



## 在模型误设的统一框架下理解双重差分方法的最新发展

林梦芸 徐阳 郭汝飞 易君健

### 摘要:

以 Angrist 为代表的经验主义范式主导了最近三十年的经济学实证研究,但近年来屡受质疑。争议的焦点之一在于双重固定效应回归模型的误用。Angrist (1998) 最初基于条件独立假设,发现“无伤大雅(Mostly Harmless)”的常系数回归模型可以近似估计出异质性处理效应的平均值。Angrist & Pischke (2009, 2014) 则将“无伤大雅”思路运用于平行趋势假设下的双重差分方法,扩展了原本用于实施  $2 \times 2$  经典双重差分的双重固定效应回归模型,提炼出该模型三个维度的推广,促进了该模型的广泛使用。本文发现,Angrist & Pischke 的三个推广都可能带来严重的模型误设。在平行趋势假设下,常系数双重固定效应回归无法估计出异质性处理效应的凸组合,即不同个体或组的处理效应的加权平均(权重非负且和为 1);此时,忽略处理效应的异质性不再是“无伤大雅”的,这是 Angrist 经验主义范式的根本问题。本文以模型误设的统一框架理解双重差分方法的最新发展,提倡研究者基于具体问题的现实背景与经济理论来设定经验分析模型,深化了中国经济实证研究的方法论基础。

**关键词:** 无伤大雅; 模型误设; 负权重; 双重差分方法; 双重固定效应回归; 因果推断

**Keywords:** Mostly harmless; Model mis-specification; Negative weights; Difference-in-differences method; Two-way fixed-effect regression; Causal inference

**JEL Codes:** C13; C23

# 在模型误设的统一框架下理解双重差分方法的最新发展\*

林梦芸 徐阳 郭汝飞 易君健

2024年8月22日

## 摘要

以 Angrist 为代表的经验主义范式主导了最近三十年的经济学实证研究，但近年来屡受质疑。争议的焦点之一在于双重固定效应回归模型的误用。Angrist (1998) 最初基于条件独立假设，发现“无伤大雅 (Mostly Harmless)”的常系数回归模型可以近似估计出异质性处理效应的平均值。Angrist & Pischke (2009, 2014) 则将“无伤大雅”思路运用于平行趋势假设下的双重差分方法，扩展了原本用于实施  $2 \times 2$  经典双重差分的双重固定效应回归模型，提炼出该模型三个维度的推广，促进了该模型的广泛使用。本文发现，Angrist & Pischke 的三个推广都可能带来严重的模型误设。在平行趋势假设下，常系数双重固定效应回归无法估计出异质性处理效应的凸组合，即不同个体或组的处理效应的加权平均（权重非负且和为 1）；此时，忽略处理效应的异质性不再是“无伤大雅”的，这是 Angrist 经验主义范式的根本问题。本文以模型误设的统一框架理解双重差分方法的最新发展，提倡研究者基于具体问题的现实背景与经济理论来设定经验分析模型，深化了中国经济实证研究的方法论基础。

关键词：无伤大雅；模型误设；负权重；双重差分方法；双重固定效应回归；因果推断

**Keywords:** Mostly harmless; Model mis-specification; Negative weights; Difference-in-differences method; Two-way fixed-effect regression; Causal inference

**JEL Codes:** C13; C23

---

\*林梦芸，厦门大学经济学院与王亚南经济研究院；徐阳，湖北经济学院财政与公共管理学院；郭汝飞，武汉大学经济与管理学院健康经济与管理研究中心；易君健，北京大学中国经济研究中心，北京大学国家发展研究院。

## 1 引言

随着 Angrist & Pischke (2009, 2014) 两本计量经济学教材在我国的传播, 中国经济学研究逐渐疏于理论构建, 转向以回归分析为主要研究手段的经验主义范式。Angrist & Pischke (2009) 在其书中开篇指出, 经验主义范式的核心思路是“无伤大雅” (Mostly Harmless), 即通过简单易行的回归模型得到平均处理效应的近似估计。该思路源于 Angrist (1998) 的论证: 在条件独立假设下, 常系数回归模型可以近似估计出异质性处理效应的平均值。<sup>1</sup> 据此, Angrist (1998) 认为在模型设定上可以忽略处理效应的异质性。

