

双重差分设计

更新: October 28, 2023

1 经典双重差分

假设 $D_i \in \{0, 1\}$ 表示处置 (treatment), 表示个体 i 是否接受处置 (参加培训), 其中 1 代表接受处置 (个体 i 位于处置组), 0 代表没有接受处置 (个体 i 位于控制组).

假设 Y_i 代表结果变量 (outcome), $Y_{i,t}(0, 0)$ 表示个体 i 在 t 时刻的潜在结果 (potential outcome), 如果 i 在两个时期都保持未处置. 类似地, $Y_{i,t}(0, 1)$ 表示个体 i 在 t 时刻的潜在结果, 如果 i 在第一期末处置而在第二期接受处置.

为了简化符号, 记 $Y_{i,t}(0) = Y_{i,t}(0, 0)$ 以及 $Y_{i,t}(1) = Y_{i,t}(0, 1)$. 由于因果推断的根本问题, 我们实际上只能观测到

$$Y_{i,t} = D_i Y_{i,t}(1) + (1 - D_i) Y_{i,t}(0)$$

换言之, 一旦个体 i 接受处置, 那就无法观测到它未接受处置的反事实结果 (counterfactual outcome), 反之亦然.

在本文剩余部分, 我们均给定个体稳定处置效应假设 (Stable Unit Treatment Value Assumption, SUTVA) 成立, 也即任何个体的潜在结果不随其它个体是否接受处置而变化, 并且每一个体所接受的处置水平是唯一的, 因此处置所导致的潜在结果也是唯一的.

在政策评估领域, 研究者感兴趣的因果效应为

$$\tau_2 = \mathbb{E}[Y_{i,2}(1) - Y_{i,1}(0)|D_i = 1]$$

它衡量了接受处置的个体在接受处置时 ($t = 2$) 的平均因果效应, 称之为**处置组平均处置效应** (Average Treatment Effect on the Treated, ATT). 为了正确识别和估计 τ_2 , 可以使用**双重差分法** (Difference-in-Differences, DiD).

在 SUTVA 成立的情况下, ATT 是良好定义的, 但由于反事实结果的存在, 我们仍然无法识别出它. 换言之, 为了正确识别 ATT, 还需要准备其它假设条件.

假设 1.1 (平行趋势) $\mathbb{E}[Y_{i,2}(0) - Y_{i,1}(0)|D_i = 1] = \mathbb{E}[Y_{i,2}(0) - Y_{i,1}(0)|D_i = 0]$.

平行趋势假设意味着, 如果没有政策冲击, 那么处置组和控制组中的平均结果变量随时间变化的趋势是平行的. 换言之, 其他因素造成的影响处置的**选择偏差** (selection bias) 在 $t = 1$ 和 $t = 2$ 是相同的.

假设 1.2 (无预期效应) 对于所有 $D_i = 1$ 的个体 i 都有 $Y_{i,1}(0) = Y_{i,1}(1)$.

无预期效应意味着, 在政策冲击到来之前, 没有任何处置组中的个体会预期到它的到来而作出反应. 换言之, 政策冲击在实施前对结果变量没有任何因果效应.

如果平行趋势和无预期效应成立, 那么在 $t = 2$ 时刻的处置组平均处置效应 τ_2 可以被成功识别, 经过简单的数学运算可得

$$\tau_2 = \mathbb{E}[Y_{i,2}(1) - Y_{i,1}(1)|D_i = 1] - \mathbb{E}[Y_{i,2}(0) - Y_{i,1}(0)|D_i = 0]$$

上式第 1 项表示由于政策冲击和其他因素造成的处置组结果变量的平均差异, 第 2 项为其他因素造成的控制组结果变量的平均差异, 因此该方法称为双重差分.

现在已经正确识别出了 τ_2 , 一个估计它的自然想法是用样本均值替代期望, 得到

$$\hat{\tau}_2 = (\bar{Y}_{t=2,d=1} - \bar{Y}_{t=1,d=1}) - (\bar{Y}_{t=2,d=0} - \bar{Y}_{t=1,d=0})$$

其中 $\bar{Y}_{t=t',D=d}$ 是组别 d 中的 Y 在 t' 时刻的样本均值. 另一方面, 还可以使用双向固定效应 (Two-Way Fixed Effects, TWFE) 回归

$$Y_{i,t} = \alpha_i + \lambda_t + (1[t = 2] \cdot D_i)\beta + \epsilon_{i,t}$$

这里的 α_i 为个体固定效应, λ_t 为时间固定效应, $\epsilon_{i,t}$ 为随机扰动项. 可以证明 β 的 OLS 估计量 $\hat{\beta}$ 等于 $\hat{\tau}_2$. 最后, 我们可以在 TWFE 中加入不受处置影响的协变量 X 得到

$$Y_{i,t} = \alpha_i + \lambda_t + (1[t = 2] \cdot D_i)\beta + X'_{i,t}\gamma + \epsilon_{i,t}$$

这样可以控制处置组与控制组中随时间变化的可观测特征差异可能导致的估计误差.

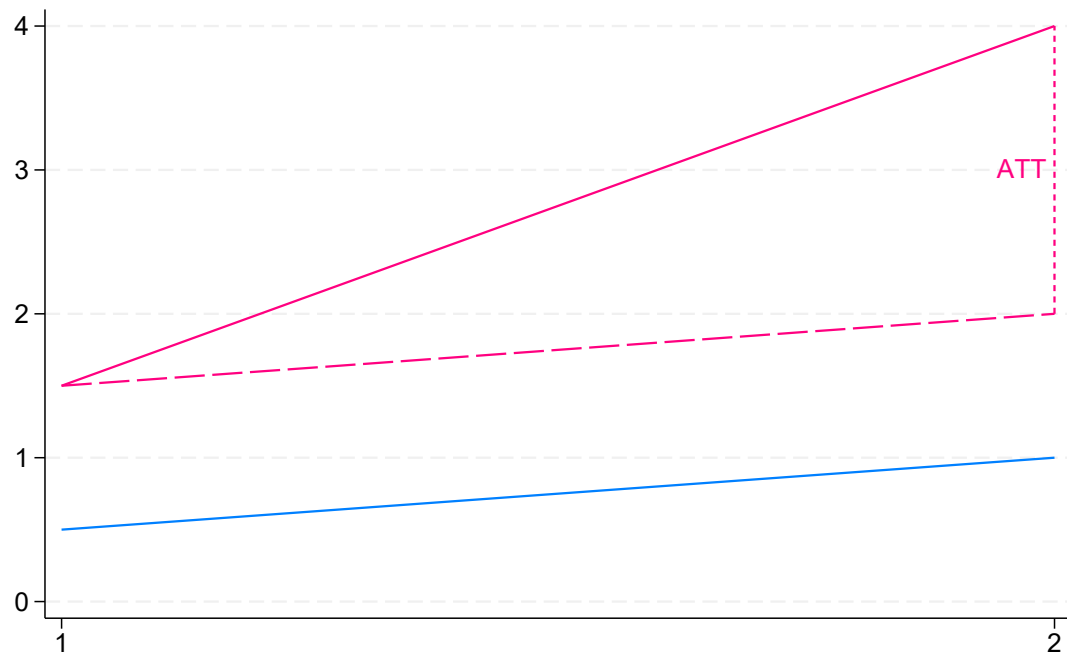


图 1: 双重差分示意图

2 交错时间处置

2.1 引子: 异质性处置效应下的 TWFE 估计量

假设观测数据可以被分割为 G 个组群和 T 个时期, 对于任意 $(g, t) \in \{1, \dots, G\} \times \{1, \dots, T\}$, 令 $N_{g,t}$ 表示当时点为 t 时的组群 g 的观测数量, $N = \sum_{g,t} N_{g,t}$ 为总的观测数量. 特别地, 如果对于一切 (g, t) 都有 $N_{g,t} = 1$, 那么一个组群仅代表一个个体.

