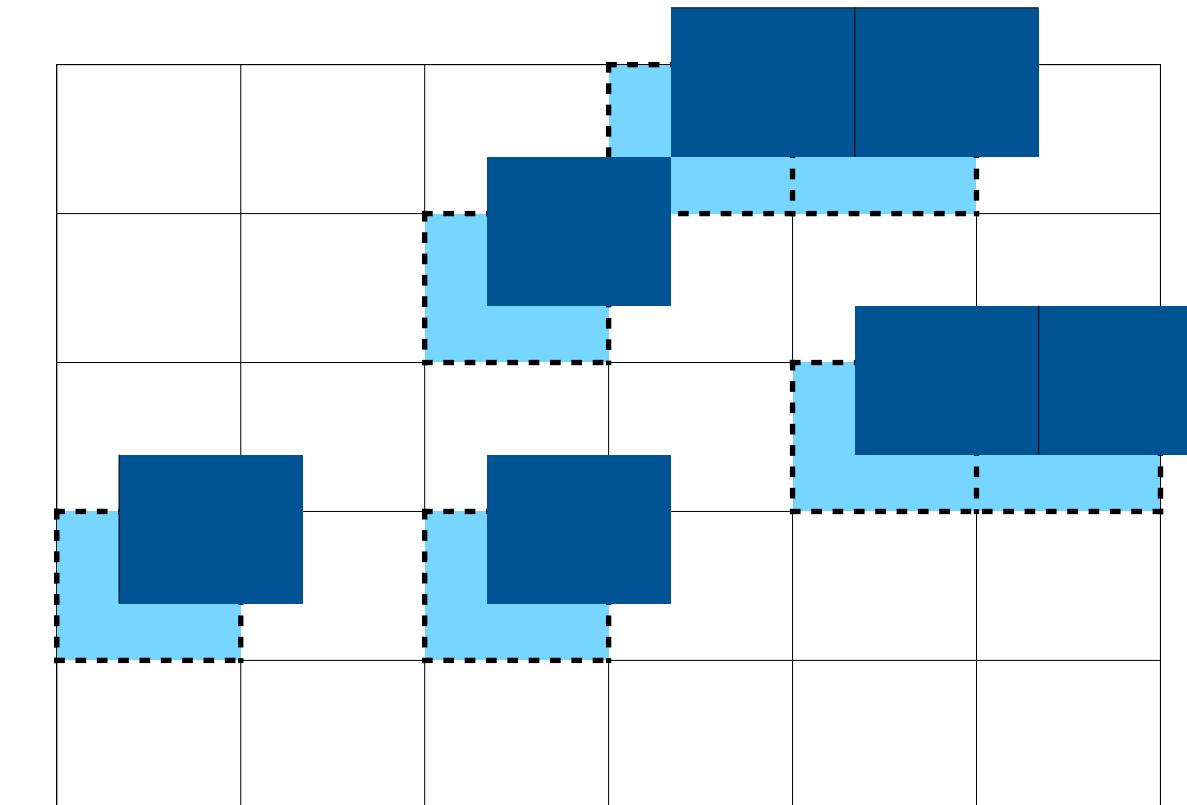
Imai, Kim & Wang (2021) “面板匹配” (PanelMatch)

- 在进入/退出前一直匹配到第 a 期
 - ▶ 用 $\{j: D_{is} = D_{js}, \text{ 对于所有 } s \in \{t-1, t-2, \dots, t-a\}\}$ 去匹配处理组 (i, t)
- 使用DID去估计未来 $l = 1, 2, \dots$ 期之后的动态效应
 - (直至有进入/退出)
- (用控制变量 X_{it} 来匹配)
- $a = l = 1$ (没有匹配) 与 DID_M 是一致的

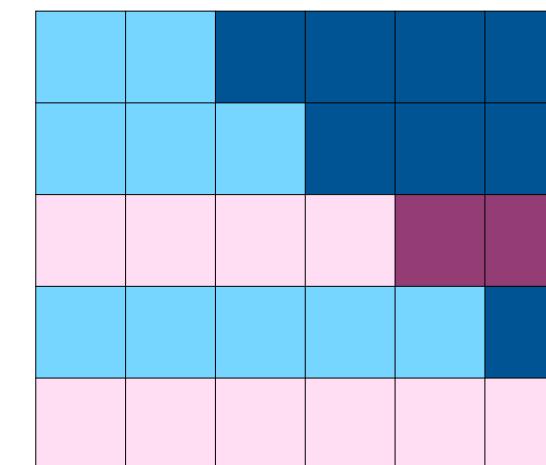
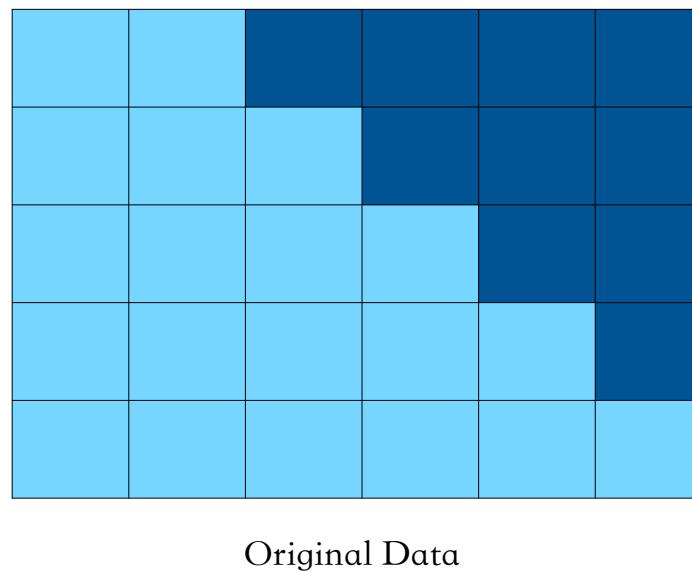


插补法 (Imputation Methods) (BJS 2021; LWX 2022)

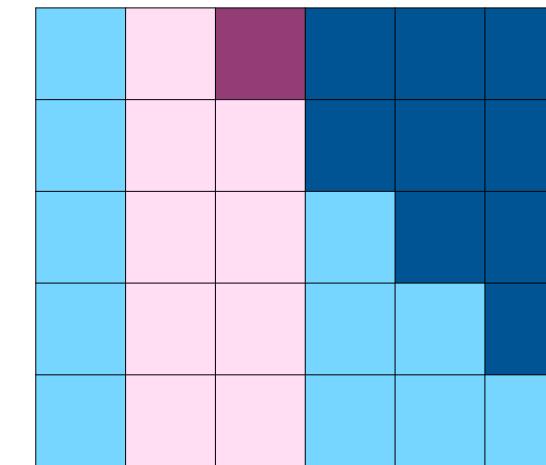
- 构建关于 $Y_{it}(0)$ 与控制变量的模型
 - 双向固定效应 (TWFE) : 固定效应反事实估计量 (Fixed effects counterfactual estimator) (fect, LWX 2022)
- 对处理组进行插补 $\hat{Y}_{it}(0)$ 估计处理组的个体处理效应 (ITE)
 - $\hat{\delta}_{it} = Y_{it} - \hat{Y}_{it}(0)$
 - 基于 $\hat{\delta}_{it}$ 进行加总
 - 在同方差假设下最有效率 (BJS 2021)



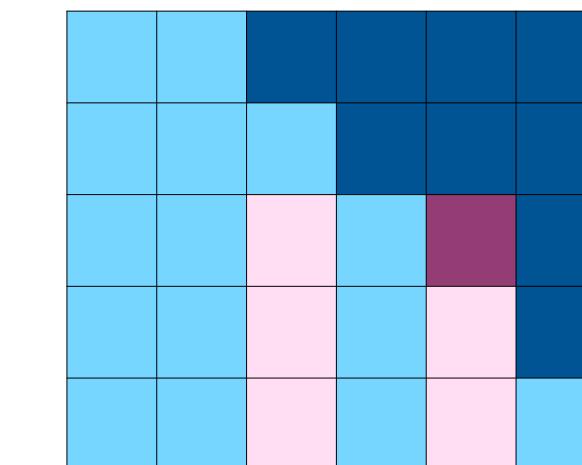
对比 — 交错设置



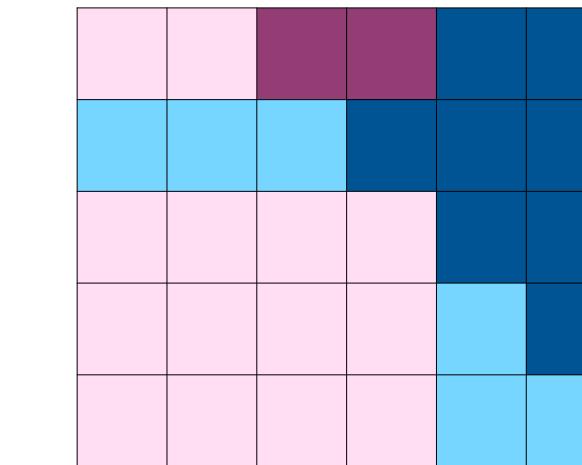
Interaction Weighted



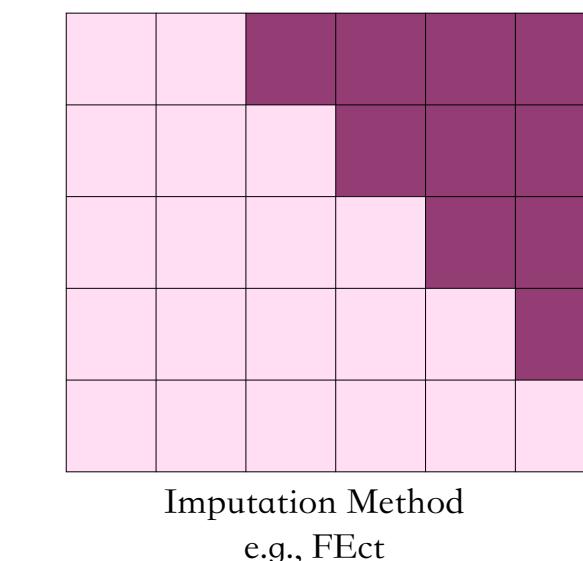
DID multiple



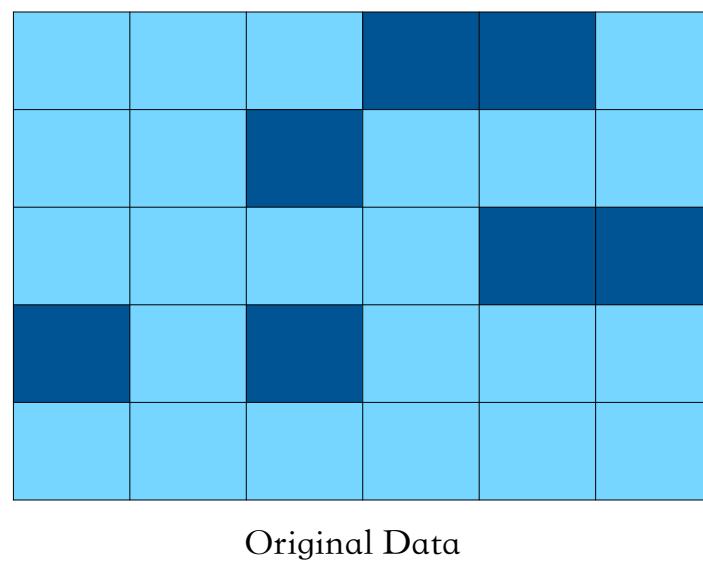
CSDID



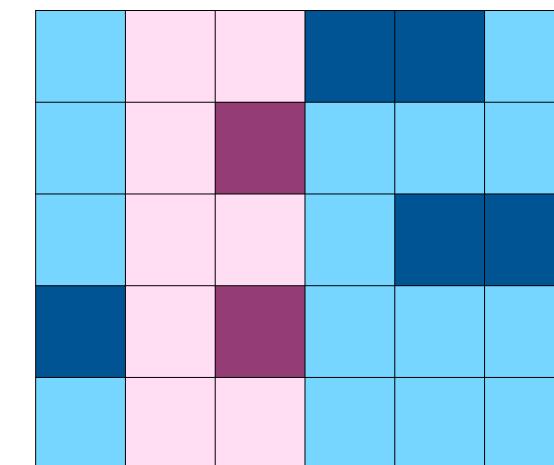
PanelMatch



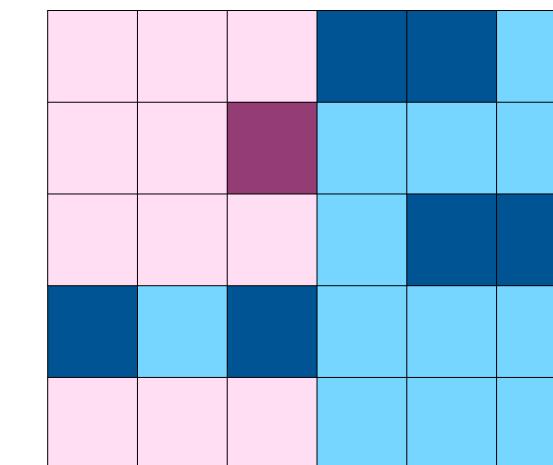
对比 — 一般化设定



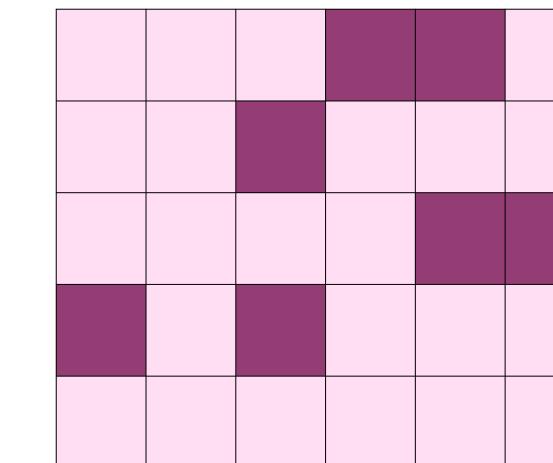
Original Data



DID multiple



PanelMatch



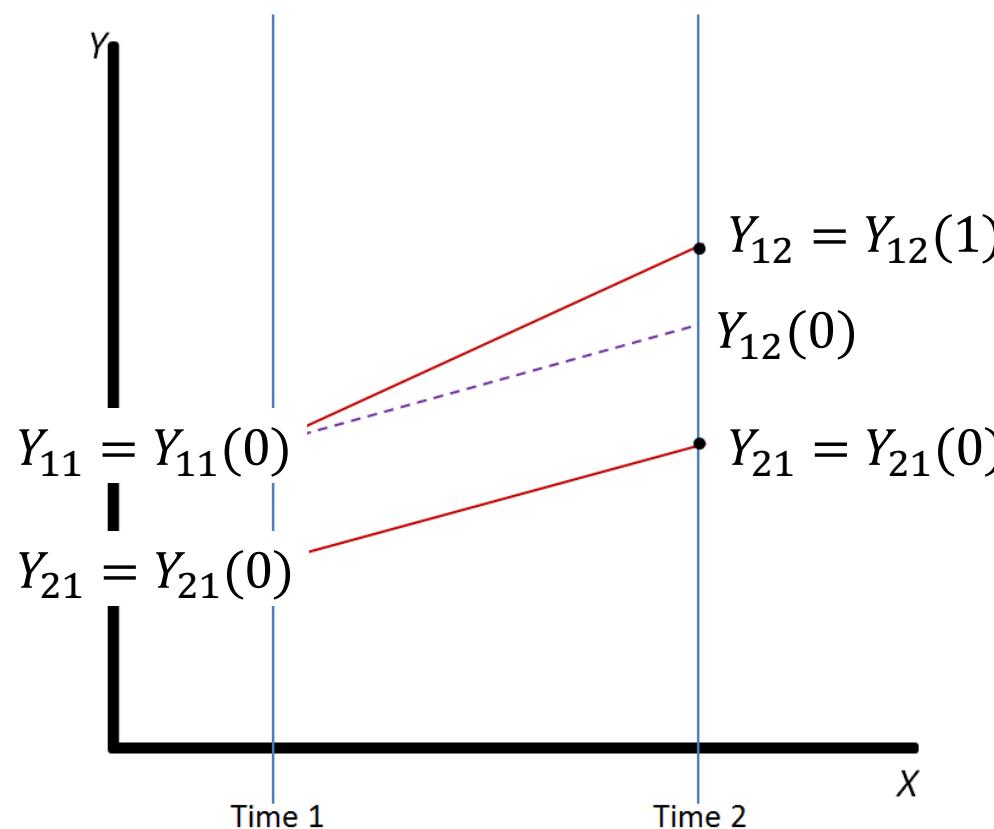
Imputation Method
e.g., FECT

异质性处理效应稳健标准误 (HTE-Robust Estimators)