Angrist & Pischke (2009) 则在条件独立假设的适用范围之外, 将“无伤大雅”思路运用于平行趋势假设下的双重差分 (Difference-in-differences, 下文简称 DID) 方法。他们通过总结已有文献, 发现双重固定效应回归 (Two-way Fixed-effect Regression, 下文简称 TWFE 回归) 不限于  $2 \times 2$  经典 DID 的情形, 提炼出常系数 TWFE 回归的三类推广。回归方法的便利性, 让 TWFE 回归的三类推广在中国经济学研究中得到广泛应用。

本文发现, TWFE 回归的三类推广都可能带来严重的模型误设。在这三类推广情形下, 仅依赖平行趋势假设, 常系数 TWFE 回归无法估计出异质性处理效应的凸组合, 即不同个体或组的处理效应的加权平均 (权重非负且和为 1); 此时, 常系数回归模型不再是“无伤大雅”的。不假思索地使用回归模型, 而不探明研究问题的现实背景和回归模型设定背后的经济学理论, 将阻碍对中国经济发展规律的深入研究。

实际上, 以 Angrist & Pischke (2009, 2014) 为代表的计量教材, 仅在  $2 \times 2$  经典 DID 的情形下证明 TWFE 回归可以识别出有清晰定义的因果效应。在  $2 \times 2$  的经典数据结构下, DID 方法依赖未受处理潜在结果变量的平行趋势假设: 个体  $i$  未受处理潜在结果 ( $Y_{i,g,t}^0$ ) 的期望值是组别固定效应 ( $\alpha_g$ ) 和时间固定效应 ( $\lambda_t$ ) 的线性组合, 即  $E[Y_{i,g,t}^0 | g, t] = \alpha_g + \lambda_t$ 。<sup>2</sup> 此时, TWFE 回归模型 (1) 是潜在结果变量的饱和模型 (Saturated Model):<sup>3</sup>

$$Y_{i,g,t} = \beta D_{g,t} + \alpha_g + \lambda_t + \epsilon_{i,g,t}, \quad (1)$$

其中  $Y_{i,g,t}$  为观测到的结果变量,  $D_{g,t}$  为处理状态的二元变量,  $\epsilon_{i,g,t}$  为误差项,  $E[\epsilon_{i,g,t}] = 0$ 。在  $2 \times 2$  经典 DID 的情形下, TWFE 回归模型 (1) 不存在模型误设, 是正确设定的饱和模型, 其系数  $\beta$  可以识别出受处理个体的平均处理效应 (Average Treatment Effect on the Treated, 下文简

---

<sup>1</sup>具体而言, 在条件独立假设下, 若在常系数回归模型中对控制变量进行离散化的饱和控制, 处理变量前的单一系数可以识别出以处理变量条件方差为额外权重的平均处理效应。由于方差的取值非负, 可以证明, 回归系数是异质性处理效应的凸组合。详细的计量理论证明出现在 Angrist (1998) 的第 256 页, 亦是 Angrist & Pischke (2009) 经典教材第三章的核心内容。更早的类似论证见于 Angrist (1991) 对 2SLS 的讨论。

<sup>2</sup>平行趋势假设还有另外一种常见写法: 对于任意的  $t \geq 2$ ,  $E[Y_{i,g,t}^0 - Y_{i,g,t-1}^0 | g]$  不随  $g$  变化, 即后文的假设 1。

<sup>3</sup>饱和模型, 即对离散解释变量的每一组取值, 模型中都有特定的参数值与之对应。

称 ATT), 即  $E[Y_{i,g,t}^1 - Y_{i,g,t}^0 | D_{g,t} = 1]$ .<sup>4</sup>

Angrist & Pischke (2009) 通过总结已有的经验研究, 发现 TWFE 回归的应用早已不限于  $2 \times 2$  经典 DID 的情形。作为一项“经验研究者指南”(An Empiricist's Companion), 他们提炼出 TWFE 回归在三个维度的推广, 极大地推动了 TWFE 回归的广泛应用。<sup>5</sup>