另一方面, 对于任意的 $(i, g, t) \in \{1, \dots, N_{g,t}\} \times \{1, \dots, G\} \times \{1, \dots, T\}$, 设 $D_{i,g,t}$ 和 $(Y_{i,g,t}(0), Y_{i,g,t}(1))$ 分别表示组群 g 中的个体 i 在时点 t 的处置阶段和潜在结果, 此时结果变量为 $Y_{i,g,t} = Y_{i,g,t}(D_{i,g,t})$. 此外, 对于一切 (g, t) , 令

$$D_{g,t} = \frac{1}{N_{g,t}} \sum_{i=1}^{N_{g,t}} D_{i,g,t}, \quad Y_{g,t}(0) = \frac{1}{N_{g,t}} \sum_{i=1}^{N_{g,t}} Y_{i,g,t}(0)$$
$$Y_{g,t}(1) = \frac{1}{N_{g,t}} \sum_{i=1}^{N_{g,t}} Y_{i,g,t}(1), \quad Y_{g,t} = \frac{1}{N_{g,t}} \sum_{i=1}^{N_{g,t}} Y_{i,g,t}$$

这里的 $D_{g,t}$ 组群 g 在时点 t 的平均处置, $Y_{g,t}(0)$ 和 $Y_{g,t}(1)$ 含义类似.

在本小节的剩余部分, 总是给定以下假设.

假设 2.1 (平衡组群) $\forall (g, t) \times \{1, \dots, G\} \times \{1, \dots, T\}, N_{g,t} > 0$.

假设2.1表明随着时间流逝, 没有组群会突然出现或消失, 这一条件通常是满足的.

假设 2.2 (清晰设计) $\forall (g, t) \times \{1, \dots, G\} \times \{1, \dots, T\}$, 以及 $i \in \{1, \dots, N_{g,t}\}$, $D_{i,g,t} = D_{g,t}$.

假设2.2表明个体的处置状态在每个胞元 (g, t) 内部是不会改变的.

假设 2.3 (独立组群) 向量 $(Y_{g,t}(0), Y_{g,t}(1), D_{g,t})_{1 \leq t \leq T}$ 两两独立.

假设2.3允许同一组群的处置与潜在结果随着时间流逝产生相关性, 但要求不同组群的处置和潜在结果是独立的.

假设 2.4 (强外生性) $\forall (g, t) \in \{1, \dots, G\} \times \{2, \dots, T\}$, $\mathbb{E}[Y_{g,t}(0) - Y_{g,t-1}(0) | D_{g,1}, \dots, D_{g,T}] = \mathbb{E}[Y_{g,t}(0) - Y_{g,t-1}(0)]$.

假设2.4要求一个组的 $Y_{g,t}(0)$ 均值独立于该组的处置序列以排除发生 Ashenfelter 沉降的可能性.

假设 2.5 (共同趋势) 对于一切 $t \geq 2$, $\mathbb{E}[Y_{g,t}(0) - Y_{g,t-1}(0)]$ 跨组群 g 保持不变.

假设2.5要求没有接受处置的预期结果随着时间流逝在每个组群中具有相同的演变趋势.

现在设 $\hat{\beta}_{\text{FE}}$ 表示 $Y_{i,g,t}$ 对组群固定效应、时间固定效应和 $D_{g,t}$ 进行线性回归的 OLS 估计量, 并设 $\beta = \mathbb{E}[\hat{\beta}_{\text{FE}}]$. 再令 $N_1 = \sum_{i,g,t} D_{i,g,t}$ 为所有接受处置的个体数量, 令

$$\Delta^{\text{TR}} = \frac{1}{N_1} \sum_{(i,g,t): D_{g,t}=1} [Y_{i,g,t}(1) - Y_{i,g,t}(0)]$$

表示跨全体收处置个体的平均处置效应, 并且 $\delta^{\text{TR}} = \mathbb{E}[\Delta^{\text{TR}}]$ 表示该参数的期望, 也即 ATT. 对于一切 $(g, t) \in \{1, \dots, G\} \times \{1, \dots, T\}$, 令

$$\Delta_{g,t} = \frac{1}{N_{g,t}} \sum_{i=1}^{N_{g,t}} [Y_{i,g,t}(1) - Y_{i,g,t}(0)]$$

为胞元 (g, t) 的平均处置效应 (Average Treatment Effect, ATE), 经过一些简单运算可知

$$\delta^{\text{TR}} = \mathbb{E} \left[\sum_{(g,t): D_{g,t}=1} \frac{N_{g,t}}{N_1} \Delta_{g,t} \right]$$

de Chaisemartin and D'Haultfœuille (2020) 证明了, 在假设2.1-2.5成立的条件下有

$$\beta_{\text{FE}} = \mathbb{E} \left[\sum_{(g,t): D_{g,t}=1} \frac{N_{g,t}}{N_1} w_{g,t} \Delta_{g,t} \right] \quad (\star)$$

其中

$$w_{g,t} = \frac{\eta_{g,t}}{\sum_{(g,t): D_{g,t}=1} \frac{N_{g,t}}{N_1} \eta_{n,t}}$$

并且

$$D_{g,t} = \alpha_i + \lambda_t + \eta_{i,t}$$

是出自 $D_{i,t}$ 对组群固定效应和时间固定效应回归的残差¹. 根据式 (★) 可知, 一般情况下 $\beta_{FE} \neq \delta^{TR}$, 也即 $\hat{\beta}_{FE}$ 是 ATT 的有偏估计.

为了看到这一点, 假设组群 1 在时点 1 和 2 未接受处置而在时点 3 接受处置, 组群 2 在时点 1 未接受处置而在时点 2 和时点 3 接受处置, 并且 $N_{g,t}/N_1$ 跨 g 保持不变, 当数据是平衡面板时, 这一要求是满足的. 此时

$$\eta_{g,t} = D_{i,t} - D_{g,\cdot} - D_{\cdot,t} + D_{\cdot,\cdot},$$

于是

$$\eta_{1,3} = 1 - 1/3 - 1 + 1/2 = 1/6$$

$$\eta_{2,3} = 1 - 2/3 - 1/2 + 1/2 = 1/3$$

$$\eta_{3,3} = 1 - 2/3 - 1 + 1/2 = -1/6$$

¹这里要求回归元不是角点的 (collinear).