	DID的拓展 (将2x2 DID 作为基础)	插补法 (固定效应的结果模型)
情境	交错	常规
估计目标 (Estimand)	CATT	ATT
估计量	IW, CSDID	DID_M , PanelMatch
对照组	一直未处理/最晚/ 还未处理	匹配的组别
关键假设	平行趋势	平行趋势
		强外生性

关键假设：平行趋势 (Parallel Trends)

- 异质性处理效应稳健估计量 (HTE-robust estimators) 放松了“同质性处理效应”假设 (homogeneous effects assumption)
- 仍然对动态 (dynamics) 有很强的假设
- 平行趋势 (Parallel trends)：
 - $\mathbb{E}[Y_{it}(0) - Y_{is}(0)|i \in \mathcal{T}] = \mathbb{E}[Y_{it}(0) - Y_{is}(0)|i \in \mathcal{C}]$



第四讲大纲

- 对异质性稳健的估计量
 - 概览
 - 交互加权法
 - CSDID
 - 匹配法
 - 插补法
- 一项全面的复制研究
 - 样本和步骤
 - 典型案例一
 - 系统性评估
 - 典型案例三则
- 融会贯通

研究目的

问题

- 政治科学领域的实证发现，在这些新的洞见面前，是否站得住脚？
- 与违反平行趋势（parallel trends）相比，异质性处理效应（HTE）导致的偏误有多严重？

我们做了什么

- 收集了发表在政治科学三本顶尖期刊的30篇论文的复现材料，这些论文使用了含有二元处理变量的面板数据
- 使用新估计方法和检验手段的每篇论文的主要结果进行重现和再分析
- 基于这些结果，为应用研究者提升实操提供建议

研究发现

主流操作

双重固定效应模型 (TWFE) (90%) + 聚类稳健标准误 (cluster-robust SEs) (几乎100%)

共性问题

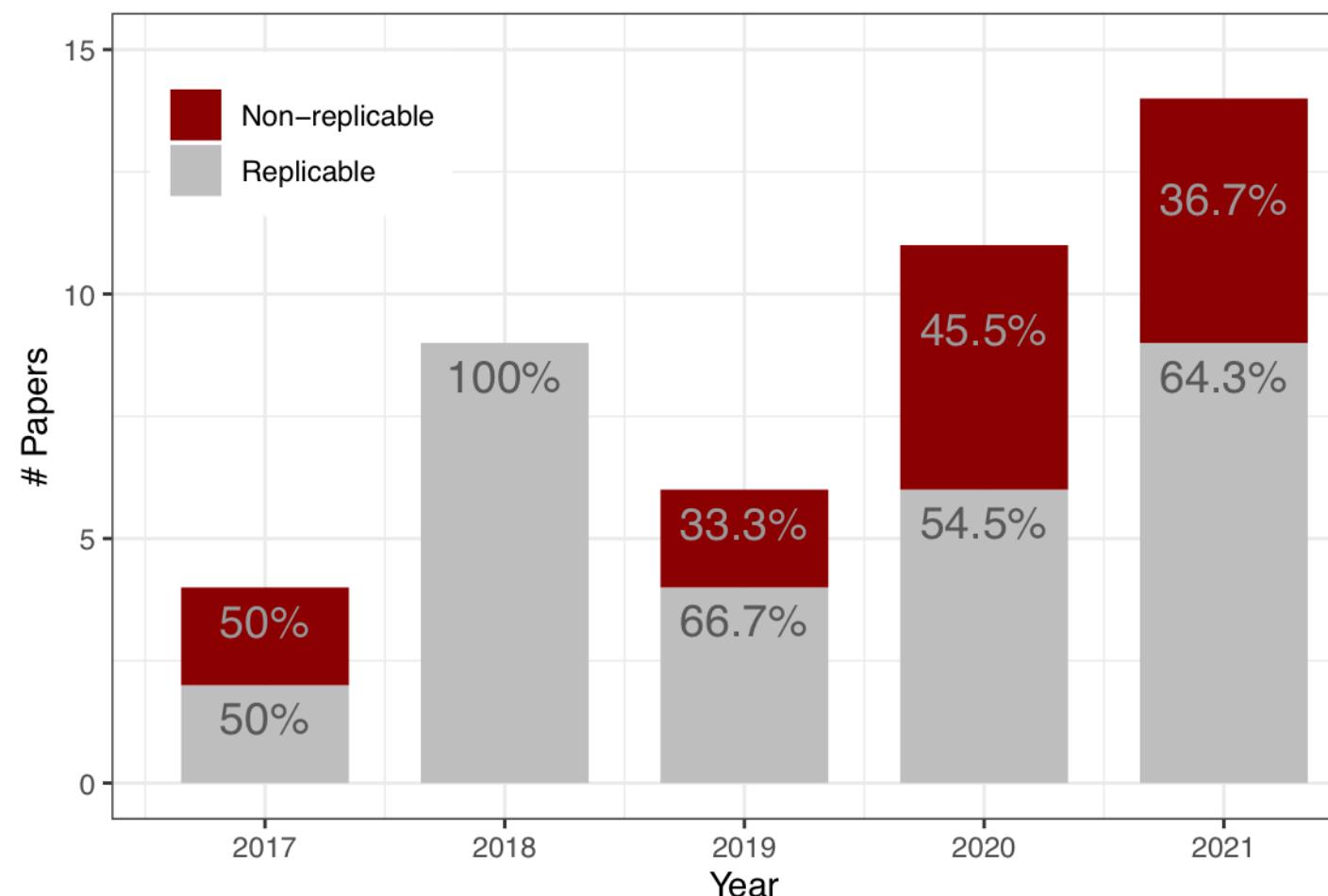
- 许多缺失值 (且缺失值并非随机)
- 约1/3的研究中发现了很强的事前趋势 (pretrends)
- 约1/3的研究中发现了延续效应 (Carryover effects)
- 低估不确定性 (Underestimating uncertainties) 仍然是一个大问题—2/5的研究检验效力不足 (Underpowered)
- 整体而言，1/3的研究发现能很好地得到数据的支撑

一线希望

- 对于观察不出明显的事前趋势 (pretrends) 的研究，异质性处理效应 (HTE) 不是一个严重的问题
- 异质性处理效应稳健标准误 (HTE-robust estimators) 之间没有非常大的差异
- 事件研究散点图 (Event study plots) / 动态处理效应平面图 (dynamic treatment effects plots) 是一个强有力的检验工具

论文复现的样本 (2017-2021)

- 论文的挑选标准：
 - 使用面板数据进行分析，这是作为一个支撑因果论证的关键证据
 - 使用至少一个包括二分处理变量的线性模型（DID, TWFE, LDV等）
 - 排除面板断点回归（panel RD）或工具变量法（IV designs）
 - 挑选作者偏好的函数形式
- “可复制的”：所选择的方程形式的点估计可以准确地复现。

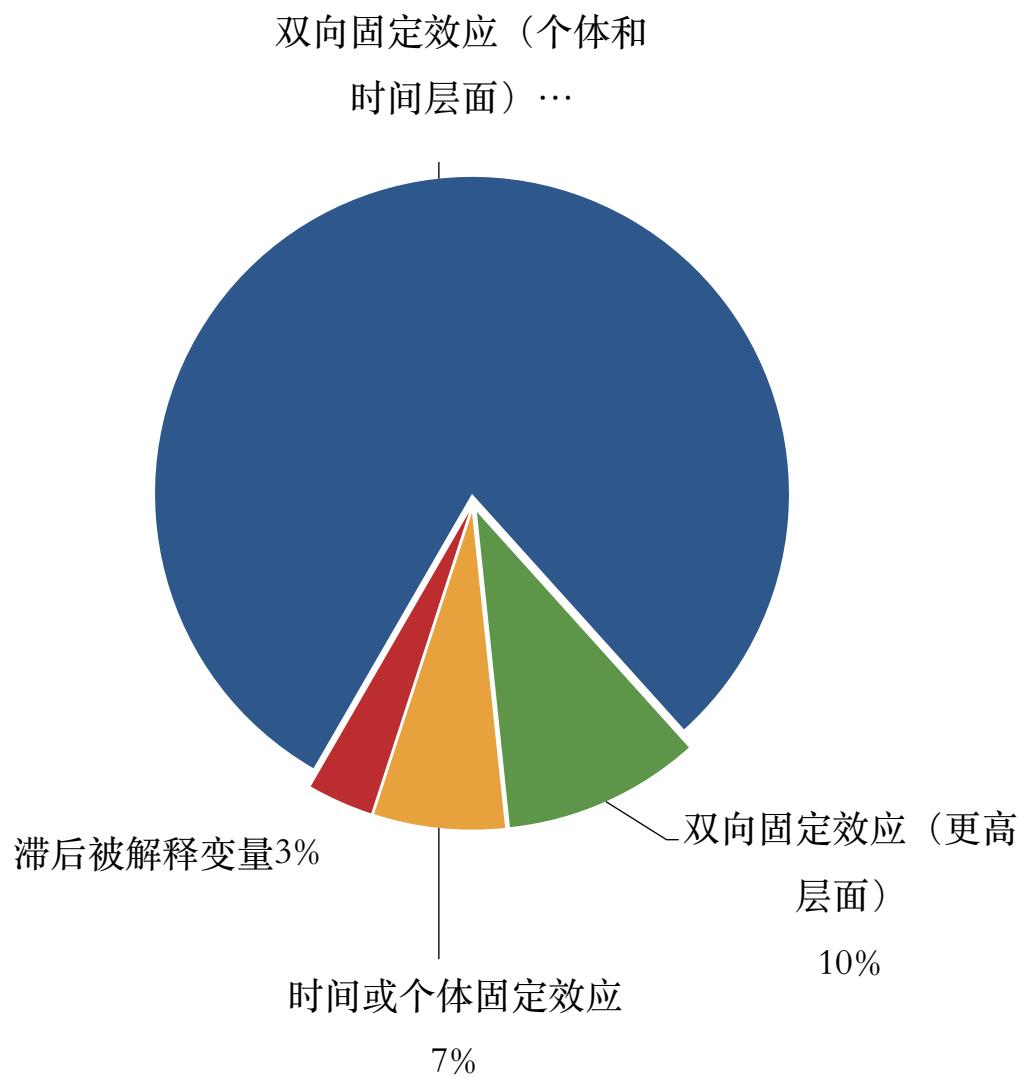
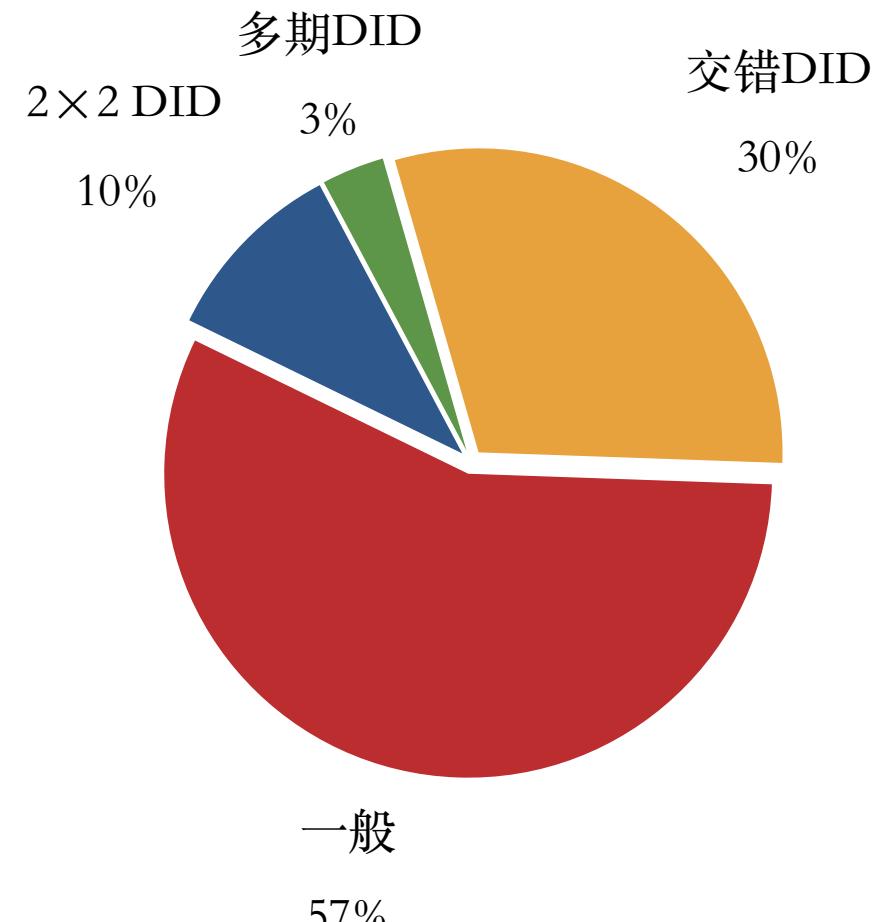


论文复现的样本 (2017-2021)

- 论文的挑选标准：
 - 使用面板数据进行分析，这是作为一个支撑因果论证的关键证据
 - 使用至少一个包括二分处理变量的线性模型（DID, TWFE, LDV等）
 - 排除面板断点回归（panel RD）或工具变量法（IV designs）
 - 挑选作者偏好的函数形式
- “可复制的”：所选择的方程形式的点估计可以准确地复现。

期刊	小计	数据不完整	代码有误	可复制的
APSR	10	4	1	5 (50%)
AJPS	15	4	2	9 (60%)
JOP	19	3	0	16 (84.2%)
总计	44	11 (25%)	3 (6.8%)	30 (68.2%)

论文的设定和实践



Variance Estimator	Count	Percentage
Cluster-robust SE or PCSE	30	100.0%
Cluster-bootstrapped SE	4	13.3%
LDVs and time trends		
LDVs	5	16.7%
Unit-specific linear time trends	6	20.0%
Higher-level time trends	9	30.0%
Data visualization		
Over-time group means	11	36.7%
Dynamic treatment effects	13	43.3%
Neither	17	56.7%

分析步骤

步骤1. 通过画出原始数据的图，了解背景、设定、数据结构；记录关键信息

步骤2. 对使用原始方差估计量（variance estimator）和聚类自抽样（cluster-bootstrap）的偏好函数形式的主要结果进行复制

步骤3. 使用双向固定效应和多种异质性处理效应稳健估计量对平均处理效应（ATT）和动态效应进行估计，包括：

- 交互加权（IW）（Sun & Abraham 2021） — 交错DID
- CSDID (Callaway and Sant'Anna 2021) — 交错DID
- 面板匹配（PanelMatch）（Imai, Kim & Wang 2021）
- 插补法（Fect）（Borusyak, Jaravel and Spiess 2021; Liu, Wang & Xu 2022）