- 推广 1: 二元交错处理变量。此时, 各组在不同时间点接受处理。例如, 有些地区较早开始实施政策, 而有些地区则较晚开始实施政策。<sup>6</sup>
- 推广 2: 非二元处理变量。此时, 模型 (1) 中的  $D_{g,t}$  不再是一个是否接受处理的二元变量, 而是一个多值离散型变量或连续型变量。例如,  $D_{g,t}$  可代表政策实施的强度。<sup>7</sup>
- 推广 3: 额外的协变量。例如, 在回归中控制同时期的其他政策变动。<sup>8</sup>

然而, 这三类推广都没有经过严格的理论证明。研究者在应用这三类推广时, 仍然依赖  $2 \times 2$  经典 DID 的平行趋势假设, 并将回归系数  $\beta$  阐释为某类平均处理效应。笔者发现, 三类推广都存在模型误设: 仅依靠未受处理潜在结果变量的平行趋势假设, 无法用回归系数  $\beta$  识别出异质性处理效应的凸组合。因此, 在 TWFE 回归的三类推广中, “无伤大雅”的思路失效了。

尽管未经严格证明, TWFE 回归的三类推广因其极高的灵活性和便利性, 在中国经济研究中得到了广泛应用。笔者逐篇阅读了 2019–2021 年间发表于五本中国经济研究旗舰期刊的论文, 发现有 473 篇着重运用 TWFE 回归, 占发文总量的 22%。<sup>9</sup> 图 1 按具体应用情形将这 473 篇文章分为四类。其中, 100 篇文章近似于  $2 \times 2$  经典 DID 情形, 仅占 TWFE 回归应用的 21% 左右。三类推广主导了 TWFE 回归的应用: 109 篇文章属于推广 1 二元交错处理变量的情形, 145 篇文章属于推广 2 非二元处理变量的情形, 而几乎所有文章都涉及推广 3 中添加额外协变量的情形。<sup>10</sup> 运用 TWFE 回归模型研究中国经济问题, 急需严谨的理论基础。

本文构建了模型误设的统一框架, 为 TWFE 回归的三类推广情形提供了理论基础。具体而言, 笔者将多种常系数回归模型对应到可以刻画异质性处理效应的饱和模型, 得出常系数是异质性处理效应的线性组合 (表 1)。然而, 在 TWFE 回归的推广情形下, 仅依赖未受处理潜在结果的平行趋势, 常系数不再是异质性处理效应的凸组合, 组合的权重可能为负, 从而无法得到“无伤大雅”的近似估计。究其根源, 常系数 TWFE 回归的推广情形不再以未受处理组为唯一控制组, 不能仅仅基于未受处理潜在结果的平行趋势就作出因果推断, 这就是模型误设的根源所在。

<sup>4</sup> 本文采用 Angrist & Pischke (2009) 的符号。令  $g = 1$  为处理组,  $g = 2$  为控制组,  $t = 1$  为处理前,  $t = 2$  为处理后。只有处理组 ( $g = 1$ ) 的个体  $i$  在处理后 ( $t = 2$ ) 受到政策处理, 即  $D_{1,2} = 1$ , 其余情况下  $D_{g,t} = 0$ 。由模型 (1) 可得:  $E[Y_{i,1,2}] = \alpha_1 + \lambda_2 + \beta$ ,  $E[Y_{i,1,1}] = \alpha_1 + \lambda_1$ ,  $E[Y_{i,2,2}] = \alpha_2 + \lambda_2$ ,  $E[Y_{i,2,1}] = \alpha_2 + \lambda_1$ 。因此, DID 估计量与模型 (1) 的系数  $\beta$  相等:  $DID = E[Y_{i,1,2} - Y_{i,1,1}] - (E[Y_{i,2,2} - Y_{i,2,1}]) = ((\alpha_1 + \lambda_2 + \beta) - (\alpha_1 + \lambda_1)) - ((\alpha_2 + \lambda_2) - (\alpha_2 + \lambda_1)) = \beta$ 。当满足平行趋势假设时,  $DID = \beta = ATT$ 。相关证明出现在 Angrist & Pischke (2009) 的第五章。此外,  $2 \times 2$  经典 DID 的情形下 TWFE 回归模型还有一种常见写法:  $Y_{i,g,t} = \beta Treat_g \times Post_t + \alpha Treat_g + \lambda Post_t + \epsilon_{i,g,t}$ 。其中, 当  $g$  为处理组时,  $Treat_g = 1$ , 否则  $Treat_g = 0$ ; 当  $t$  为处理后时期时,  $Post_t = 1$ , 否则  $Post_t = 0$ 。