根据式 (★) 可知

$$\beta_{\text{FE}} = 1/2\mathbb{E}[\Delta_{1,3}] + \mathbb{E}[\Delta_{2,2}] - 1/2\mathbb{E}[\Delta_{2,3}]$$

组群 2 在时点 3 出现了负权重, 如果设 $\mathbb{E}[\Delta_{1,3}] = \mathbb{E}[\Delta_{2,2}] = 1$, $\mathbb{E}[\Delta_{2,3}] = 4$, 那么 $\beta_{\text{FE}} = -1/2 < 0$. 换言之, 即使每个 $\mathbb{E}[\Delta_{g,t}] > 0$, 但只要它们跨 g 或 t 是异质性的 (heterogeneous), 那么 β_{FE} 也可能严格为负, 这将导致对处置效应估计的严重错误!

2.2 Goodman-Bacon 分解

现在回到正题, 在经典 DiD 的框架下, 政策在同一时刻发生, 且只包含了处置组和控制组, 然而政策可能具有滞后性, 因此经典 DiD 不再适用. 有鉴于此, 我们可以将 2×2 的经典 DiD 推广至多维, 也即存在多个组群在不同时点受到政策冲击.

假设时间跨度为 $T + 1$, 每个时点由 $t = 0, \dots, T$ 标记, 个体 i 可以在任意 $t \geq 0$ 的时点受到二元处置, 并且处置是吸收态 (absorbing state), 也即一旦个体接受处置, 那么它在剩余的时间上保持被处置的状态. 我们使用 $D_{i,t}$ 作为个体 i 在时点 t 是否接受处置的二元指示符, $G_i = \min\{t : D_{i,t} = 1\}$ 为个体 i 首次接受处置的时点, 同在 G_i 接受处置的个体 i 构成一个组群, $G_i = \infty$ 表示个体 i 从未接受处置. 由于处置是吸收的, 因此对于一切 $t \geq G_i$ 都有 $D_{i,t} = 1$.

我们需要推广潜在结果的表示方法. 设 $\mathbf{0}_s$ 和 $\mathbf{1}_s$ 分别为 s 维的 0 向量和 1 向量, 如果个体 i 在时点 g 首次接受处置, 那么它在 t 时刻的潜在结果为 $Y_{i,t}(\mathbf{0}_{g-1}, \mathbf{1}_{T-g+1})$, 而用 $Y_{i,t}(\mathbf{0}_T)$ 表示从未接受处置的潜在结果, 分别简写为 $Y_{i,t}(g)$ 以及 $Y_{i,t}(\infty)$, 并且

$$Y_{i,t}(g) = Y_{i,t}(\infty) + \sum_{0 \leq g \leq T} [Y_{i,t}(g) - Y_{i,t}(\infty)] \cdot 1[G_i = g]$$

同样地, 为了使用 DiD, 我们需要新的平行趋势和无参与效应的假设, 它们是前文所述假设的推广.

假设 2.6 (平行趋势) 对于一切 $t \neq t', g \neq g'$ 都有

$$\mathbb{E}[Y_{i,t}(\infty) - Y_{i,t'}(\infty) | G_i = g] = \mathbb{E}[Y_{i,t}(\infty) - Y_{i,t'}(\infty) | G_i = g']$$

上式表示, 不管个体 i 在何时接受处置, 如果它们从未接受处置, 那么它在任意时点上的潜在结果将具有平行趋势.

假设 2.7 (无参与效应) 对于一切 i 和 $t < g$ 都有 $Y_{i,t}(g) = Y_{i,t}(\infty)$.

在以上假设成立的情况下, 仍然考虑使用静态 TWFE 模型

$$Y_{i,t} = \alpha_i + \lambda_t + D_{i,t} \beta_{\text{post}} + \epsilon_{i,t}$$

Goodman-Bacon (2021) 证明了以下概率极限

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}_{\text{post}} = \text{VWATT} + \text{VWCT} + \Delta \text{ATT}$$

其中 VWATT 是 TWFE-DiD 估计量所能估计的我们感兴趣的因果效应, VWCT 在平行趋势假设成立的情况下为 0, ΔATT 等于各组群在处置前后的处置结果变化的加权和.

特别地, 如果处置效应是同质的 (也即 ATT 跨组群和时点保持不变), 那么 $\Delta \text{ATT} = 0$, 此时 $\hat{\beta}_{\text{post}}$ 正确识别了因果关系, 使用 OLS 估计静态 TWFE 模型是合理的.

根据上一节的论述, 如果在不同时点受到的处置效应存在异质性, 那么 TWFE 估计是有偏的 (de Chaisemartin and D'Haultfœuille, 2020), 甚至会出现完全相反的因果效应. 事实上, 根据 FWL 定理可知 $\hat{\beta}_{\text{post}}$ 的 OLS 估计量为

$$\hat{\beta}_{\text{post}} = \frac{\sum_{i,t} (D_{i,t} - \hat{D}_{i,t}) Y_{i,t}}{\sum_{i,t} (D_{i,t} - \hat{D}_{i,t})^2}$$

其中 $\hat{D}_{i,t}$ 是出自回归 $D_{i,t} = \alpha_i + \lambda_t + u_{i,t}$ 的 $D_{i,t}$ 的拟合值. 由于上式分母始终为正, 以及 $Y_{i,t} = Y_{i,t}(\infty) + \tau_{i,t}(g)$, 因此如果 $D_{i,t} = 1$ 且 $D_{i,t} - \hat{D}_{i,t} < 0$, 那么 $\tau_{i,t}(g)$ 在 $\hat{\beta}_{\text{post}}$ 中取得负权重, 即使个体 i 在 t 时刻接受处置.

进一步, 经过某些代数运算可得

$$\hat{D}_{i,t} = \bar{D}_i + \bar{D}_t - \bar{D}$$

其中 $\bar{D}_i = T^{-1} \sum_t D_{i,t}$ 表示个体 i 的处置 D 在时间上的平均, $\bar{D}_t = N^{-1} \sum_i D_{i,t}$ 表示时点 t 的处置 D 在截面上的平均, $\bar{D} = (NT)^{-1} \sum_{i,t} D_{i,t}$ 表示处置 D 在整个截面和时间上的平均.

如果个体在几乎所有时点都是处置状态, 那么 $\bar{D}_i \approx 1$, 如果某时点几乎所有个体都被处置, 那么 $\bar{D}_t \approx 1$, 在二者同时成立的情况下有 $\hat{D}_{i,t} \approx 2 - \bar{D}$, 如果在某段时间内有部分未接受处置的个体存在, 那么 $\hat{D}_{i,t} > 1$. 由此可见, 早期接受处置的个体在后期更容易产生负权重.