步骤4. 基于Fect来检验（Liu, Wang & Xu 2022）

- 可视化检验（Visual inspection）
- 检验事前趋势（F检验和安慰剂检验）
- 检验延迟效应（carryover effects）

案例：同族群的政治动员

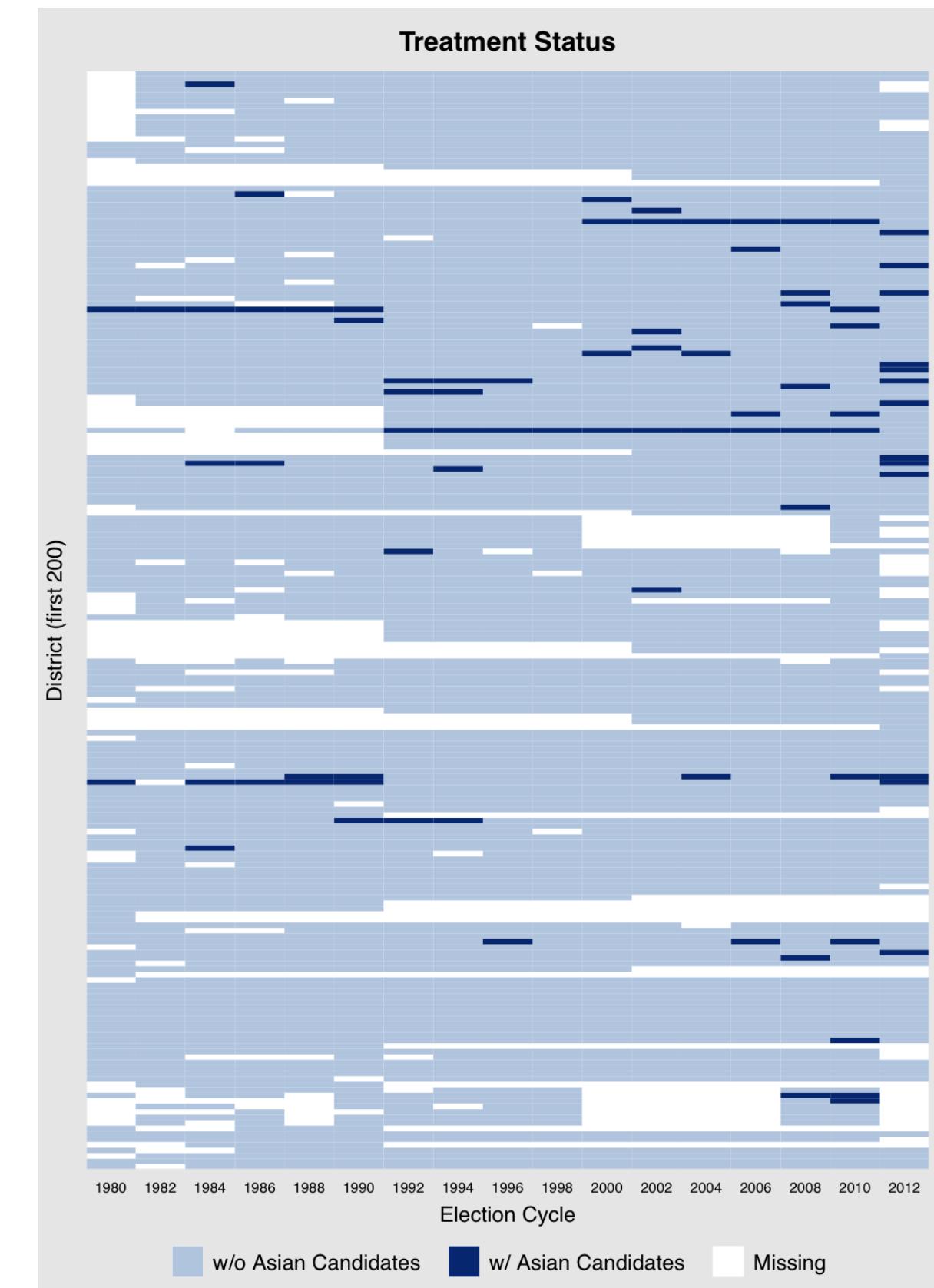
在美国国会选举中的少数族裔候选人会去动员同族的捐款人吗? (Grumbach & Sahn 2020)

处理变量：少数族裔候选人

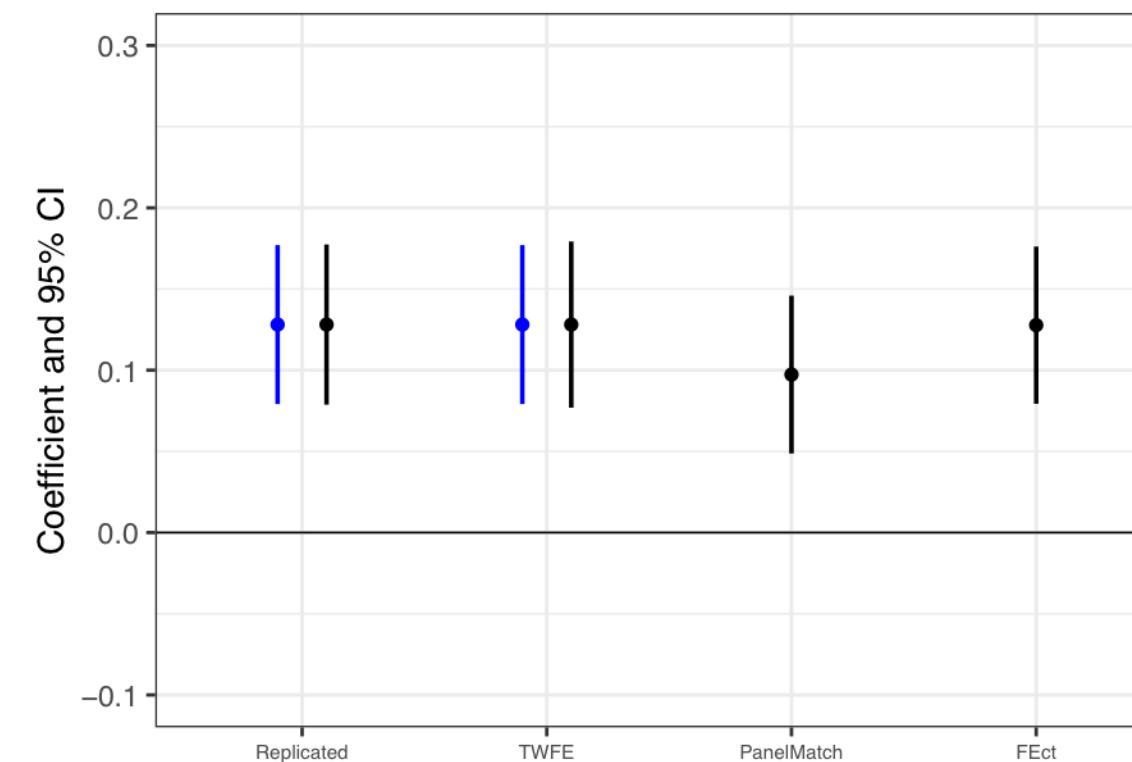
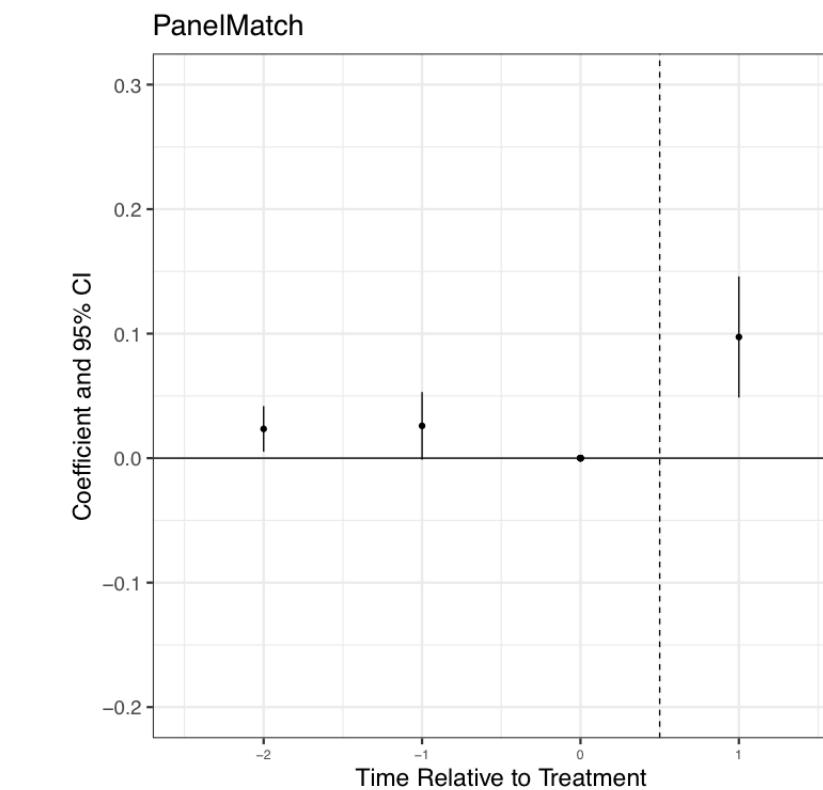
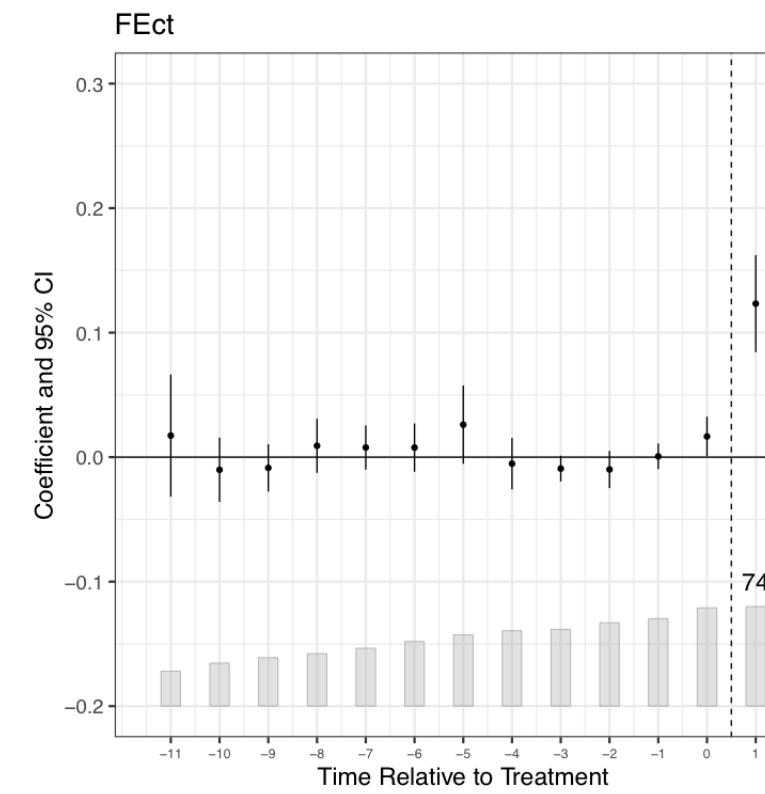
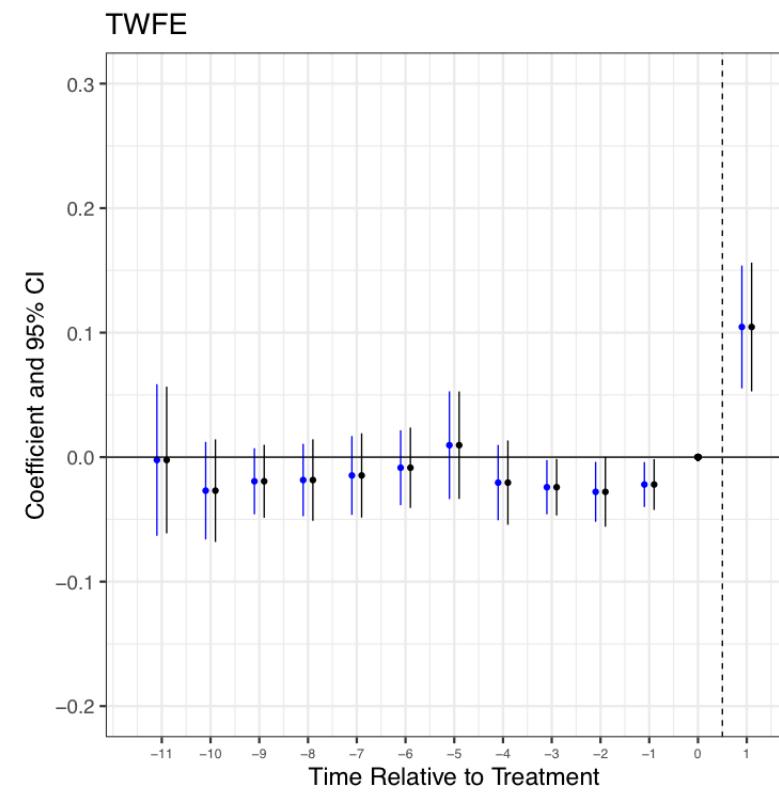
结果变量：同族捐款人在选举中的比例