<sup>5</sup> 这三类推广, 被统称为广义双重差分。此外, 还有一支文献围绕平行趋势假设展开。在附录 C 中, 笔者详细讨论了有关平行趋势假设的理论分析和最新研究, 为广义双重差分方法的应用提供参考。

<sup>6</sup> “可以方便地在回归模型中加入多个处理组和多个政策处理时点”(Angrist & Pischke, 2009, 第 175 页)

<sup>7</sup> “用回归来估计双重差分方法的第二个优势是, 它能够更方便地在实证研究中使用除了二元变量之外的其他处理变量。”(Angrist & Pischke, 2009, 第 175 页)

<sup>8</sup> “Card (1992) 的分析展示了用回归来估计双重差分方法的另一个优势: 在这个框架下, 添加额外的协变量非常容易。”(Angrist & Pischke, 2009, 第 176 页)

<sup>9</sup> 这五本期刊为《经济研究》、《管理世界》、《经济学》(季刊)、《世界经济》和 *China Economic Review*。

<sup>10</sup> 还有 119 篇文章不涉及因果推断框架, 构成图 1 中的“其他”类别。

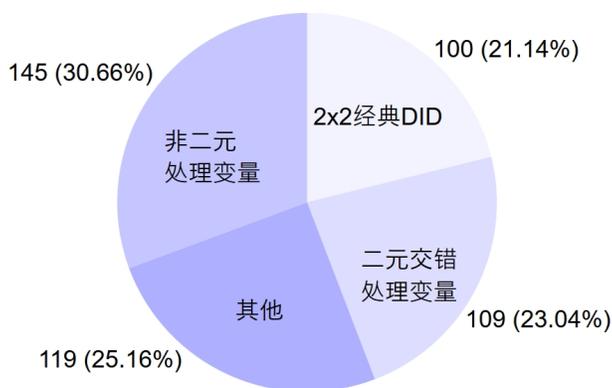


图 1: 不同文章类型的数量与占比

<sup>11</sup> 对误设的 TWFE 回归模型进行估计，会出现各类“禁止的对比”，从而产生“负权重”问题，最终可能使估计结果出现严重偏误（如 de Chaisemartin & D’Haultfœuille, 2020, 2023, 2024a; Goodman-Bacon, 2021; Lin and Zhang, 2022; Borusyak et al., 2023; Caetano and Callaway, 2023）。

基于模型误设的统一框架，笔者为 TWFE 回归和广义 DID 方法的应用提供了全面的研究指南、操作手册以及教学课件，深化了中国经济研究的方法基础。现有文献很好地总结了 DID 方法的理论发展，但尚无一个统一的理论框架来理解 TWFE 回归在广义 DID 方法中的适用性、局限性和应用范围。<sup>12</sup> 在统一框架下，本文进一步做出两点贡献。第一，已有研究指南聚焦于二元交错处理变量，即推广 1 的情形。本文进一步研究了推广 2 非二元处理变量的情形，提炼出基于工具变量法的 TWFE 回归模型改进思路。第二，本文还讨论了添加额外协变量所需的基本假设，明确了推广 3 的应用条件。模型误设的统一框架，可以引导研究者深入了解研究问题的现实背景，以经济理论指导实证模型设定，从而深化对中国经济发展规律的探索。