Goodman-Bacon (2021) 进一步指出, TWFE-DiD 估计量是所有可能的 2×2 DiD 估计量的加权平均, 各个小的 DiD 估计量的处置组和控制组是混乱的, 有的将从未接受处置的组群作为控制组, 有的将尚未处置 (not-yet-treated) 的组群作为控制组, 还有的将先接受处置的组群作为控制组而将后接受处置的组作为处置组. 当存在随时间变化的处置效应异质性时, 某些小的 DiD 估计量就会产生负权重. 特别地, 使用较早接受处置的组作为控制组被称为**禁止比较** (forbidden comparison).

为了解决异质性处置效应下的 TWFE-DiD 估计偏误, 一方面可以在无异质性处理

路径的情况下使用**事件研究法** (event study method), 另一方面则是使用更具有普适性的稳健的 DiD 估计量, 下面将分别介绍.

2.3 事件研究法

既然情况已经由 2×2 DiD 扩展到了更一般的多个处置时点, 为了利用更多信息, 一个自然考虑是使用动态模型.

假设数据结构为平衡面板数据, 包含个体 $i \in \{1, \dots, N\}$, 每个个体可观测时期为 $t \in \{0, \dots, T\}$, 个体 i 接受处置的时期为 G_i , 如果从未接受处置则 $G_i = \infty$, 个体 i 关于 G_i 的相对处置时间变量为 $D_{i,t} = 1[t - G_i = l]$, 其中 l 表示相对处置时点, 此外还定义集合 e , 使得

$$1[t - G_i \in e] = \sum_{l \in e} 1[t - G_i = l] = \sum_{l \in e} D_{i,t}^l$$

以及关于相对处置时点 $l \in [-T, T]$ 的互不相交的 e 构成的集族 \mathcal{E} .

实践中通常考虑如下动态 TWFE 模型

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{l=-K}^{-2} \mu_l D_{i,t}^l + \sum_{l=0}^L \mu_l D_{i,t}^l + \epsilon_{i,t}$$

其中相对时间变量 $D_{i,t}$ 被分为了两组, 一组代表处置前 ($l \leq -2$), 一组代表处置后 ($l \geq 0$), 各个 μ_g 体现了动态效应 (dynamic effects). 此外, 为了避免完全多重共线性, 选取 $l = -1$ 作为基期. 这里的 $K = G^{\max} - 1$, $L = T - G^{\min}$, 并且 G^{\max} 和 G^{\min} 分别表示最晚和最早接受处置的时间.

Sun and Abraham (2021) 证明了, 再无需任何假设的情况下, μ_g 可以被分解为

$$\begin{aligned}\mu_g = & \sum_{l \in e} \sum_{g \neq \infty} w_{g,l}^e (\mathbb{E}[Y_{i,g+l} - Y_{i,0}(\infty) | G_i = g] - \mathbb{E}[Y_{i,g+l} - Y_{i,0}(\infty)]) \\ & + \sum_{e' \neq e, e' \in \mathcal{E}} \sum_{l \in e'} \sum_{g \neq \infty} w_{g,l}^{e'} (\mathbb{E}[Y_{i,g+l} - Y_{i,0}(\infty) | G_i = g] - \mathbb{E}[Y_{i,g+l} - Y_{i,0}(\infty)]) \\ & + \sum_{l \in e^{\text{excl}}} \sum_{g \neq \infty} w_{g,l}^e (\mathbb{E}[Y_{i,g+l} - Y_{i,0}(\infty) | G_i = g] - \mathbb{E}[Y_{i,g+l} - Y_{i,0}(\infty)]) \\ & + \sum_t w_{\infty,t}^e (\mathbb{E}[Y_{i,g+l} - Y_{i,0}(\infty) | G_i = \infty] - \mathbb{E}[Y_{i,g+l} - Y_{i,0}(\infty)])\end{aligned}$$

其中 $w_{g,l}^e (g \neq \infty)$ 是出自以下辅助回归的回归系数

$$D_{i,t}^l \cdot 1[G_i = g] = \alpha_i + \lambda_t + \sum_{e \in \mathcal{E}} w_{g,l}^e 1[t - G_i \in e] + u_{i,t}$$

并且 $e^{\text{excl}} = \{-T, \dots, -K-1, -1, L+1, \dots, T\}$, 以及

$$\begin{aligned} \sum_{l \in e} \sum_g w_{g,l}^e &= 1, \quad \forall l \in e \\ \sum_{l \in e'} \sum_g w_{g,l}^e &= 0, \quad \forall l \in e', e' \neq e, e' \in \mathcal{E} \\ \sum_{l \in e^{\text{excl}}} \sum_g w_{g,l}^e &= -1, \quad \forall l \in e^{\text{excl}} \end{aligned} \quad (*)$$

可以看出, 事件研究法的回归系数是一系列基本的 2×2 DiD 估计量的加权平均. 另一方面, 定义

$$\text{ATT}_{g,l} = \mathbb{E}[Y_{i,g+l} - Y_{i,g+l}(\infty) | G_i = g]$$

表示在 $G_i = g$ 时首次接受处置的那批个体经过 l 期后的平均处置效应. Sun and Abraham (2021) 还证明了, 在合适的平行趋势假设下有

$$\mu_g = \sum_{l \in e} \sum_g w_{g,l}^e \text{ATT}_{g,l} + \sum_{e \neq e', e' \in \mathcal{E}} \sum_{l' \in e'} \sum_g w_{g,l'}^e \text{ATT}_{g,l'} + \sum_{l' \in e^{\text{excl}}} \sum_g w_{g,l'}^e \text{ATT}_{g,l'}$$

如果处理效应的异质性仅随时间流动产生动态变化且路径对每个 g 相同 (称之为同质性处置效应路径), 也即 $\text{ATT}_{g,l} = \text{ATT}_l \neq \text{ATT}$, 那么

$$\mu_g = \sum_{l \in e} w_l^e \text{ATT}_l + \sum_{e \neq e', e' \in \mathcal{E}} \sum_{l' \in e'} w_{l'}^e \text{ATT}_{l'} + \sum_{l' \in e^{\text{excl}}} w_{l'}^e \text{ATT}_{l'}$$

其中 $w_l^e = \sum_g w_{g,l}^e$. 特别地, 如果模型是完全动态的 (fully dynamic), 也即 $e^{\text{excl}} = \{-1\}$, 那么每个 e 只包含一个元素, 于是根据式 (*) 可知对于一切 $l \neq l'$ 都有 $w_{l'}^l = 0$, 因此

$$\mu_g = \text{ATT}_l + w_{-1}^l \text{ATT}_{-1}$$

此时若追加无预期效应假设, 立即可得 $\mu_g = \text{ATT}_l$.

综上所述, 如果是完全动态下的事件研究设计, 在平行趋势假设、无预期效应假设、同质性处置效应路径的假设下, 回归系数 μ_g 可以正确识别平均处置效应. 相反, 单纯使用 DiD 则无法在这些假设下得到正确的估计.

2.4 Callaway-Sant'Anna 估计量

上述提到的事件研究设计可以解决同质性处置效应路径的问题, 然而对于 $\text{ATT}_{g,l} \neq \text{ATT}_l \neq \text{ATT}$ 这种一般的异质性处置效应问题, 事件研究仍然会得到有偏估计, 此时需要用到异质性处置效应稳健的估计量.