样本规模：

- N: 489
- T: 17 (1980-2012)
- #obs: 7,141

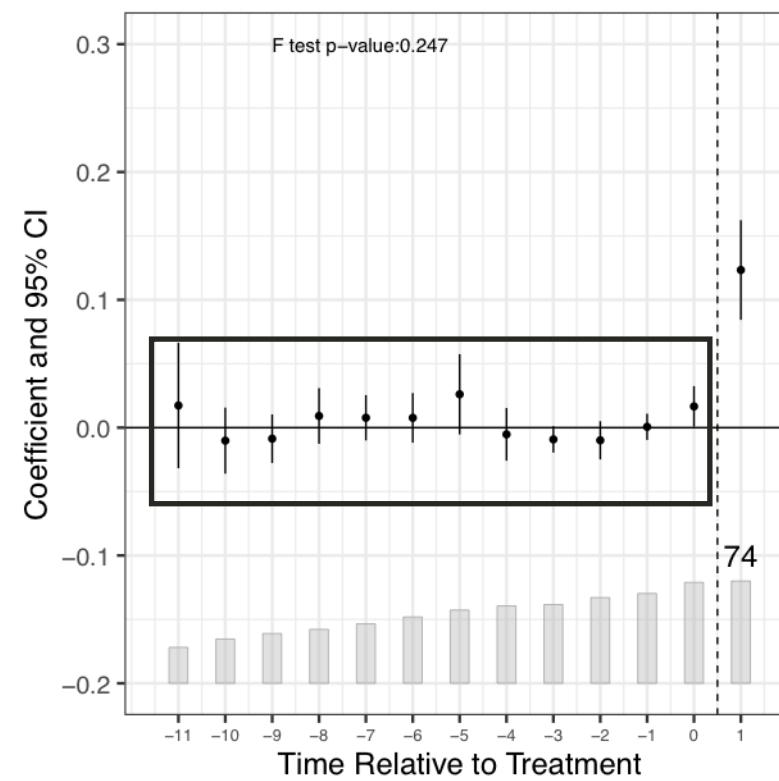


案例：同族动员



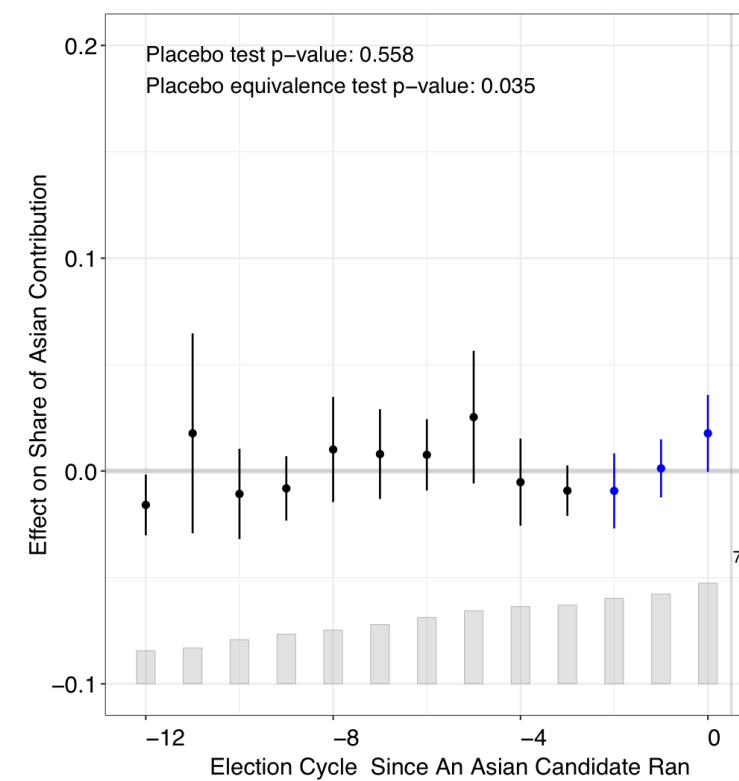
比较估计结果

检验



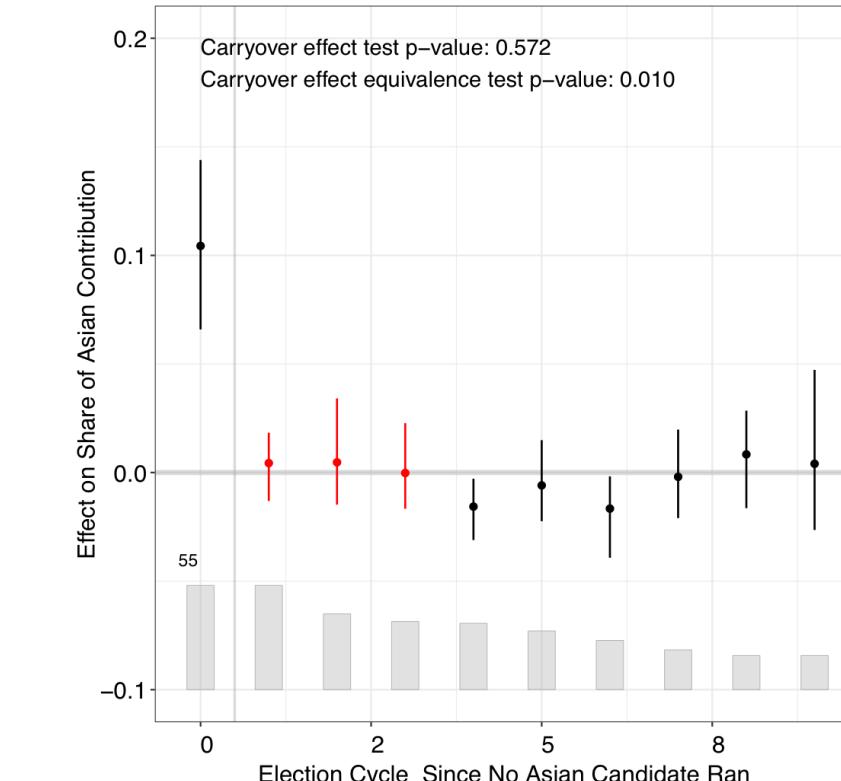
对“没有事前趋势”做F检验

$p = 0.247$



安慰剂检验

$p = 0.558$



检验延迟效应

$p = 0.572$

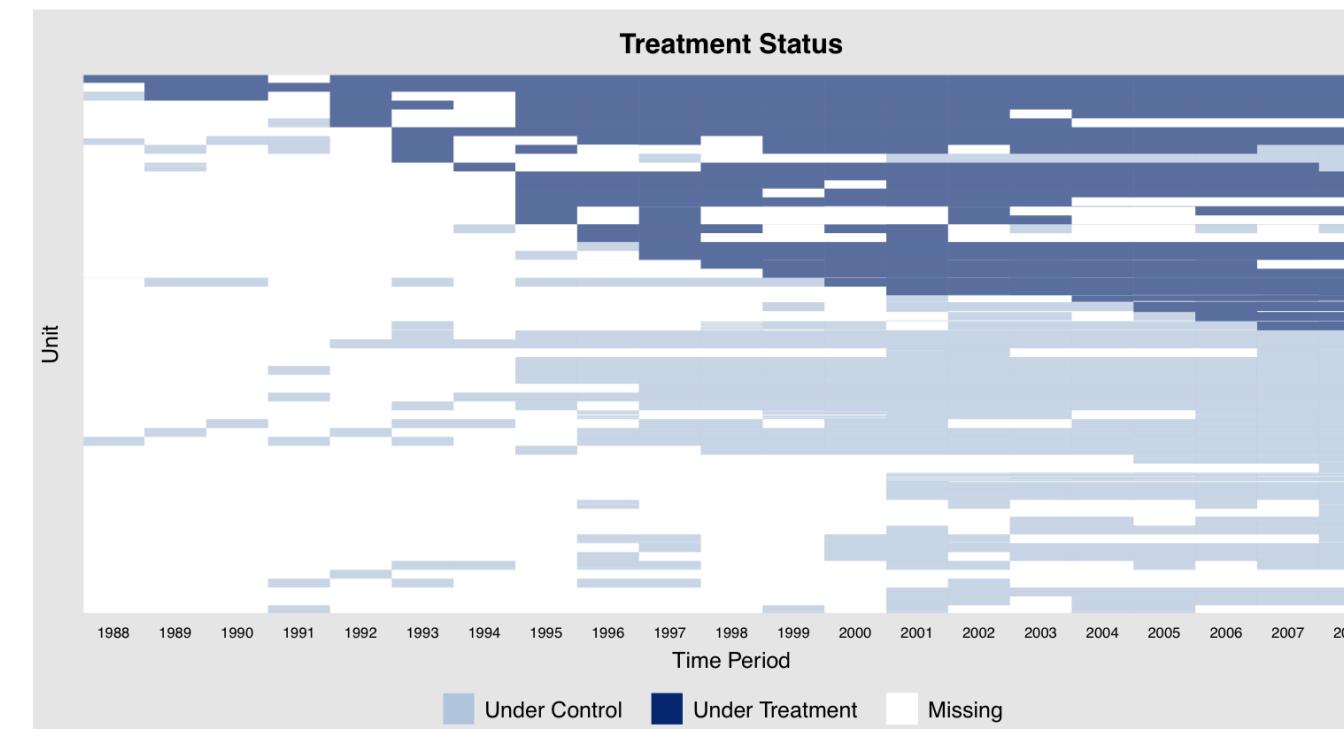
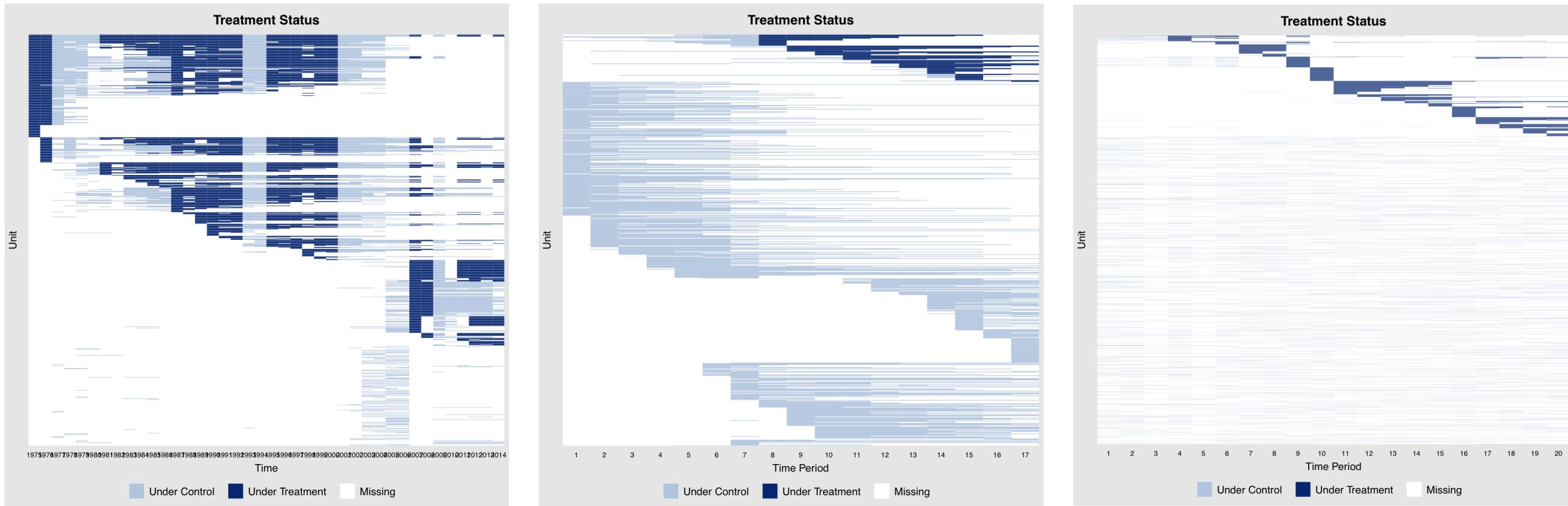
评估事前趋势 (Pretrend)

系统性评估

共性问题

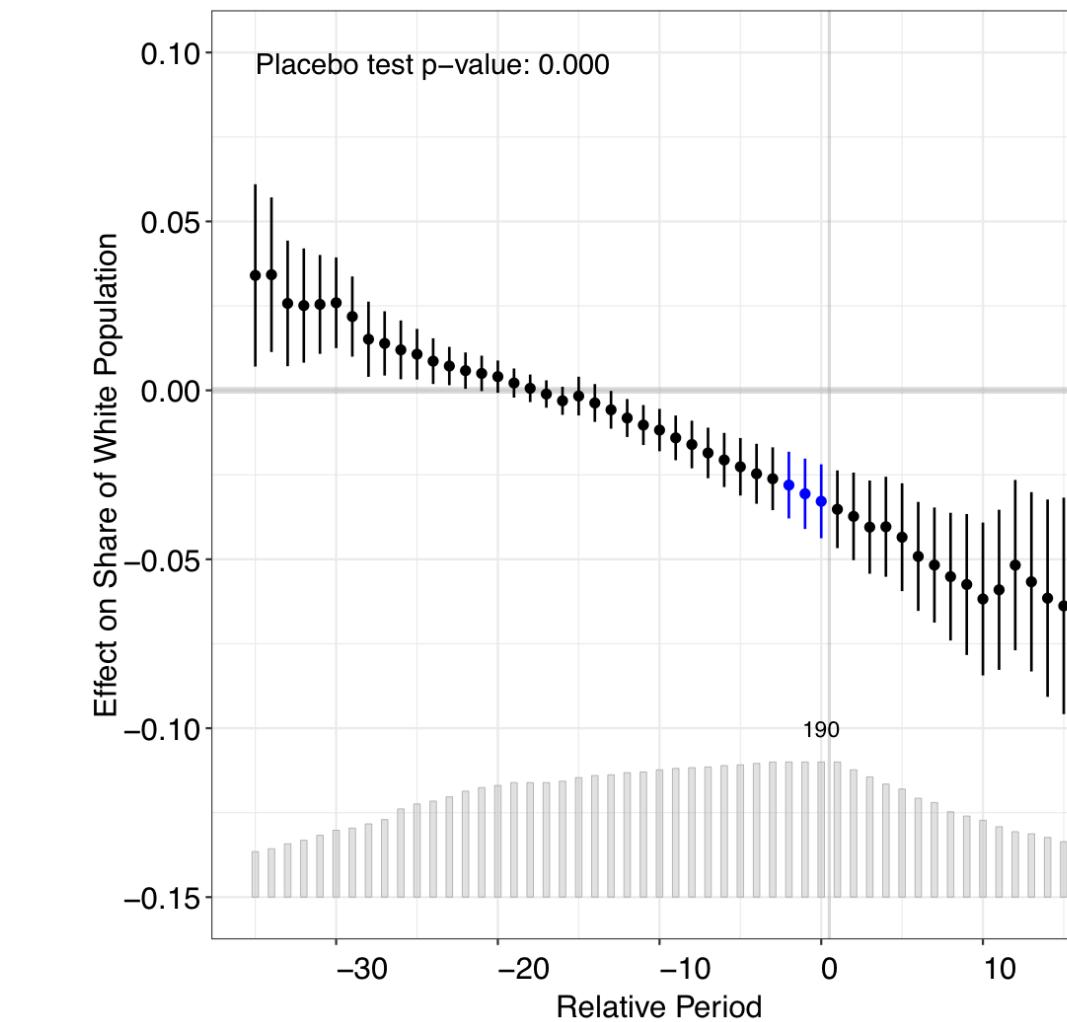
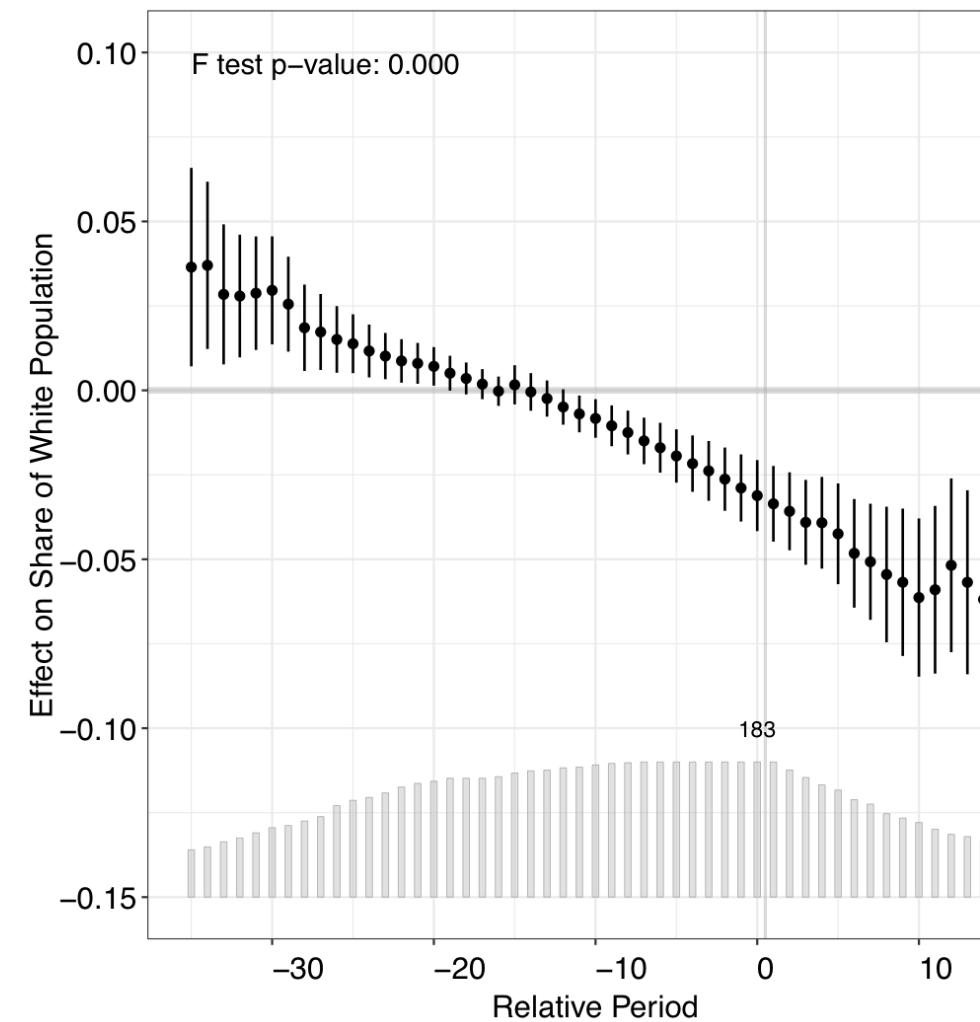
1. 数据缺失的普遍性及模式
2. 明显的事前趋势
3. 延续效应
4. 低估不确定性

问题1：数据缺失



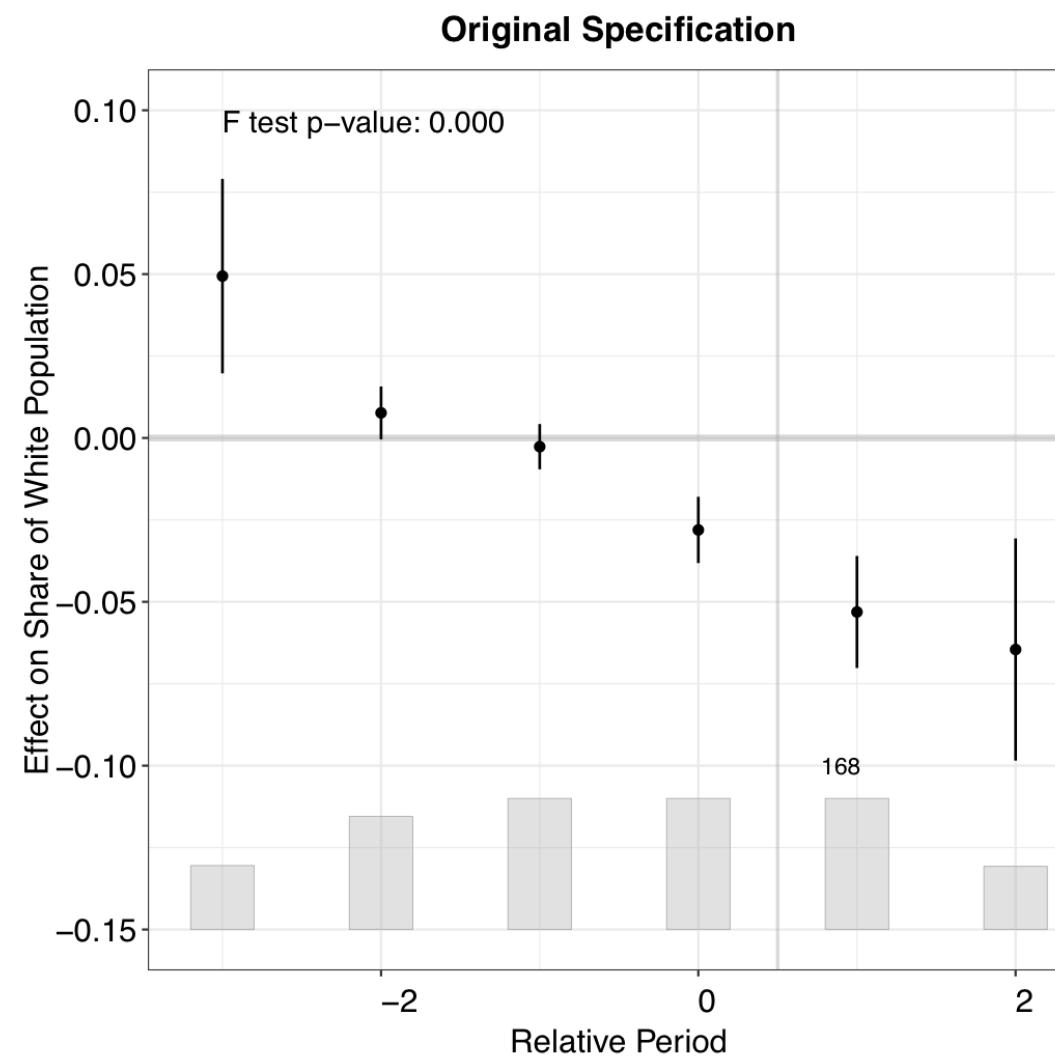
问题2：明显的事前趋势

- 作者研究城市土地使用限制的诉讼对城市居民结构的影响
- 基于人口普查的人口统计学数据很多是被插值的 (interpolated)