本文的其余部分安排如下。第 2 节追溯了“无伤大雅”思路的缘起与推广。第 3–5 节分别论述了 TWFE 回归三类推广下模型误设的具体形式，如何引入“禁止的对比”并导致“负权重”问题，以及可能的模型改进方法。第 6 节是总结性评述。

<sup>11</sup>例如，在推广 1 二元交错处理变量的情形下，较早开始实施政策的组会作为控制组（Goodman-Bacon, 2021）；在推广 2 非二元处理变量的情形下，处理强度较低的组会作为控制组（de Chaisemartin & D’Haultfœuille, 2024a）；在推广 3 添加额外控制变量的情形下，额外控制变量为另一个处理变量时，较早开始实施其他政策的组会作为控制组（de Chaisemartin & D’Haultfœuille, 2023）。

<sup>12</sup>如 de Chaisemartin & D’Haultfœuille (2022)、Chiu et al. (2023)、Roth et al. (2023)、刘冲、沙学康和张妍 (2022)、许文立 (2023) 以及张子尧和黄炜 (2023) 等文章，对 DID 方法进行了精彩的理论梳理。

表 1: 模型误设框架: 常数回归模型与处理效应异质性

基本设定	简单 OLS	二元交错处理变量 (静态模型)	二元交错处理变量 (动态模型)	非二元交错处理变量 (静态模型)
参考文献	Angrist (1998)	de Chaisemartin & D'Haultfoeulle (2020)	Sun & Abraham (2021)	de Chaisemartin & D'Haultfoeulle (2024a)
主要识别假设	条件独立假设	平行趋势假设	平行趋势假设	平行趋势假设
常回归模型/TWFE 回归模型	$Y_i = \beta D_i + \sum_{x=x_1}^{x_m} \lambda_x I_i^x + \epsilon_i$	$Y_{i,g,t} = \beta D_{g,t} + \alpha_g + \lambda_t + \epsilon_{i,g,t}$	$Y_{i,g,t} = \sum_{l \neq -1} \beta^l D_{g,t}^l + \alpha_g + \lambda_t + \epsilon_{i,g,t}$	$Y_{i,g,t} = \beta D_{g,t} + \alpha_g + \lambda_t + \epsilon_{i,g,t}$
异质效应回归模型/异质效应 TWFE 回归模型	$Y_i = \sum_{x=x_1}^{x_m} \beta_x D_i I_i^x + \sum_{x=x_1}^{x_m} \lambda_x I_i^x + \epsilon_i$	$Y_{i,g,t} = \sum_{(g,t) \in \Omega} \beta_{g,t} I_{g,t} D_{g,t} + \alpha_g + \lambda_t + \epsilon_{i,g,t}$	$Y_{i,g,t} = \sum_{l \in L_{in}} \sum_e \beta_e^l I_e D_{g,t}^l + \sum_{l \in L_{ex}} \sum_e \beta_e^l I_e D_{g,t}^l + \alpha_i + \lambda_t + \epsilon_{i,g,t}$	$Y_{i,g,t} = \sum_{(g,t) \in \Omega} \sum_{d_1}^{d_m} \beta_{g,t}^{d_1} I_{g,t}^{d_1} \{D_{g,t} > 0\} + \alpha_g + \lambda_t + \epsilon_{i,g,t}$
参数分解公式	$\beta = \sum_{x=x_1}^{x_m} \omega_x \beta_x$	$\beta = \sum_{(g,t) \in \Omega} \omega_{g,t} \beta_{g,t}$	$\beta^l = \sum_{l' \in L_{in} \cup L_{ex}} \sum_e w_{e,l'}^l \beta_{e,l'}^l$	$\beta = \sum_{(g,t) \in \Omega} \sum_{d_1}^{d_m} \omega_{g,t}^{d_1} \beta_{g,t}^{d_1}$
权重	$\omega_x = \frac{\frac{N_x}{N} \text{Var}(D_i   X_i = x)}{\sum_{x=x_1}^{x_m} \frac{N_x}{N} \text{Var}(D_i   X_i = x)}$	$\omega_{g,t} = \frac{N_{g,t} \bar{D}_{g,t}}{\sum_{(g,t) \in \Omega_1} N_{g,t} \bar{D}_{g,t}} D_{g,t}$	<ol style="list-style-type: none"> <li><math>\sum_e w_{e,l}^l = 1</math>;</li> <li>对任意 <math>l' \in L_{in}</math>, 并且 <math>l' \neq l</math>, <math>\sum_e w_{e,l'}^l = 0</math>;</li> <li><math>\sum_{l' \neq l, l' \in L_{ex}} \sum_e w_{e,l'}^l = -1</math></li> </ol>	$\omega_{g,t}^d = \frac{N_{g,t}^d \bar{D}_{g,t}}{\sum_{(g,t) \in \Omega_2} N_{g,t} D_{g,t} \bar{D}_{g,t}} \mathbf{1}\{D_{g,t} > 0\}$