Callaway and Sant'Anna (2021)²考虑了

$$\theta(g, t) = \mathbb{E}[Y_t(g) - Y_t(\infty) | G_g = 1]$$

这里为了简单起见省略了下标 i , G_g 是个体 i 在时点 g 是否接受处置时的指示符, 上式表示在时点 g 首次接受处置的那批个体在时点 t 的 ATT.

由于反事实结果的存在, $\theta(g, t)$ 是无法直接估计的, 但是可以由回归调整 (Regression Adjustment, RA), 逆概率加权 (Inverse Probability Weighting, IPW) 和双重稳健 (Double Robust, DR) 方法来估计.

为了通过上述方法得到识别 ATT, 需要先定义控制组, 而它的选择有两种: 其一是从未处置 $C^{\text{NEV}} = G_\infty$, 其二是尚未处置 $C^{\text{NY}} = (1 - G_g)(1 - D_t)$, 以下统一记作 $C_{g,t}^*$. 通常而言, 如果有相当部分的个体从未接受处置, 并且他们与最终接受处置的个体相似时, 研究者会倾向于选择 C^{NEV} .

²其它的替代估计量包括 Sun and Abraham (2021) 提出的 SA 估计量以及堆叠回归估计量 (stacked regression estimator), 它们的比较可见 Baker *et al.* (2022).

考虑加入用以控制可观测特征的协变量 X , 下面给出 **RA**, **IPW** 和 **DR** 的待估计量

$$\begin{aligned}\theta_{\text{RA}}(g, t) &= \mathbb{E} \left[\frac{G_g}{\mathbb{E}[G_g]} \{Y_t - Y_{g-1} - m_{g,t}(X)\} \right] \\ \theta_{\text{IPW}}(g, t) &= \mathbb{E} \left[\left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\frac{p_{g,t}(X)C_{g,t}^*}{1-p_{g,t}(X)}}{\mathbb{E} \left[\frac{p_{g,t}(X)C_{g,t}^*}{1-p_{g,t}(X)} \right]} \right) (Y_t - Y_{g-1}) \right] \\ \theta_{\text{DR}}(g, t) &= \mathbb{E} \left[\left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\frac{p_{g,t}(X)C_{g,t}^*}{1-p_{g,t}(X)}}{\mathbb{E} \left[\frac{p_{g,t}(X)C_{g,t}^*}{1-p_{g,t}(X)} \right]} \right) \{Y_t - Y_{g-1} - m_{g,t}(X)\} \right]\end{aligned}$$

其中 $m_{g,t}(X) = \mathbb{E}[Y_t - Y_{g-1} | X, C_{g,t}^* = 1]$, $p_{g,t}(X) = \mathbb{P}[G_g = 1 | X, G_g + C_{g,t}^* = 1]$. 在温和的正则条件下, Callaway and Sant'Anna (2021) 证明了

$$\theta(g, t) = \theta_{\text{RA}}(g, t) = \theta_{\text{IPW}}(g, t) = \theta_{\text{DR}}(g, t)$$

由于 **RA**, **IPW** 和 **DR** 只依赖于 $(Y, X, G_g, C_{g,t}^*)$, 因此可以通过它们得到 $\theta(g, t)$ 的估计量.

然而, 在长时间跨度和多处理时点的情况下, 报告每一个 $\theta(g, t)$ 是很麻烦的事情, 并且可能不准确, Callaway and Sant'Anna (2021) 提供了将不同 $\theta(g, t)$ 进行加总的机制

$$\theta = \sum_g \sum_{t=2}^T w(g, t) \cdot \theta(g, t)$$

其中 $w(g, t)$ 是某个精心挑选的权重函数, 用以解决特定的实证问题. 特别地, 设 $l = t - g$ 为表示在接受处置后逝去的时间, 则关于 l 的一种加总方式为

$$\theta_{\text{es}}(l) = \sum_g 1[g + l \leq T] \cdot \mathbb{P}[G = g | G + l \leq T] \cdot \theta(g, t)$$

称为**事件研究参数** (event-study parameter), 它给出了在不同处置组群中, 在接受处置后 e 个时期的处置效应的加权平均值.

现在来看对 $\theta(g, t)$ 的估计, 根据 Callaway and Sant'Anna (2021) 的研究, 采取以下步骤是可行的

1. 设 $t_0 = (g - 1)1[t \geq g] + (t - 1)1[t < g]$, 将样本限制在时点 t 和 t_0 上, 也即用于估计 $\theta(g, t)$ 的个体 i 必须在时点 t 和 t_0 上可观测;
2. 使用参数模型估计 $p_{g,t}(X)$ 和 $m_{g,t}(X)$, 具体而言:
 - a. 当 $C_{g,t}^* = 1$ 时用 $Y_t - Y_{t_0}$ 对 X 进行线性回归, 得到预测值 $\hat{m}_{g,t}(X)$,
 - b. 用 G_g 对 X 进行 Logit 回归, 得到预测值 $\hat{p}_{g,t}(X)$;
3. 将估计得到的 $\hat{m}_{g,t}(X)$, $\hat{p}_{g,t}(X)$ 插入 RA, IPW 或 DR 的表达式中, 并用样本均值替代期望算子.

最后, 对于 $\hat{\theta}(g, t)$ 的协方差矩阵的估计可以通过影响函数法 (influence function approach) 计算, 它在数值上等价于 GMM 得到的结果, 但运算速度更快.

值得注意的是, 如果协变量 X_g 可以决定组群接受处置的时点, 那么这种处置时点的内生性可能导致平行趋势假设不成立, 此时使用 CS 估计量仍然是有偏的.

3 违反平行趋势

3.1 理论简介

同样考虑将 DiD 扩展到更一般的事件研究框架, 该框架一方面可以解决同质性处置效应路径下的 TWFE-DiD 估计偏误, 另一方面可以用来检验平行趋势是否成立³.

首先将整个时间窗口设置为 $[-K, L]$, 仍然考虑基本的 TWFE 模型

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{s \neq -1} 1[s = t] \times D_i \times \beta_s + \epsilon_{i,t}$$

这里为了避免完全多重共线性剔除了 $s = -1$ 基期. 如果个体 i 在 $t = 1$ 时接受处置就将 D_i 设置为 1, 否则为 0, 则经典 DiD 估计量

$$\hat{\beta}_s = (\bar{Y}_{s1} - \bar{Y}_{s0}) - (\bar{Y}_{01} - \bar{Y}_{00})$$

数值上等于 TWFE 估计量, 其中 \bar{Y}_{sd} 是 $D_i = 1$ 的那些个体在时点 s 的结果变量的样本均值.

³平行趋势假设在根本上无法检验, 这里检验的实际上是处置前平行趋势是否成立.

现在使用全体 $\hat{\beta}_s$ 构成列向量 $\hat{\beta} = (\hat{\beta}'_{\text{pre}}, \hat{\beta}'_{\text{post}})' \in \mathbb{R}^{K+L}$, 其中 $\hat{\beta}_{\text{pre}} = (\hat{\beta}_{-L}, \dots, \hat{\beta}_{-1})$, 以及 $\hat{\beta}_{\text{post}} = (\hat{\beta}_1, \dots, \hat{\beta}_L)$, 它们分别收集了处置前和处置后的估计量. 在温和的条件下, 当样本容量 $N \rightarrow \infty$ 时有 $\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \Sigma)$.