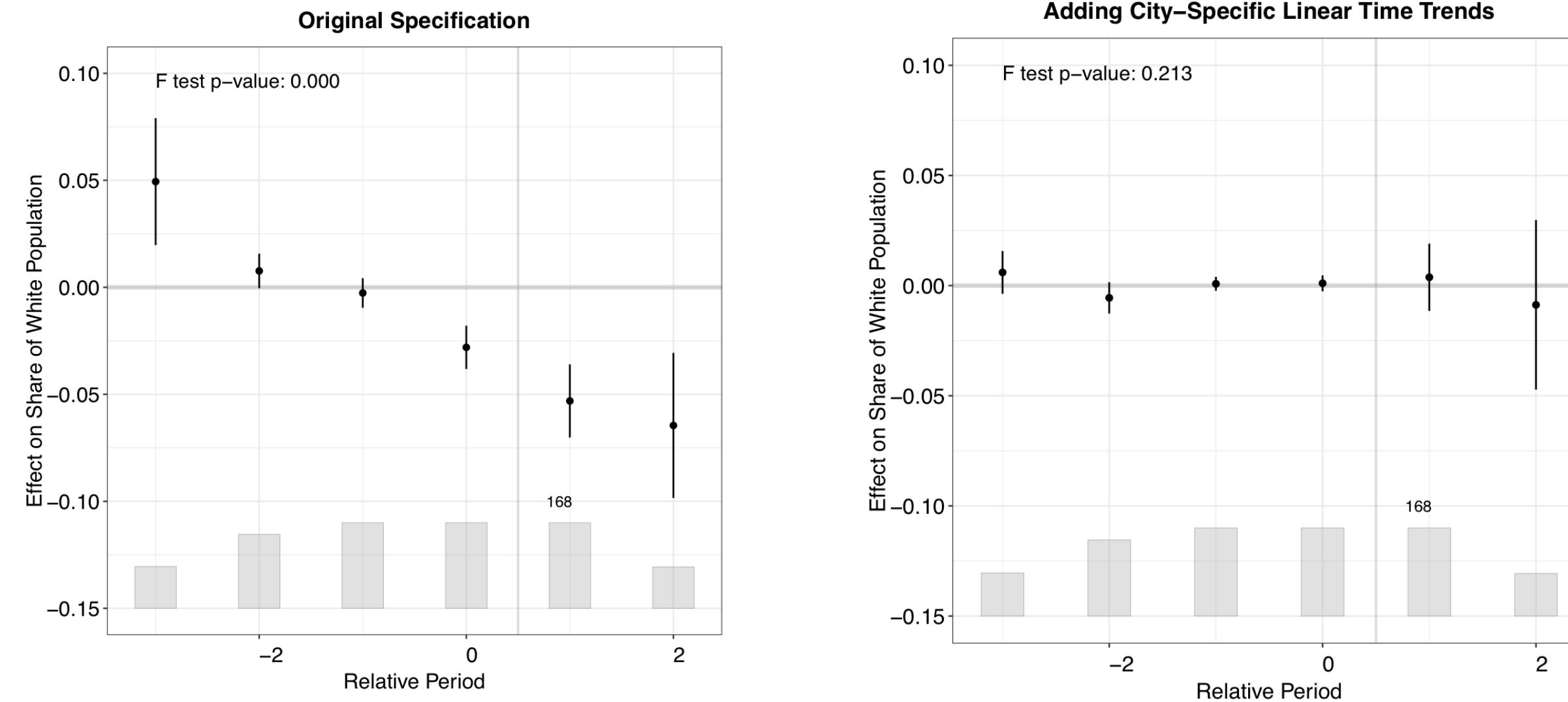
问题2：明显的事前趋势

- 基于人口普查的人口统计学数据很多是被插值的 (interpolated)
- 当我们移除插值数据 (interpolated data) 时，研究结果是很相似的

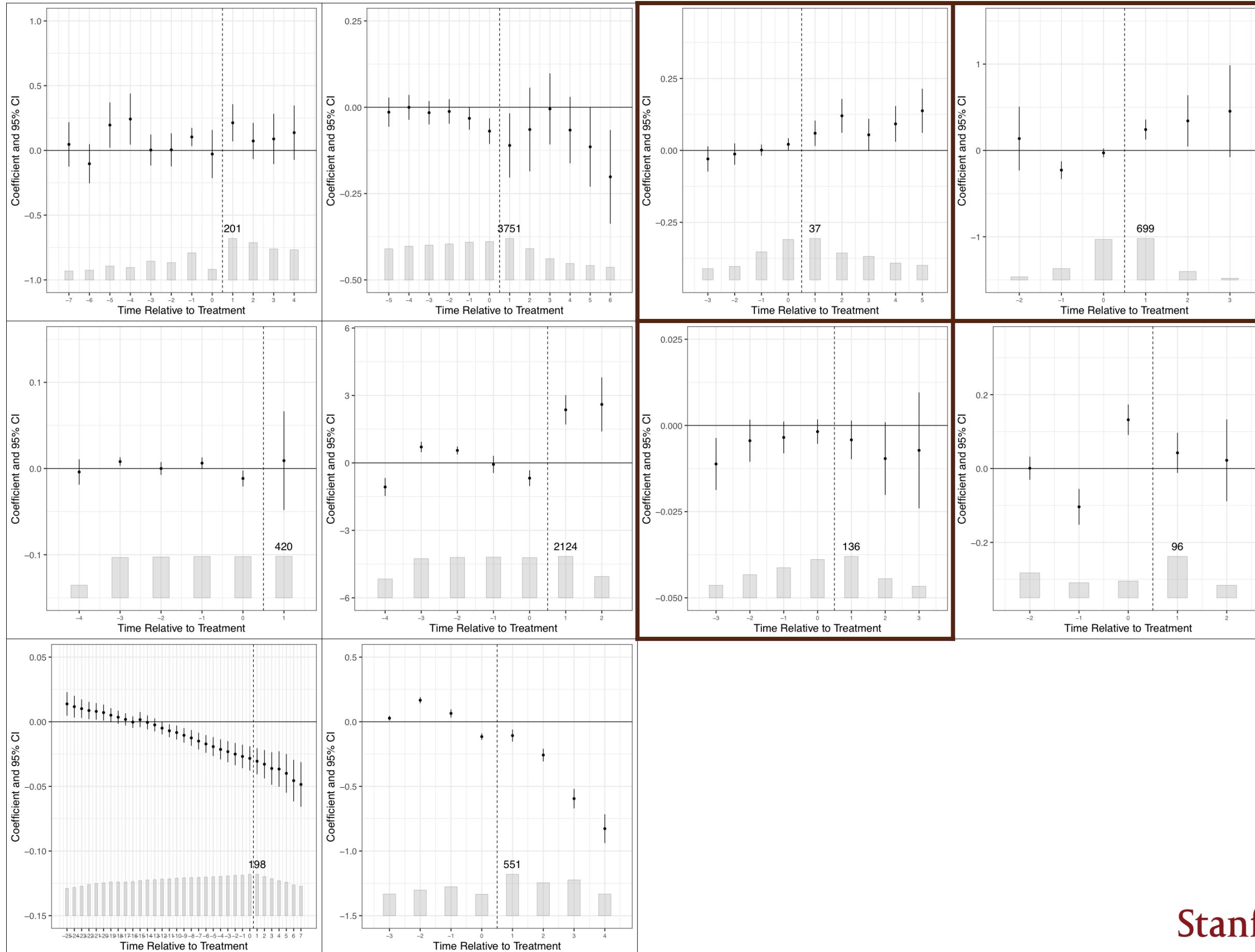


问题2：明显的事前趋势

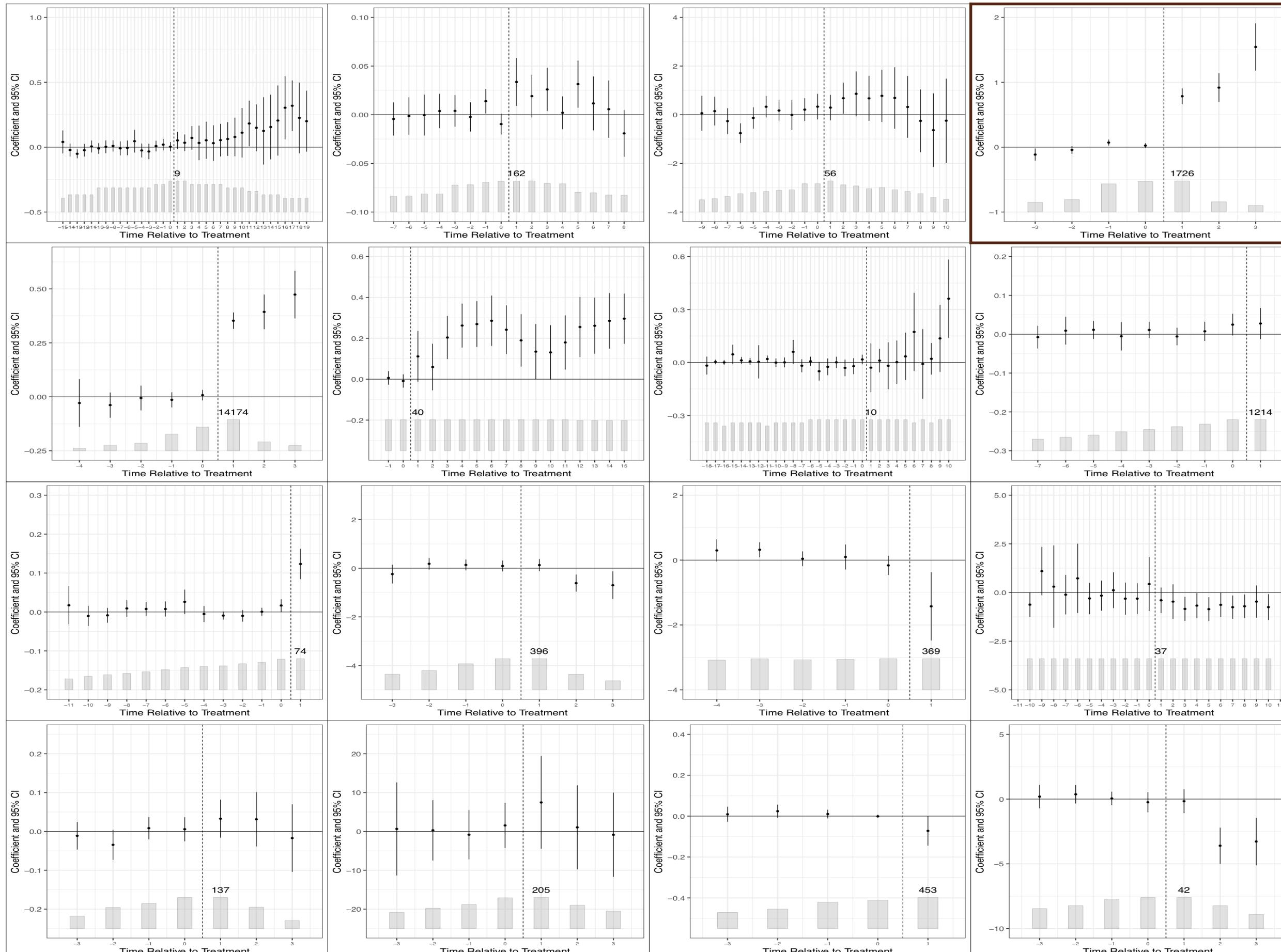
- 当我们移除插值数据 (interpolated data) 时，研究结果是很相似的
- 当我们增加了城市的时间趋势变量，负向的结果消失了