注: 非二元处理变量的动态模型参见脚注 34, 其系数分解公式可参见 de Chaisemartin and D'Haultfoeulle (2024a); 包含控制变量的 TWFE 回归模型的系数分解公式可参见 Caetano and Callaway (2023)、Lin and Zhang (2022) 和张征宇、林丽花、曹思力和周亚虹等 (2024)。由于文章篇幅限制, 笔者不在此展示。

## 2 “无伤大雅”思路的缘起与推广

### 2.1 缘起：条件独立假设下常系数回归的模型误设

经验主义范式下“无伤大雅”的分析思路，起源于 Angrist (1998) 研究志愿参军经历影响劳动力市场表现的经典论文。Angrist 基于美国八十年代的志愿兵役制，认为报名入伍的申请者是否被选中具有一定的随机性。Angrist 假设，在控制了美军用于筛选申请者的变量后，申请者是否被选中入伍与其未来在劳动力市场上的潜在收入不相关。也就是说，是否入伍满足条件独立假设： $\{Y_i^0, Y_i^1\} \perp D_i | X_i$ 。其中， $D_i$  为是否入伍的指示变量； $X_i$  代表美军用于筛选申请者的变量，如种族、年龄、入伍考试成绩等； $Y_i^1$  ( $Y_i^0$ ) 是有 (无) 入伍经历时的未来劳动力市场收入。

在条件独立假设下，Angrist (1998) 研究了常系数回归模型 (2)：<sup>13</sup>

$$Y_i = \alpha + \beta D_i + X_i' \lambda + \epsilon_i. \quad (2)$$

为了简化对控制变量的讨论，Angrist (1998) 假设所有控制变量都是离散的，可以进行饱和控制。这等价于只有一个离散控制变量  $X_i$ ，其取值  $x_1, x_2, \dots, x_m$  涵盖离散控制变量所有可能取值的组合。此时，常系数回归模型 (2) 可以表达为如下形式：<sup>14</sup>

$$Y_i = \beta D_i + \sum_{x=x_1}^{x_m} \lambda_x I_i^x + \epsilon_i, \quad (3)$$

其中  $I_i^x = \mathbf{1}\{X_i = x\}$  为指示变量，若  $X_i = x$ ，则  $I_i^x$  取值为 1，反之为 0。由于控制变量已经做了饱和控制，如果不存在处理效应异质性，模型 (3) 是正确设定的饱和回归模型。此时， $\beta$  可以识别平均处理效应 (Average Treatment Effect, 下文简称 ATE)，即  $E[Y_i^1 - Y_i^0]$ 。如果处理效应具有异质性， $\beta$  还能识别 ATE 吗？Angrist (1998) 回答道：不能，但近似能。因此常系数回归模型是“无伤大雅”的。