现在我们假设 β 可以被分解为

$$\beta = \underbrace{\begin{pmatrix} \tau_{\text{pre}} \\ \tau_{\text{post}} \end{pmatrix}}_{=:\tau} + \underbrace{\begin{pmatrix} \delta_{\text{pre}} \\ \delta_{\text{post}} \end{pmatrix}}_{=:\delta}, \quad \tau_{\text{pre}} = 0$$

其中 τ 表示感兴趣的处置效应, δ 表示如果没有政策冲击 (处置), 控制组和处置组的趋势差异. 当无预期效应成立时有 $\tau_{\text{pre}} = 0$, 而当平行趋势成立时有 $\delta_{\text{post}} = 0$, 因此 $\beta_{\text{post}} = \tau_{\text{post}}$. 当然, 检验平行趋势的通常做法则是检验**事前趋势** (pre-trend) 假设⁴ $\mathbb{H}_0 : \beta_{\text{pre}} = 0$.

可以证明, 在经典 DiD 和事件研究的框架下, β 都可以被这样分解. 例如

$$\beta_s = \tau_{\text{ATT},s} + \underbrace{\mathbb{E}[Y_{i,s}(0) - Y_{i,0}(0)|D_i = 1] - \mathbb{E}[Y_{i,s}(0) - Y_{i,0}(0)|D_i = 0]}_{=:\delta_s}$$

⁴值得注意的是, 这一检验的检验势较低, 也即可能无法检测出违反平行趋势的情况, 并且基于处置前趋势的分析可能会扭曲结论, 加剧点估计和推论的偏差 (Roth, 2022). 因此不检验事前趋势, 而考虑使用以下方法进行敏感性分析 (sensitive analysis) 或许是更好的选择.

其中 $\tau_{\text{ATT},s} = \mathbb{E}[Y_{i,s}(1) - Y_{i,s}(0) | D_i = 1]$, 在无参与效应条件, 对任意 $s < 0$ 都有 $\tau_{\text{ATT},s} = 0$, 这就完成了分解.

现在我们关注的目标参数是处置后因果效应的线性组合 $\theta = l' \tau_{\text{post}}$, 其中 l 是已知的 $L \times 1$ 维向量. 通过将 δ 置于可能的趋势差异集 $\Delta \subset \mathbb{R}^{K+L}$ 中, 参数 θ 可以被部分识别 (partial identified). 例如, 将 δ 限制在 $\Delta = \{\delta : \delta_{\text{post}} = 0\}$ 中就意味着平行趋势假设成立.

在违反平行趋势假设 $\delta \in \Delta \neq \{\delta : \delta_{\text{post}} = 0\}$ 的情况下, 参数 θ 被部分识别, 识别集是在 $\delta \in \Delta$ 的限制条件下, 与 β 的给定值相一致的 θ 值构成的集合

$$\mathcal{S}(\beta, \Delta) = \left\{ \theta : \exists \delta \in \Delta, \tau_{\text{post}} \in \mathbb{R}^L \text{ s.t. } l' \tau_{\text{post}} = \theta, \beta = \delta + \begin{pmatrix} 0 \\ \tau_{\text{post}} \end{pmatrix} \right\}$$

可以证明, 如果 Δ 是闭和凸的, 那么识别集 $\mathcal{S}(\beta, \Delta)$ 是 \mathbb{R} 上的区间 $[\theta^{\text{lb}}(\beta, \Delta), \theta^{\text{ub}}(\beta, \Delta)]$, 其中

$$\begin{aligned} \theta^{\text{lb}}(\beta, \Delta) &= l' \beta_{\text{post}} - \left(\max_{\delta} l' \delta_{\text{post}} \text{ s.t. } \delta \in \Delta, \delta_{\text{pre}} = \beta_{\text{pre}} \right) \\ \theta^{\text{ub}}(\beta, \Delta) &= l' \beta_{\text{post}} - \left(\min_{\delta} l' \delta_{\text{post}} \text{ s.t. } \delta \in \Delta, \delta_{\text{pre}} = \beta_{\text{pre}} \right) \end{aligned}$$

Rambachan and Roth (2023) 给出了如何通过选取合适的 Δ , 得到合适的 DiD 估计区间.

第一, 研究者可能愿意假设造成处置后非平行趋势的混杂因素在量级上不会比处理前的混杂因素大太多, 这可以正式地表述为 $\delta \in \Delta^{\text{RM}}(\overline{M})$, $\overline{M} \geq 0$, 其中

$$\Delta^{\text{RM}}(\overline{M}) = \left\{ \delta : \forall t \geq 0, |\delta_{t+1} - \delta_t| \leq \overline{M} \cdot \max_{s < 0} |\delta_{s+1} - \delta_s| \right\}$$

这里的 $\Delta^{\text{RM}}(\overline{M})$ 用 \overline{M} 乘以处置前最大平行趋势违反来约束处置后连续时期之间的最大平行趋势违反. 比方说, 如果造成平行趋势不成立的混杂经济冲击在前后期相似, 那么选取 $\Delta^{\text{RM}}(\overline{M})$ 可能是合理的.

第二, 研究者可能担心长期趋势带来的对平行趋势的干扰 (例如劳动力供给的长期变化), 但这种趋势随着时间推动而平稳演变, 此时的趋势差异集为

$$\Delta^{\text{SD}}(M) = \{ \delta : |(\delta_{t+1} - \delta_t) - (\delta_t - \delta_{t-1})| \leq M, \forall t \}$$

其中参数 $M \geq 0$ 控制了 δ 的斜率在连续周期之间可以变化的幅度, 由此限制了离散二阶导数的范围. 当 $M = 0$ 时, $\Delta^{\text{SD}}(0)$ 要求趋势差异是线性的, 这与实际应用中常见的参数线性假设相对应.

第三, 考虑多面限制 (polyhedral restriction) 的情况以便研究者依赖于特定知识来施加限制, 此时 $\Delta = \{ \delta : A\delta \leq d \}$, 这里的 A 和 d 分别是已知的矩阵和向量.

为了进行统计推断, Rambachan and Roth (2023) 介绍了两种方法, 其一是 Andrews

et al. (2023) 提出的条件混合方法 (简称 ARP), 其二是固定长度置信区间 (Fixed Length Confidence Intervals, FLCI).

根据 Monte Carlo 模拟的结果, 对于一般和多面形式的 Δ , 推荐使用 ARP 条件混合方法, 而对于 $\Delta^{\text{SD}}(M)$ 和其它满足一致性和有限样本近最优性条件的特殊情况, 则推荐使用 FLCI. 这些方法的使用具体见作者提供的[HonestDiD](#)命令包.