“坏的”事前趋势

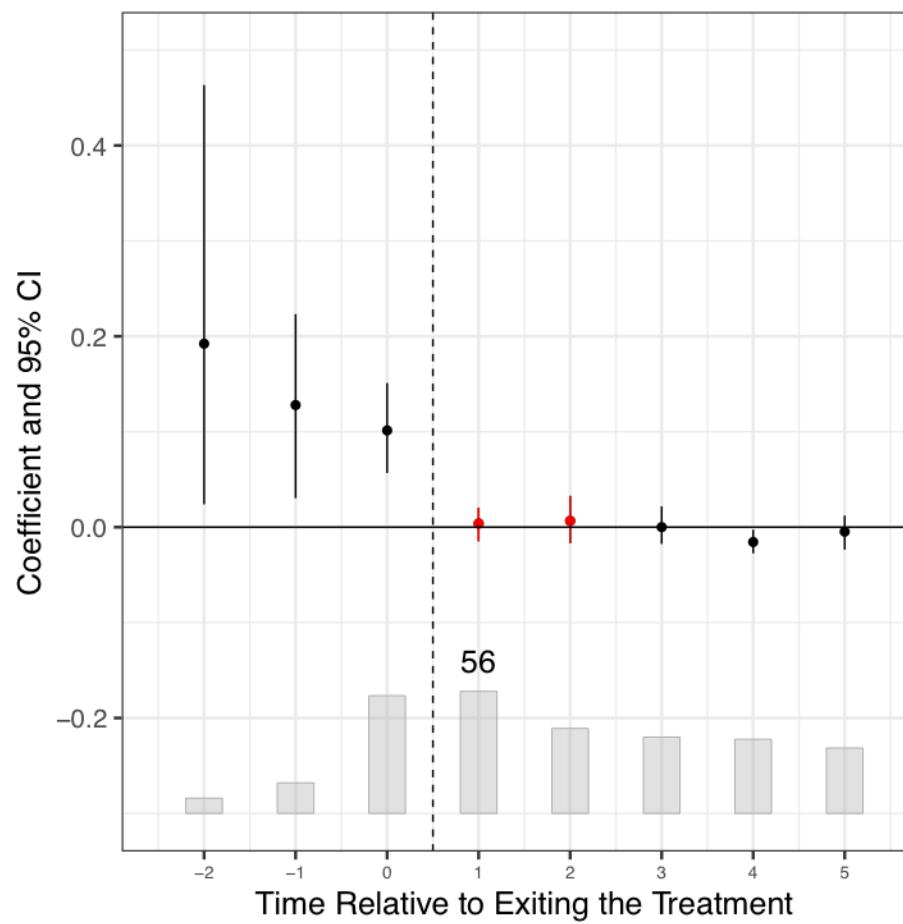


“好的”事前趋势



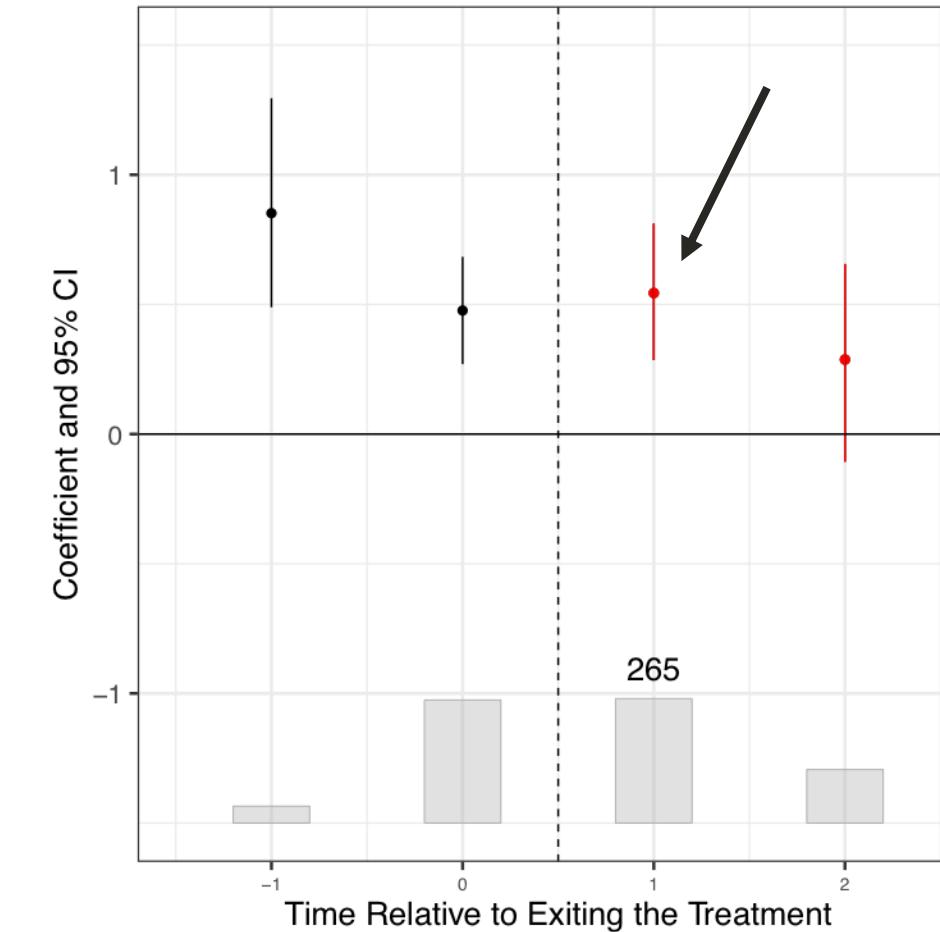
问题3. 延续效应

没有延续效应



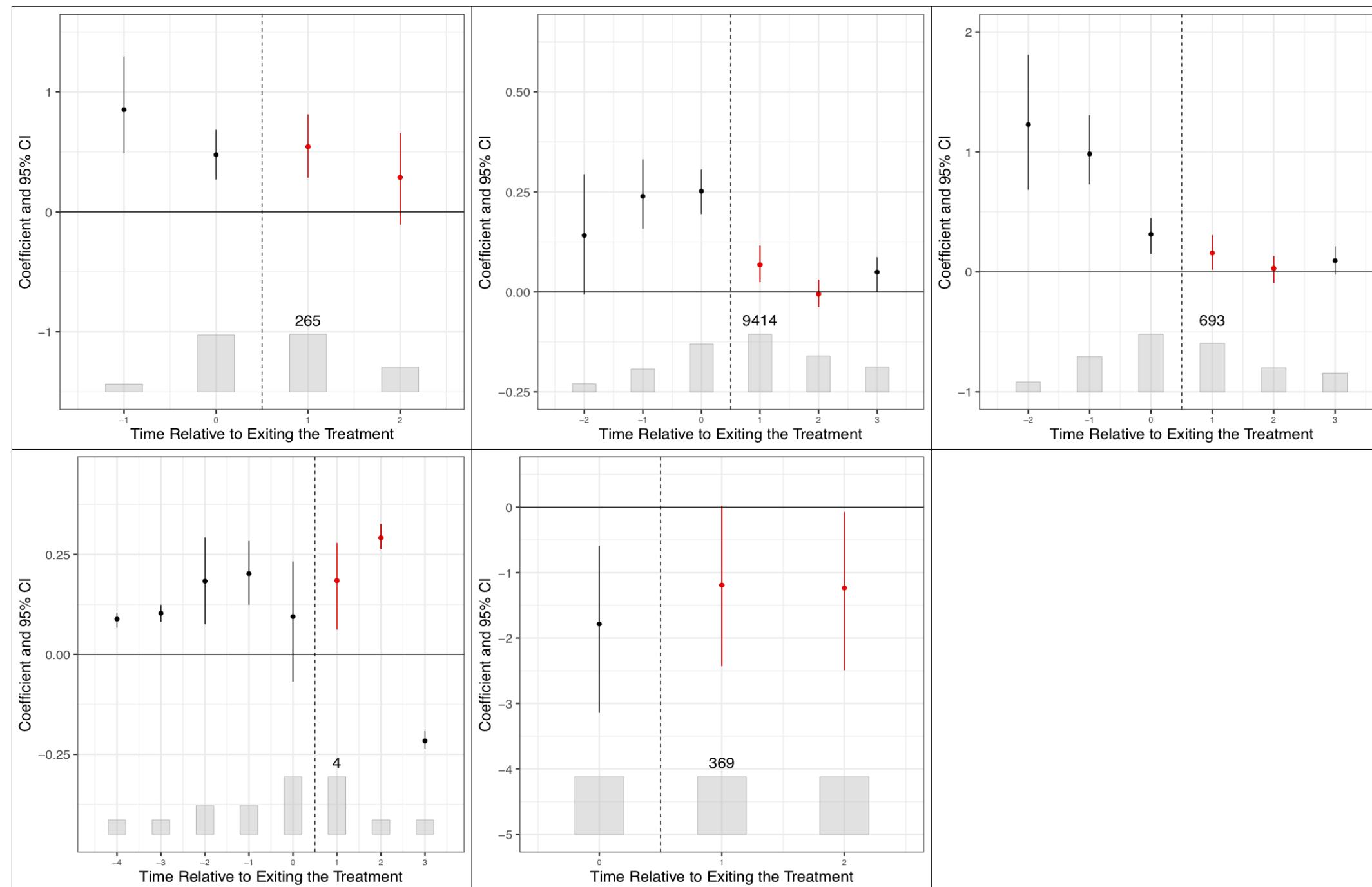
Grumbach and Sahn (2020)
 P 值 : 0.329

延续效应

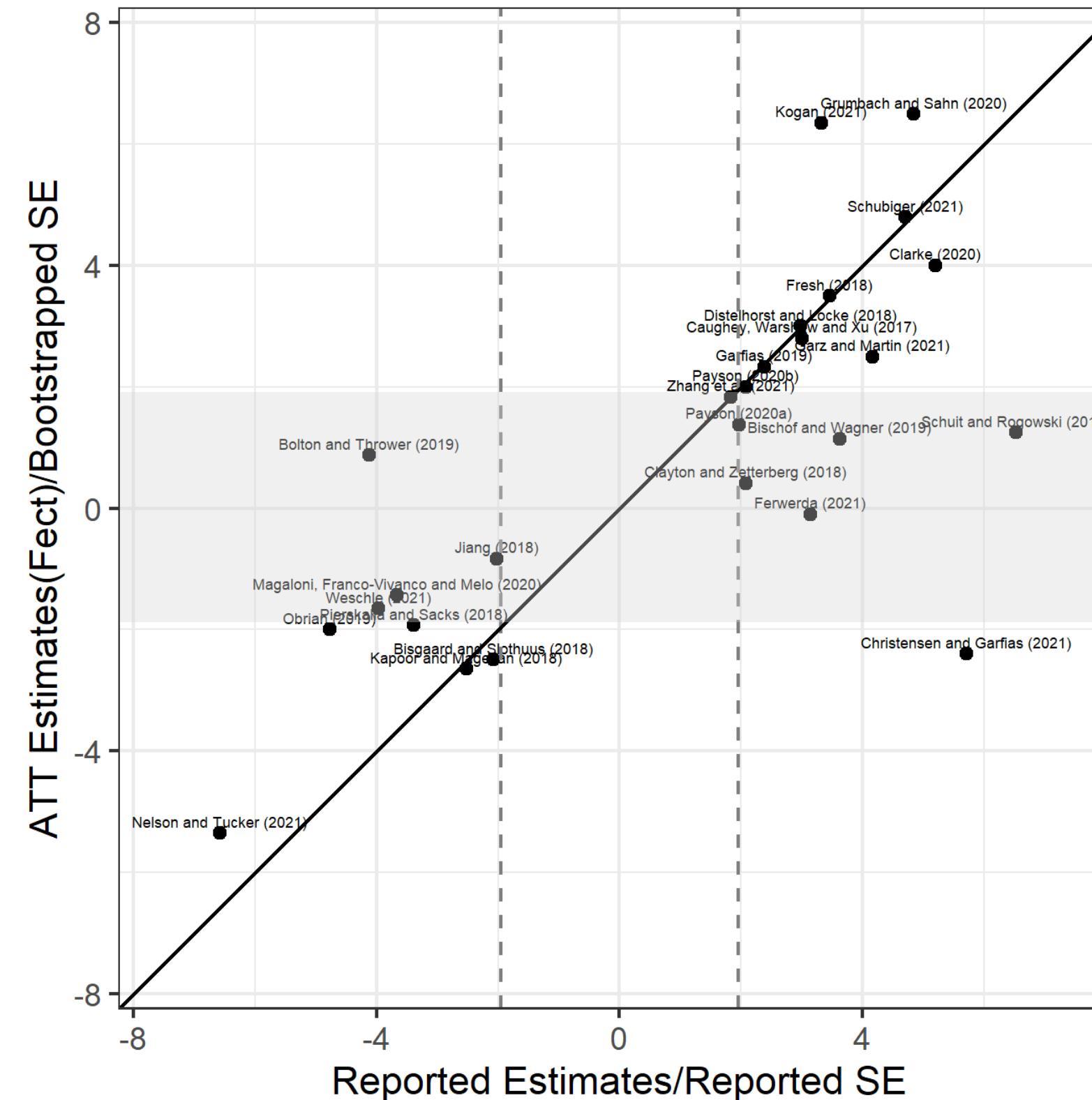


Distelhorst and Locke (2018)
 P 值 : 0.000

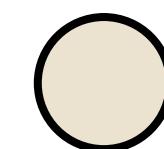
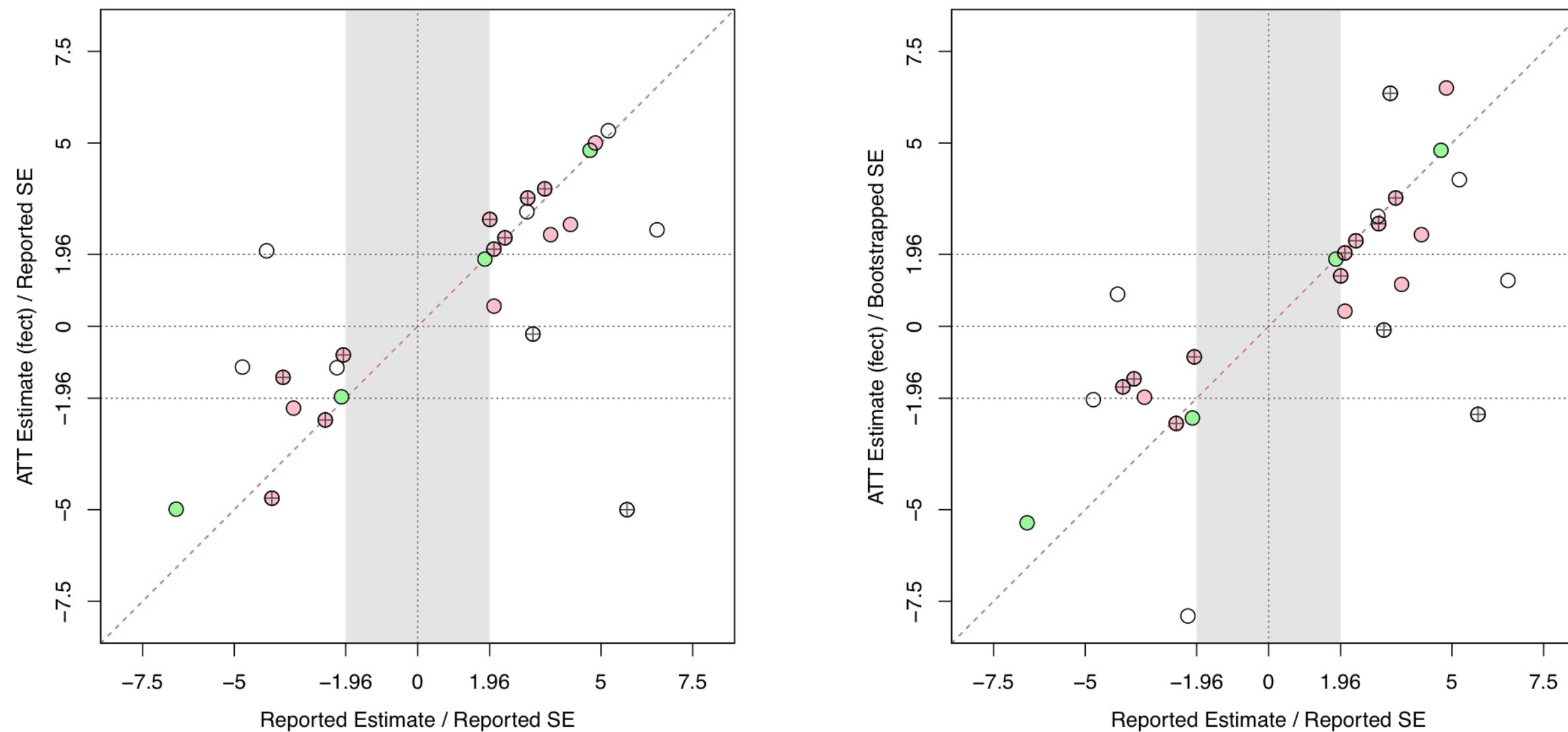
问题3. 延续效应



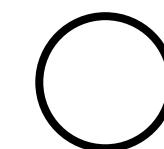
问题4：低估不确定性



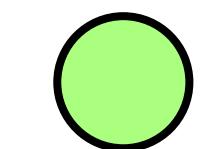
TWFE与异质性处理效应稳健估计量的结论接近



= “好的” 事前趋势



= “坏的” 事前趋势

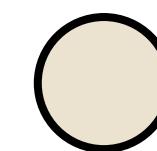
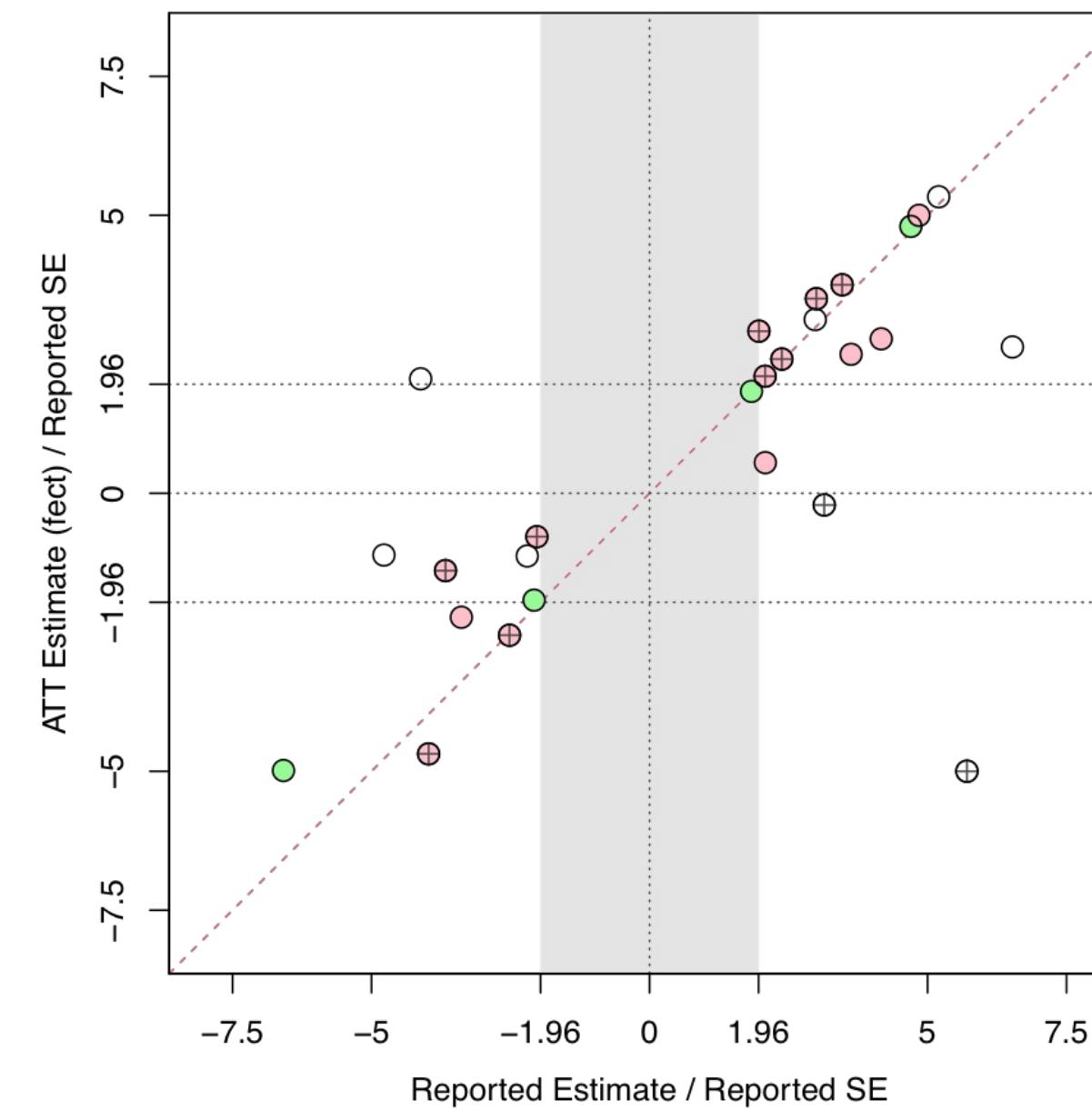