实际上，当处理效应具有异质性时，常系数回归模型 (3) 存在误设。在条件独立假设下，正确设定的回归模型为：

$$Y_i = \sum_{x=x_1}^{x_m} \beta_x D_i I_i^x + \sum_{x=x_1}^{x_m} \lambda_x I_i^x + \epsilon_i. \quad (4)$$

与模型 (3) 相比，模型 (4) 允许了处理效应随控制变量取值的变动而变动。可以证明，在条件独立假设下，模型 (4) 中的  $\beta_x$  识别了控制变量取值为  $x$  的群体的平均处理效应  $E[Y_i^1 - Y_i^0 | X_i = x]$ ，因此模型 (4) 的设定是正确的。

在模型误设的统一框架下，笔者推导出了  $\beta$  和  $\beta_x$  的关系。根据遗漏变量偏误 (Omitted Variable Bias, 下文简称 OVB) 公式和 Frisch-Waugh-Lovell (下文简称 FWL) 定理，可以得到分解

<sup>13</sup>在分析实际数据时，研究者一般基于某个抽样样本，运用普通最小二乘法 (Ordinary Least Square, 下文简称 OLS) 对模型 (2) 进行估计。笔者参考 Angrist & Pischke (2009)，数学推导一般直接对总体回归模型展开论证。

<sup>14</sup>模型 (3) 中包含  $X_i$  的所有可能取值下的虚拟变量，因此没有截距项。

公式：

$$\beta = \sum_{x=x_1}^{x_m} \omega_x \beta_x, \quad (5)$$

其中  $\omega_x$  的表达式为：

$$\omega_x = \frac{\frac{N_x}{N} \text{Var}(D_i|X_i = x)}{\sum_{x=x_1}^{x_m} \frac{N_x}{N} \text{Var}(D_i|X_i = x)},$$

$N$  为个体总数， $N_x$  是  $X_i$  取值为  $x$  的个体数量。可见， $\omega_x \geq 0$  且  $\sum_{x=x_1}^{x_m} \omega_x = 1$ 。<sup>15</sup> 在条件独立假设下，根据分解公式(5)，只有当处理状态的条件方差  $\text{Var}(D_i|X_i = x)$  为常数时， $\beta$  才能识别 ATE。

一般而言，处理状态条件方差是可变的，常系数回归模型(3)中的  $\beta$  不能识别 ATE。在条件独立假设下， $\beta$  识别了一种以处理状态条件方差为额外权重的平均处理效应。此时， $\beta$  的估计量是一种非参匹配估计量，所有权重非负 ( $\omega_x \geq 0$ )，且权重之和为 1 ( $\sum_{x=x_1}^{x_m} \omega_x = 1$ )，满足用于因果推断的综合参数的基本条件 (Blandhol et al., 2022)。回归模型(3)以牺牲估计参数的准确性为代价，换取了模型的简洁性，但仍能满足权重非负的基本要求，可谓“无伤大雅”。<sup>16</sup> 尽管常系数回归模型存在模型误设，并不是估计 ATE 的完美工具，但它已然成为众多经验研究的起始点。

## 2.2 推广：平行趋势假设下双重固定效应回归的模型误设

与条件独立假设类似，平行趋势假设是基于观测数据使用双重差分方法进行因果推断的常用基本假设。Angrist & Pischke (2009) 在分析平行趋势假设下的双重差分方法时，进一步贯彻了“无伤大雅”思路。具体而言，Angrist & Pischke (2009) 总结已有文献，发现 TWFE 回归不限于  $2 \times 2$  经典 DID 情形，并基于“无伤大雅”思路提炼出三类推广，默认 TWFE 回归在三类推广情形下仍然能够近似估计出某种平均处理效应。在平行趋势假设下，目标估计量一般为 ATT。然而，近年来的计量经济学研究表明，TWFE 回归的三类推广都难以得到 ATT 的近似估计，甚至可能得出与 ATT 相反的估计量 (Borusyak et al., 2023; Caetano & Callaway, 2023; de Chaisemartin & D’Haultfoeuille, 2020, 2023, 2024a; Goodman-Bacon, 2021; Lin & Zhang, 2022)。