3.2 实例: 医疗补助扩张对保险覆盖率的影响

考虑无交错时间处置的经典 DiD 模型

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{s \neq 2013} 1[s = t] \times D_i \times \beta_s + \epsilon_{i,t}$$

其中 $Y_{i,t}$ 代表美国第 i 个州在第 t 年的低收入无孩成人的保险覆盖率, 如果 i 在 2014 年受到医疗补助扩张的影响, 那么 $D_i = 1$, 否则 $D_i = 0$ ⁵.

根据图2可知, 在 $t < 2014$ 的时期, 95% 的置信区间都包含了 0, 事前趋势检验通过, 似乎可以认定平行趋势假设成立. 但正如 Roth (2022) 提到的, 事前趋势检验的检验势较低, 难以真正检验平行趋势是否成立. 为了稳健考虑, 我们转而进行敏感性分析.

⁵为了简单起见, 这里将样本限制在 2015 年及以前, 并且剔除掉在 2015 年前首次受到政策冲击的少数州.

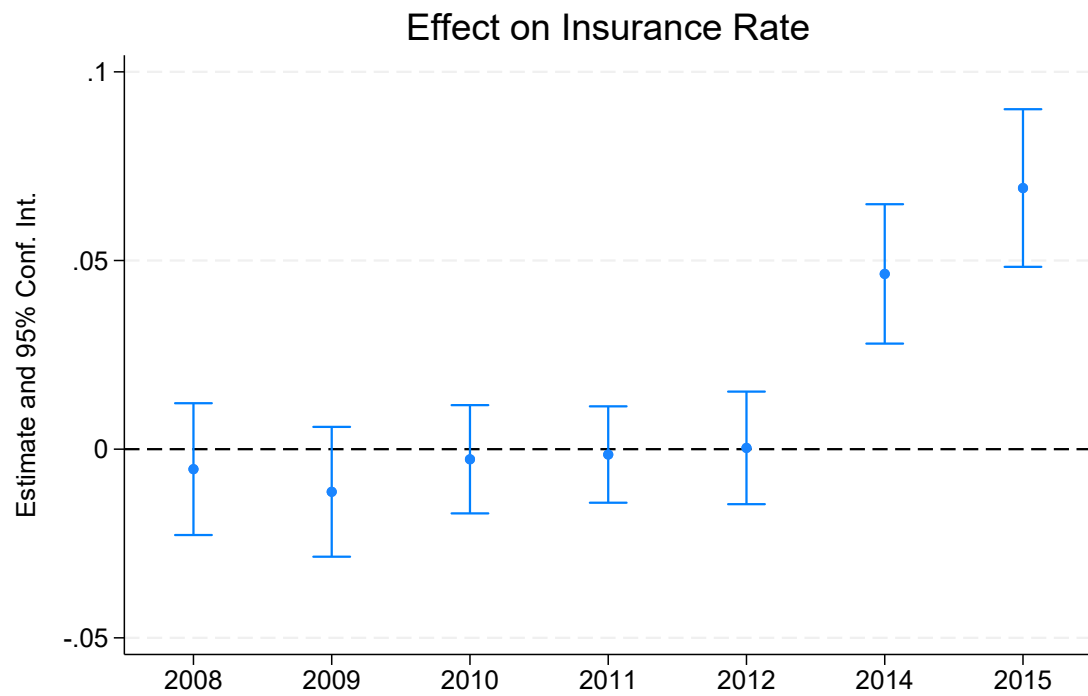


图 2: 事前趋势检验

考虑 $\Delta = \Delta^{\text{RM}}(\overline{M})$, 图3展示了不同的 \overline{M} 的设置下的稳健置信区间, 当 $\overline{M}=0.5$ 时, 违反平行趋势下的稳健置信区间和 OLS 的结果相似, 只是稍宽一些, 但仍然排除了医疗补助扩张的零效应 (null effect); 当 \overline{M} 约等于 2 时, 出现了关于零效应的“击穿值” (breakdown value), 表明如果违反平行趋势的程度达到处置前最大违反程度的两倍, 这一显著结果也是稳健的.

考虑 $\Delta = \Delta^{\text{SD}}(M)$, 图4展示了敏感性分析的另一结果, 可知显著影响的击穿值大约为 $M \approx 0.03$, 意味着除非允许跨连续时期的线性外推偏差超过 0.03, 否则就可以拒绝零效应.

最后, 假设处置是交错时间发生的, 也即医疗补助扩张不仅只发生在 2014 年, 还发生在接下来的年份, 根据前文可知在异质性处置效应下直接使用事件研究是有偏的, 此时可以将HonestDiD与csdid2搭配以使用 CS 估计量, 具体结果如图5所示.