= 缺少事前趋势

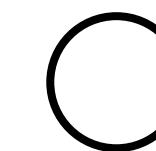


= 包含事件研究法法的原始研究

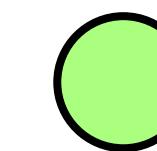
TWFE与异质性处理效应稳健估计量的结论接近



= “好的” 事前趋势



= “坏的” 事前趋势



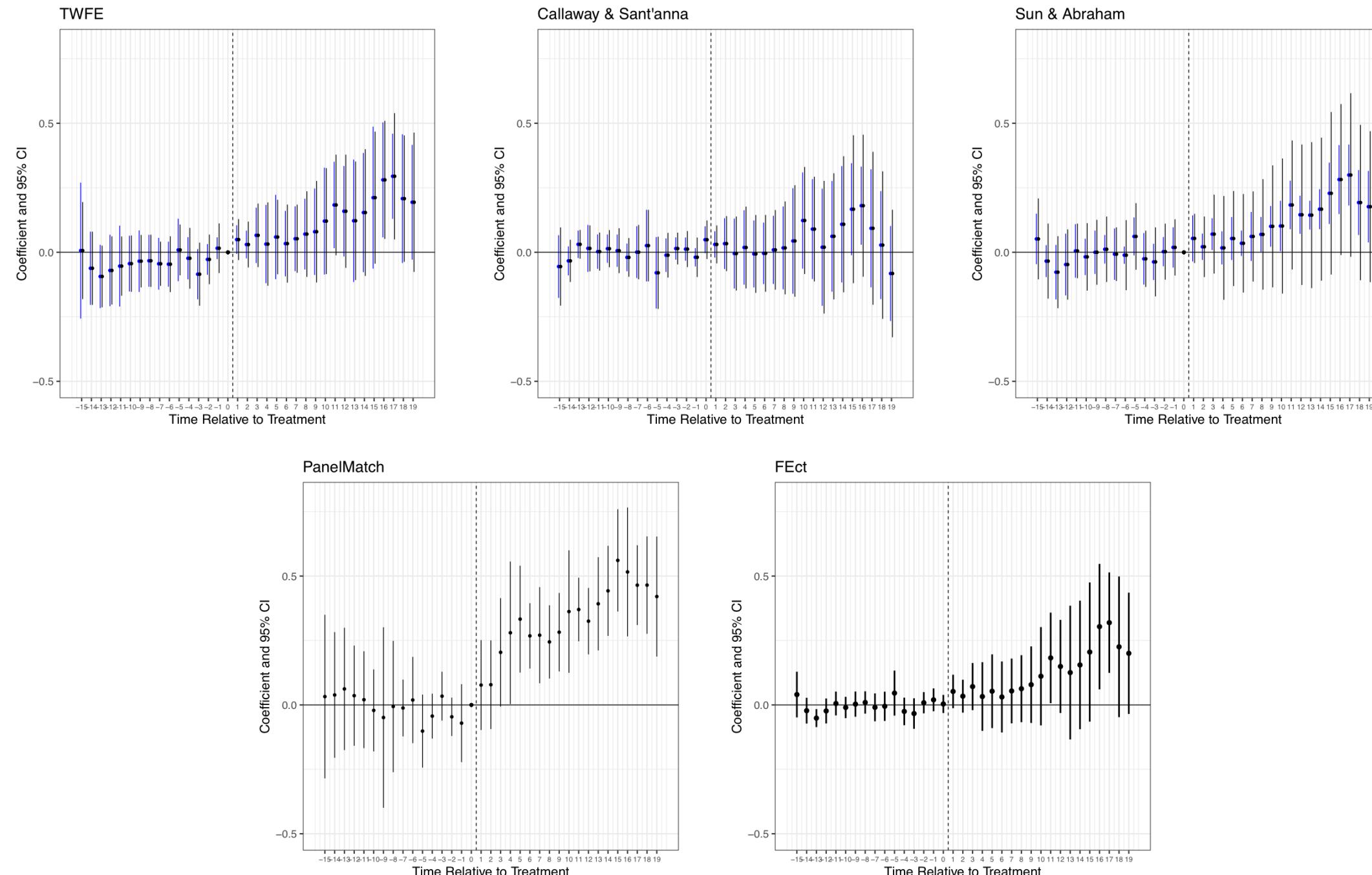
= 缺少事前趋势



= 包含事件研究法法的原始研究

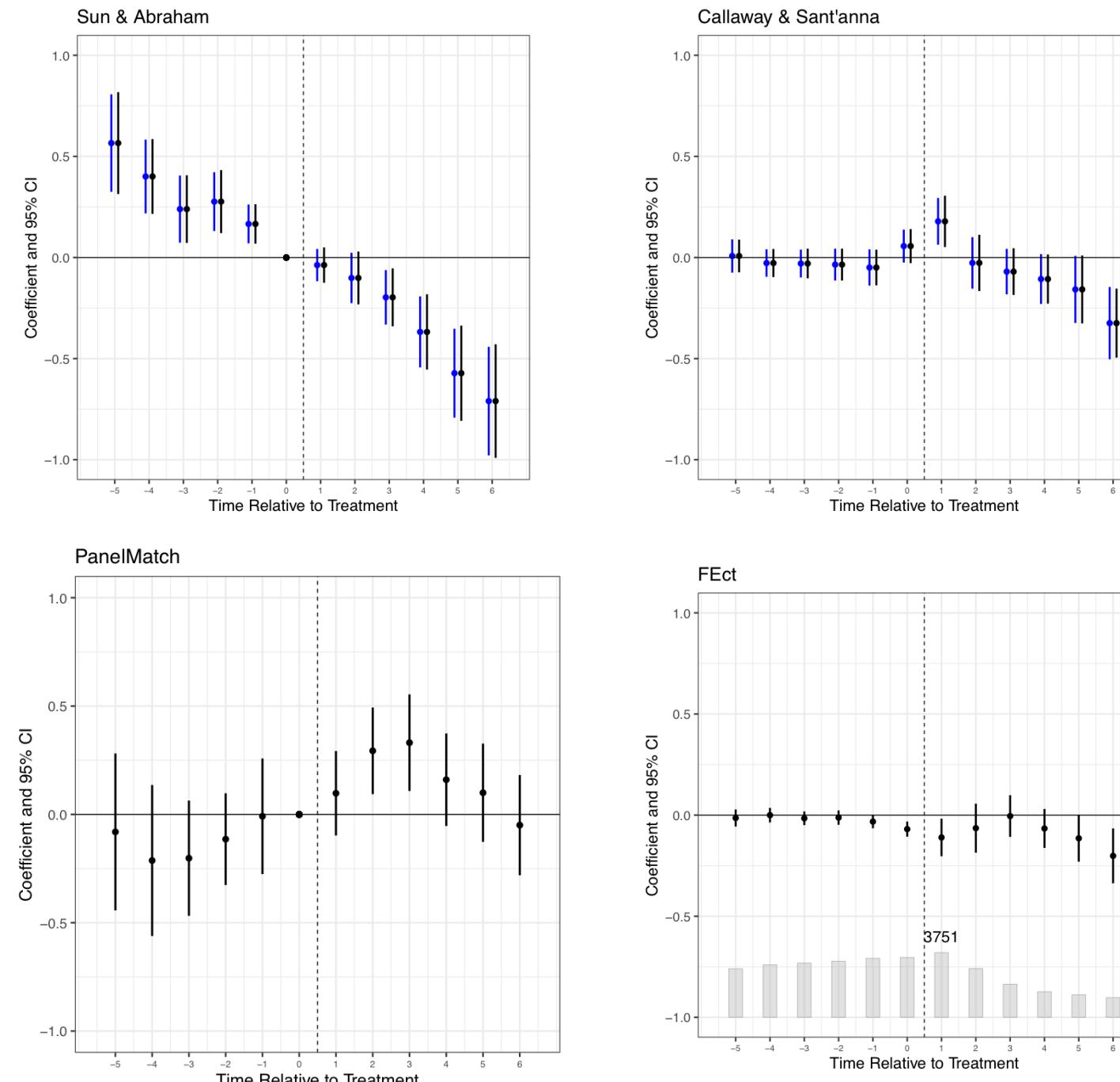
异质性处理效应稳健估计量得到的结论比较接近

- 大体上, 异质性处理效应并不是结论被颠覆的主要原因
- 当没有观察到事前趋势时, 异质性处理效应稳健估计量 (HTE-Robust Estimates) 通常会有相似的结果

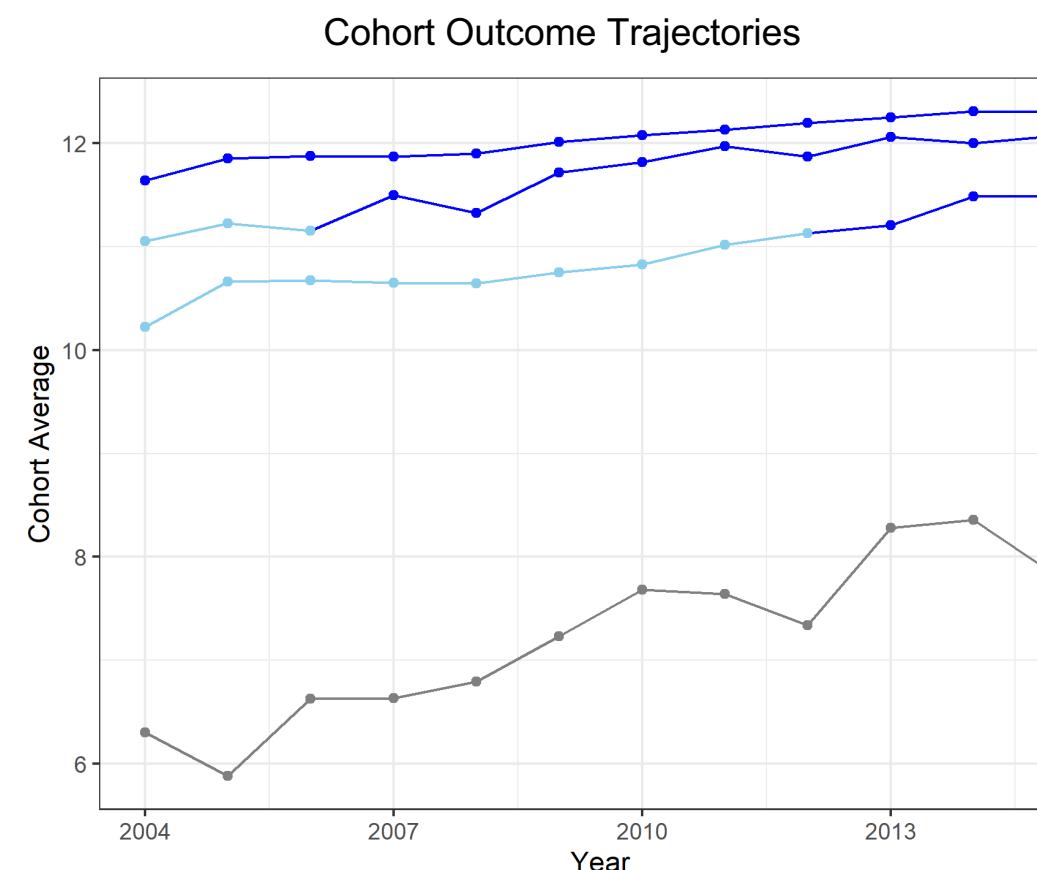
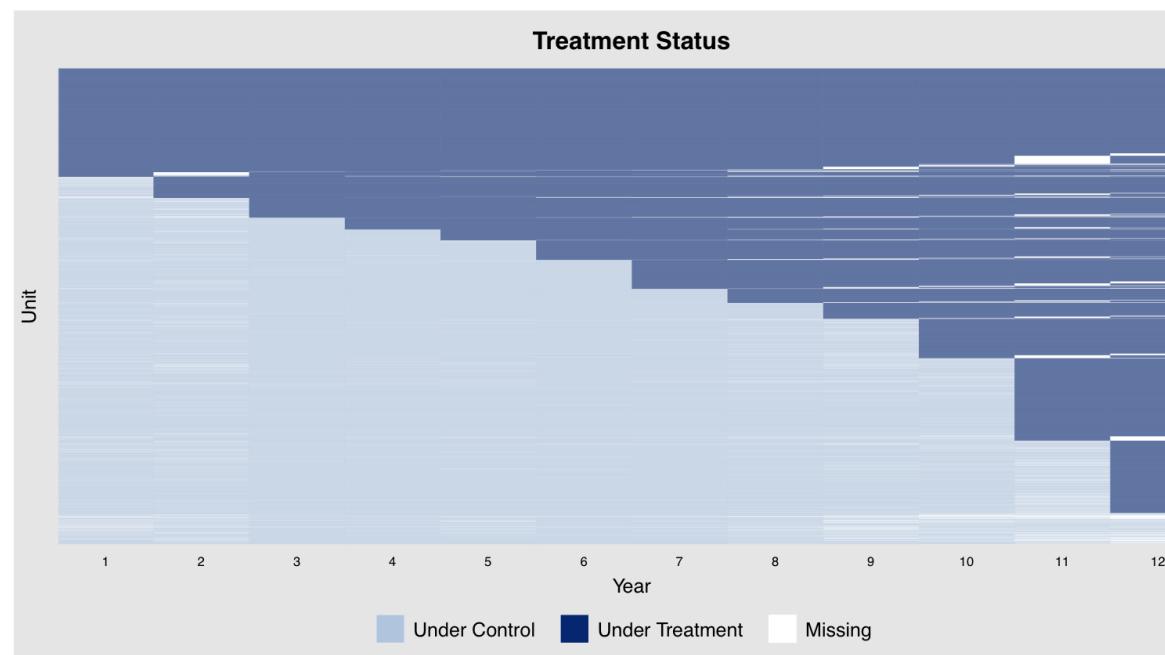


但是也有例外…

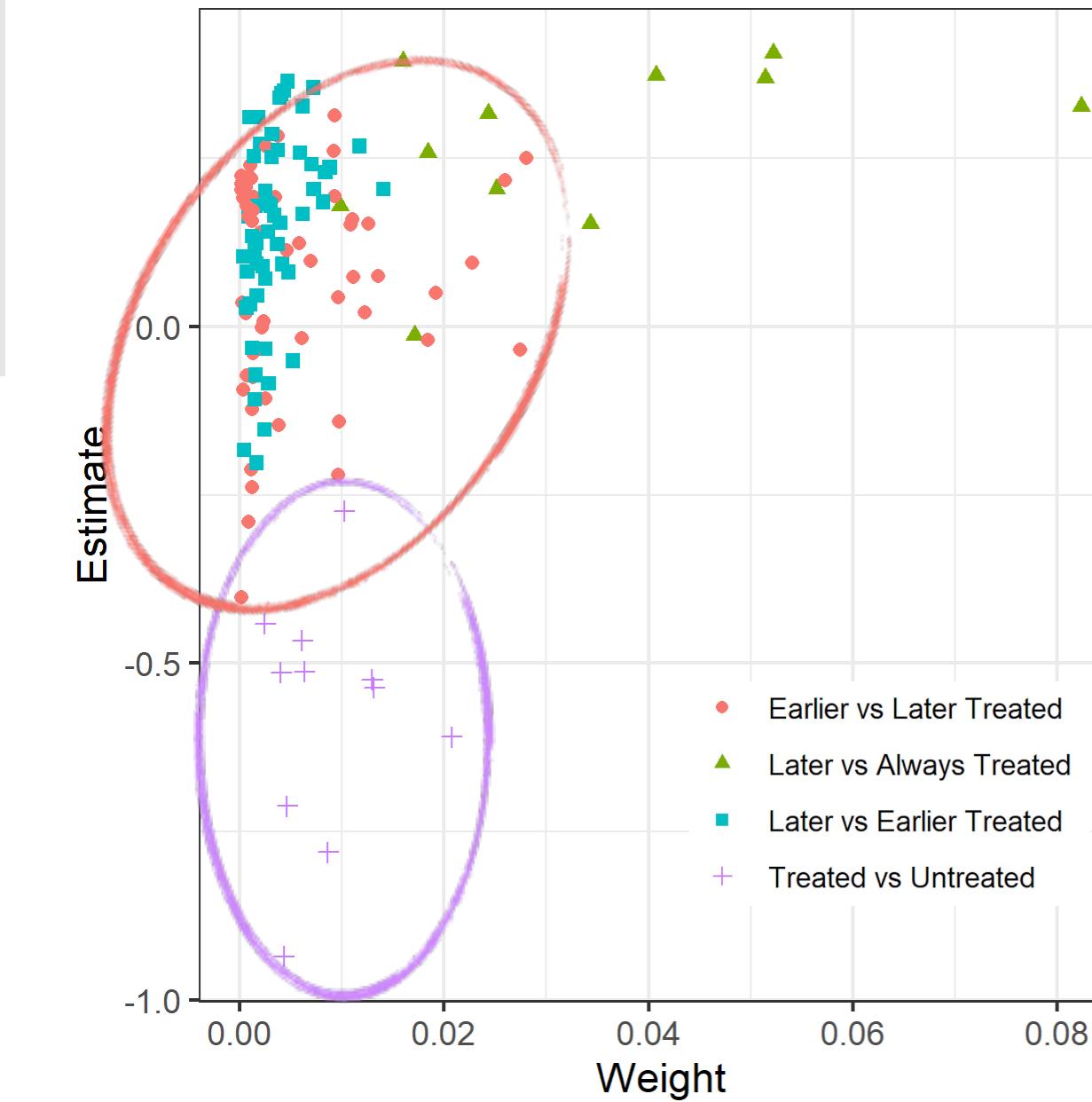
- Christensen & Garfias (2021) 研究了房产证对房地产税收收入的影响
- 估计量之间的不一致