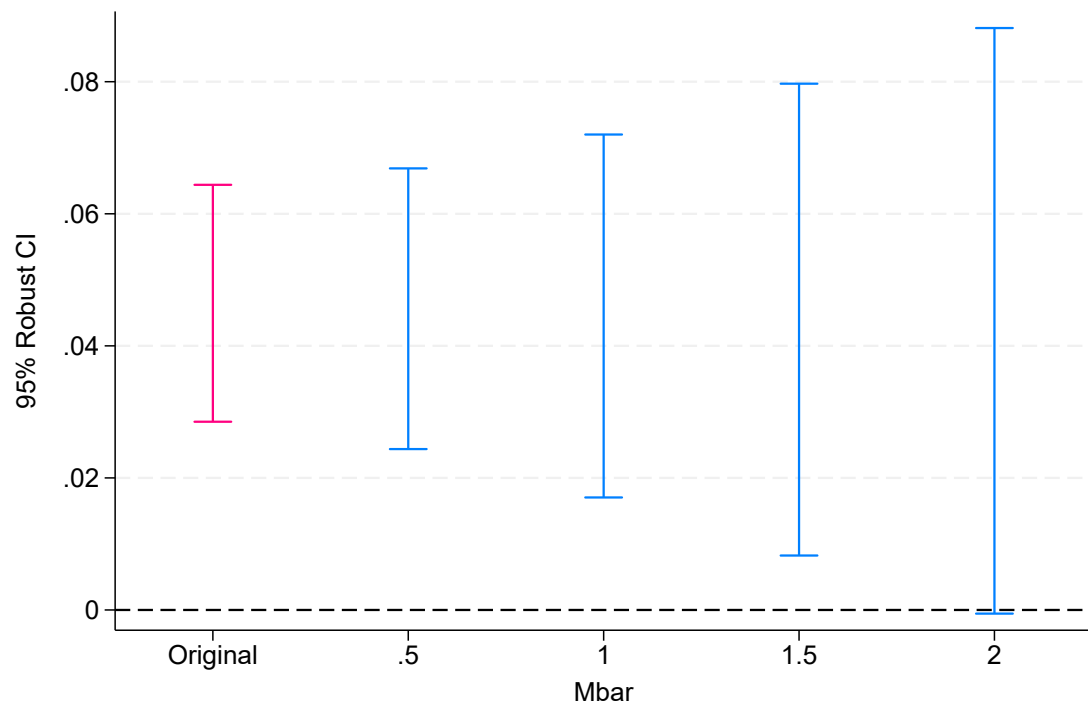


图 3: $\Delta = \Delta^{\text{RM}}(\overline{M})$ 下的敏感性分析

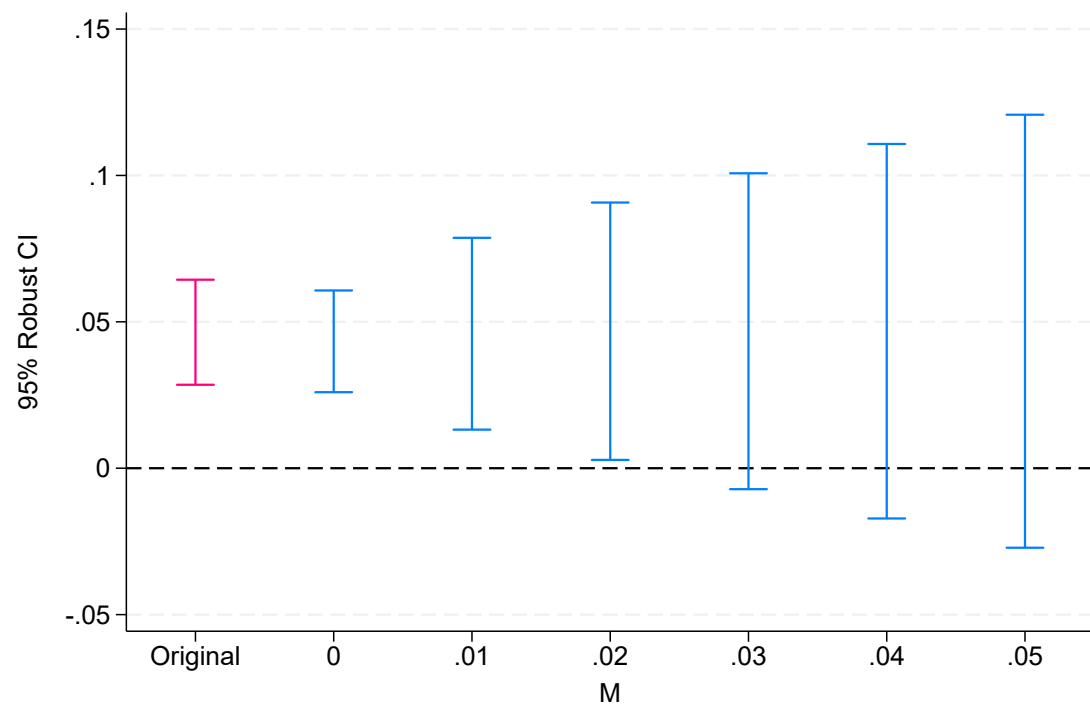


图 4: $\Delta = \Delta^{\text{SD}}(M)$ 下的敏感性分析

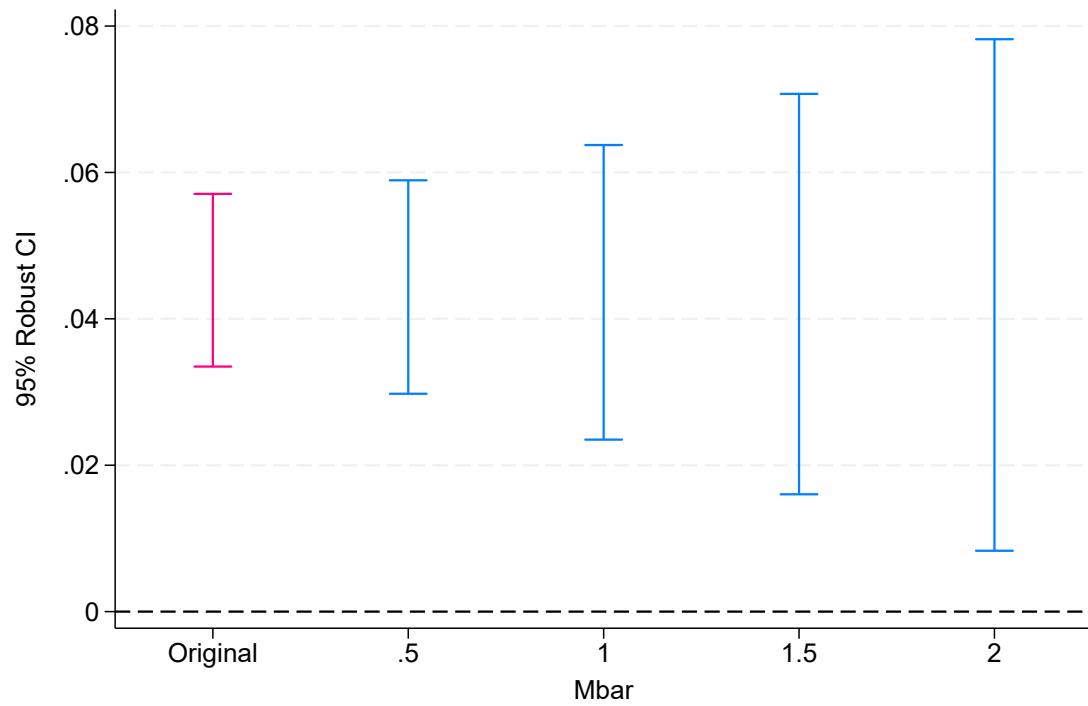


图 5: 交错时间处置下的敏感性分析

参考文献

- [1] ANDREWS I, ROTH J, PAKES A. Inference for Linear Conditional Moment Inequalities[J]. Review of Economic Studies, 2023: Forthcoming.
- [2] BAKER A C, LARCKER D F, WANG C C. How much should we trust staggered difference-in-differences estimates?[J]. Journal of Financial Economics, 2022, 144(2): 370-395.
- [3] CALLAWAY B, SANT'ANNA P H. Difference-in-differences with multiple time periods[J]. Journal of Econometrics, 2021, 225(2): 200-230.
- [4] DE CHAISEMARTIN C, D'HAULTFOEUILLE X. Two-way fixed effects estimators with heterogeneous treatment effects[J]. American Economic Review, 2020, 110(9): 2964-2996.
- [5] GOODMAN-BACON A. Difference-in-differences with variation in treatment timing[J]. Journal of Econometrics, 2021, 225(2): 254-277.
- [6] RAMBACHAN A, ROTH J. A more credible approach to parallel trends[J]. Review of Economic Studies, 2023, 90(5): 2555-2591.
- [7] ROTH J. Pretest with caution: Event-study estimates after testing for parallel trends[J]. American Economic Review: Insights, 2022, 4(3): 305-322.
- [8] ROTH J, SANT'ANNA P H, BILINSKI A, et al. What' s trending in difference-in-differences? A synthesis of the recent econometrics literature[J]. Journal of Econometrics, 2023, 235(2): 2218-2244.
- [9] SUN L, ABRAHAM S. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects[J]. Journal of Econometrics, 2021, 225(2): 175-199.
- [10] 张子尧, 黄炜. 事件研究法的实现、问题和拓展[J]. 数量经济技术经济研究, 2023, 40(9): 71-92.