平行趋势假设和异质性问题都很重要



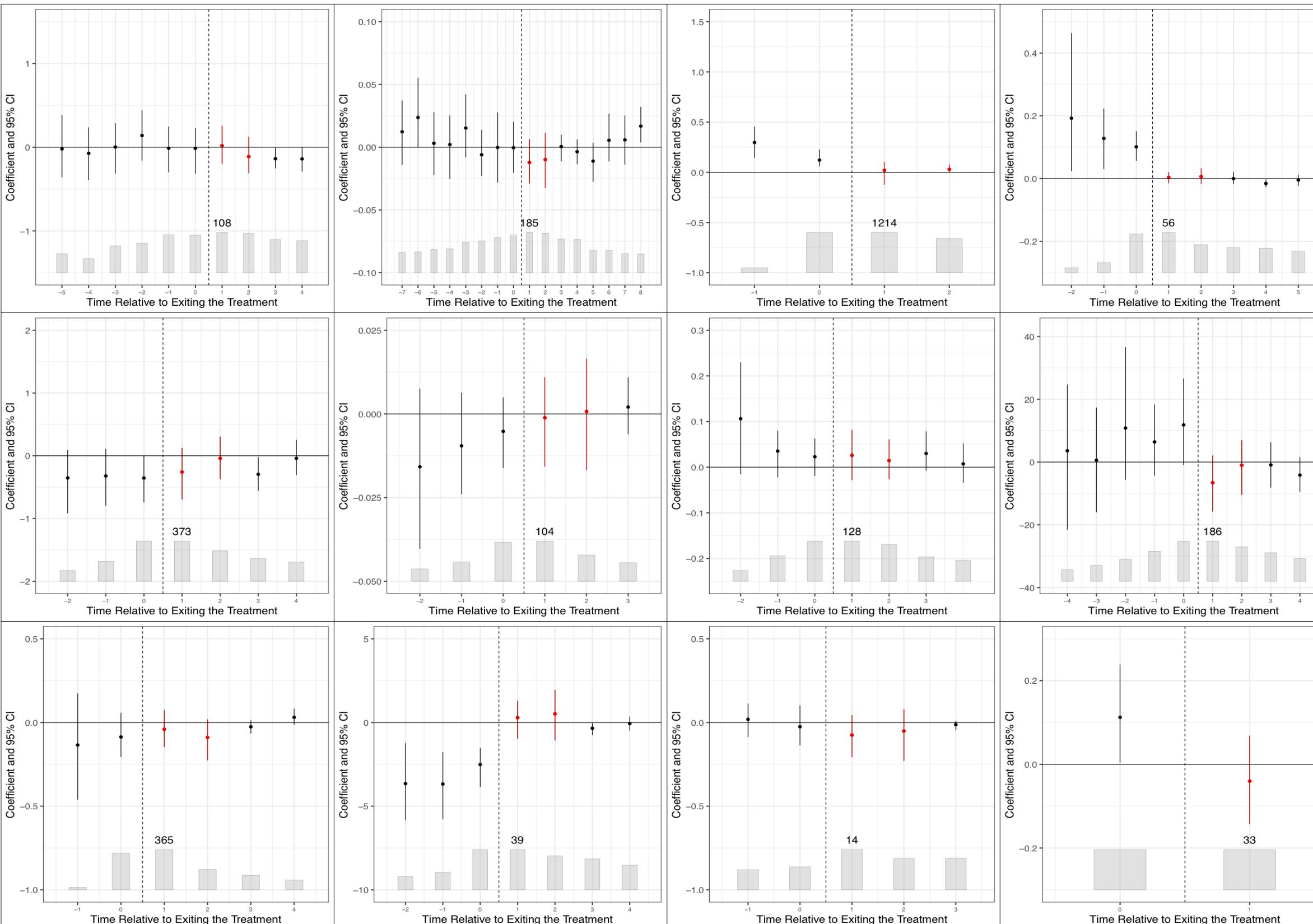
Goodman-Bacon Decomposition



整体评估

	#的研究	分母	比例 %
“坏的” 事前趋势	10	26	38.5
延迟效应	5	17	29.4
缺乏检验效力	13	30	43.3
“好的” 事前趋势， 有/没有足够检验效力	9	26	34.6

没有延续效应



对实证研究的建议

该做和不该做的事

	要	不做
设计胜于分析	带着研究设计来开始实证分析； 如果与干预分配 (treatment assignment) 有关的前期结果不是主要担心的问题，那就继续推进研究吧	盲目使用现有的数据跑回归
对研究设计的讨论	清晰地阐明研究设计及相关等识别假设	将研究设计等同于结果模型 (outcome models)
画出原始数据的图	画出原始数据的图，以更好地理解研究情境、数据缺失情况、处理变量和结果变量数值变化的来源，以及单变量或双变量的分布	不看数据，直接跑回归
估计量	选择异质性稳健估计量 (HTE-robust estimators)，时常画出估计的动态处理效应 (dynamic treatment effects) 图	仅仅基于你的想法去选择模型； 仅仅报告回归系数； 没有任何可视化或检验的结果
检验	进行可视化 (visual) 和统计检验两种方式去保证识别和模型假设等有效性	

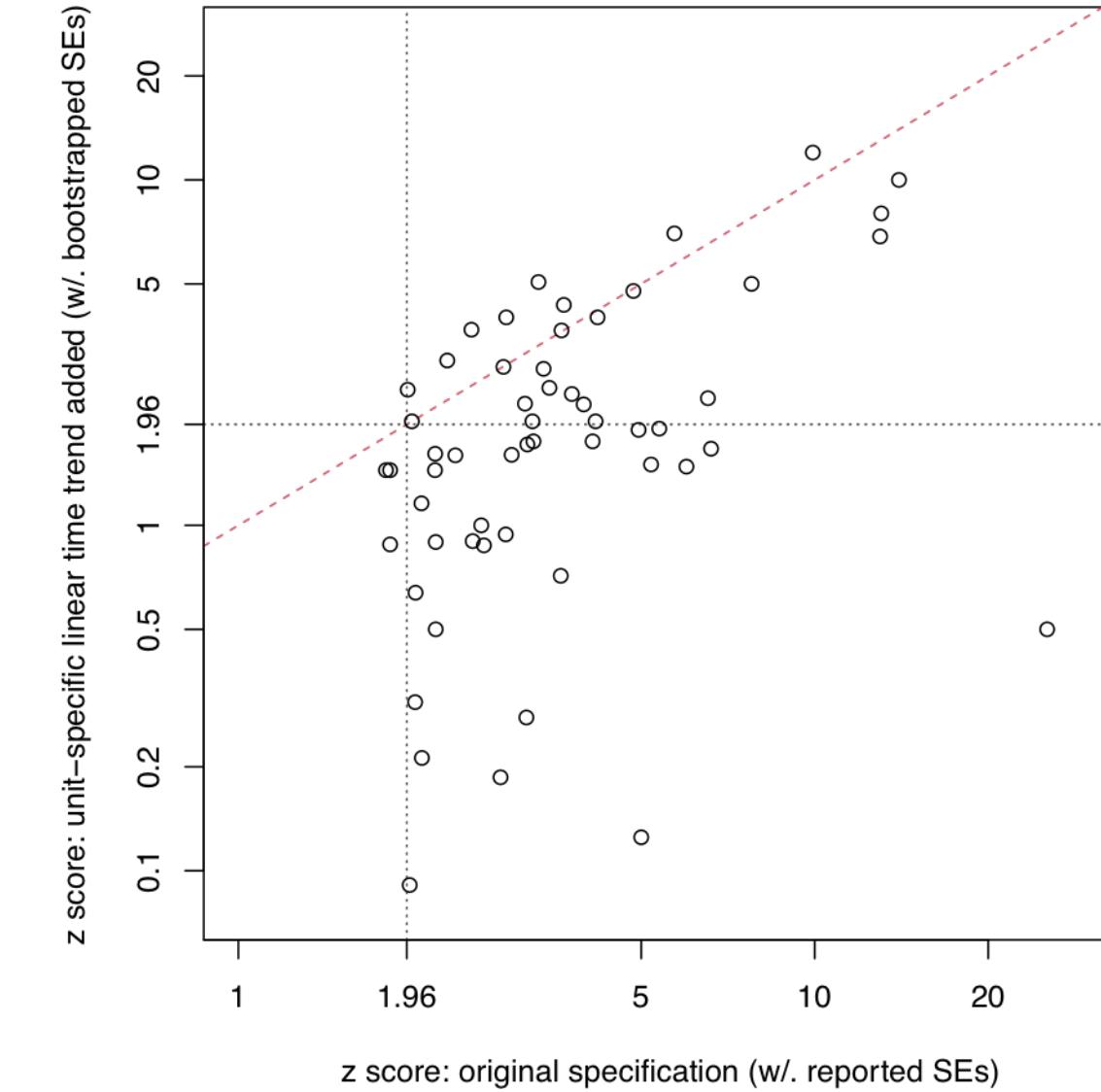
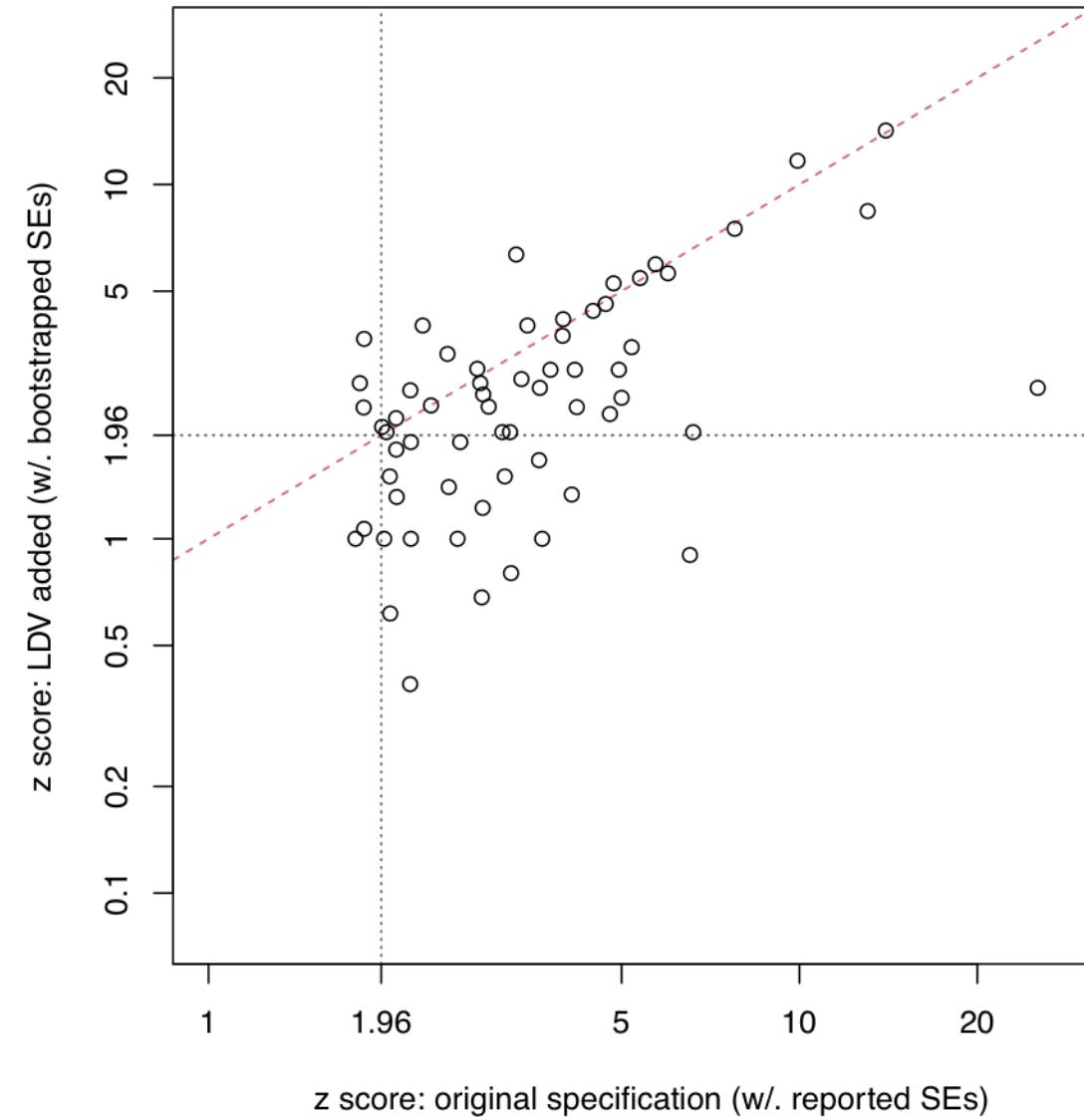
该做和不该做的事

	要	不做
聚类 (clustering) 层面	标准误聚类到干预 (treatment assignment) 或更高的层面，去解释潜在的空间溢出 (spatial spillover)	标准误聚类到比干预更低的层面
重复抽样 (Bootstrapping)	当聚类数很小 (例如小于50) 时，使用聚类自抽样 (cluster-bootstrap)	当聚类数很小时，使用渐进标准误 (asymptotic SEs)
探索异质性处理效应 (HTE)	用灵活的估计策略来探索在理论上很重要的、干预前的控制变量的异质性处理效应 (HTE)，以及可视化你的研究结果	通过死板的、带有交互项的、没有可视化手段的回归模型来探索异质性处理效应 (HTE)

第四讲小结

- 介绍了四种对异质性稳健的估计量，比较了他们的异同
- 我们复制和重新分析了30篇发表在政治科学顶尖期刊的、使用面板数据去建立因果关系的论文
- 使用多种异质性稳健估计量和检验方法，我们发现大约1/3的研究结果得到了比较强的实证支撑
- 在剩下的那些论文里，违反平行趋势（Parallel trends violations）是首要考虑的问题；异质性处理效应（HTE）是重要的第二考虑的问题（参考Baker et al. 2021）
- 低估不确定性仍然是一个重要的、被忽视的问题
- 下一讲，如何处理平行趋势被违反的情况？

连续型处理变量更麻烦



参考文献

- Haber, Noah A., Emma Clarke-Deelder, Joshua A. Salomon, Avi Feller, and Elizabeth A. Stuart. 2021. “COVID-19 Policy Impact Evaluation: A Guide to Common Design Issues.” *American Journal of Epidemiology*, June. <https://doi.org/10.1093/aje/kwab185>.
- Athey, Susan, and Guido W. Imbens. 2018. “Design-Based Analysis in Difference-in-Differences Settings with Staggered Adoption.” Working Paper Series. National Bureau of Economic Research. <https://doi.org/10.3386/w24963>.
- Imai, Kosuke and In Song Kim (2019). “When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?” *American Journal of Political Science*, Vol. 62, Iss. 2, April 2019, pp. 467 – 490.
- Sun, Liyang, and Sarah Abraham. 2018. “Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects.” arXiv [econ.EM]. arXiv. <http://arxiv.org/abs/1804.05785>.
- Goodman-Bacon, Andrew. 2021. “Difference-in-Differences with Variation in Treatment Timing.” *Journal of Econometrics*, June.
- Chaisemartin, Clément de, and Xavier D’Haultfœuille. 2020. “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects.” *The American Economic Review* 110 (9): 2964 – 96.
- Chiu, Albert, Xingchen Lan, Ziyi Liu,, and Yiqing Xu. 2022. “What To Do (and Not to Do) with Causal Panel Analysis: Parallel Trends, Heterogeneous Treatment Effects, and Diagnostics.” Working Paper, Stanford University.