接下来，本文将使用模型误设的统一框架解析 TWFE 回归在三类推广情形下引致偏误的底层原因。为方便论述，笔者对 TWFE 回归应用环境的数据结构和基本假设进行界定。考虑一个可以分为  $G$  个组、 $T$  个时期的总体，其中具体的组与时期分别由  $g$  和  $t$  表示。每一对  $(g, t) \in \{1, \dots, G\} \times \{1, \dots, T\}$  中包含的个体  $i$  的数量为  $N_{g,t}$ ，且  $N = \sum_{g,t} N_{g,t}$  为总体中的个体总数。<sup>17</sup> 组  $g$  当中的个体  $i$  在  $t$  期的结果变量为  $Y_{i,g,t}$ ，处理变量为  $D_{g,t} \in \mathbf{D}_t$ ，其中  $\mathbf{D}_t$  为  $t$  期中处理变量所有取值的集合。<sup>18</sup> 双重差分方法的基本假设如下：

<sup>15</sup>推导证明见附录B.1。

<sup>16</sup>Słoczyński (2022) 在 Angrist (1998) 的基础上，提出了对常系数回归模型中  $\beta$  的 OLS 估计量的新解释，认为其可以表示为对处理组的平均处理效应和控制组的平均处理效应的加权平均，尽管权重之和为 1，但权重的大小与群体大小成反比关系，即更小的群体在 OLS 估计量中占据更大的权重。

<sup>17</sup>总体也可能来自同一个日历时间，由出生队列扮演时期  $t$  的角色；也有可能对于所有的  $(g, t)$  对，都有  $N_{g,t} = 1$ ，例如，一个组就是一个个体、公司或地区。

<sup>18</sup>这里笔者假设精准设计 (Sharp Design)，同一组内个体所接受的处理是一致的， $D_{i,g,t} = D_{g,t}$ 。

假设 1 平行趋势：对于任意的  $t \geq 2$ ,  $E \left[ Y_{i,g,t}^0 - Y_{i,g,t-1}^0 \middle| g \right]$  不随  $g$  变化。

假设 1 是未受处理潜在结果的平行趋势假设，要求未受处理潜在结果变量的变化趋势没有组间差异。在假设 1 之下，Angrist & Pischke (2009) 提炼出常系数 TWFE 回归模型的三个推广，扩展模型如下：

$$Y_{i,g,t} = \beta D_{g,t} + \mathbf{X}'_{i,t} \boldsymbol{\delta} + \boldsymbol{\alpha}_g + \lambda_t + \epsilon_{i,g,t}, \quad (6)$$

其中，不同组处理状态  $D_{g,t}$  的变动时点可以不同，允许多个处理组和多个处理时点（推广 1）； $D_{g,t}$  也可以是非二元变量，即个体或组接受的处理强度不同（推广 2）；回归中还可以添加额外的协变量  $\mathbf{X}_{i,t}$ （推广 3）。尽管这些推广显著扩大了常系数 TWFE 回归模型的适用范围，但是都存在严重的模型误设。在三类推广下，仅依赖未受处理组潜在结果的平行趋势假设，模型 (6) 无法得到 ATT 的合理近似。具体而言：

推广 1 允许多个处理组和多个处理时点，此时基于 TWFE 回归模型估计不仅将未受处理组作为控制组，还会将较早接受处理的组作为控制组 (Goodman-Bacon, 2021)。这与仅依赖假设 (1) 的 DID 方法不一致，从而导致了模型误设。

推广 2 允许处理变量为非二元变量，此时未受处理组不再是唯一控制组。当处理时点为单时点时，处理强度较低的组会被用作控制组；当处理时点为多时点时，较早接受处理且处理强度较低的组会被用作控制组 (de Chaisemartin & D'Haultfoeuille, 2024a)。因此，不管是单时点还是多时点，仅依赖假设 (1) 进行 TWFE 回归均会导致模型误设。

推广 3 允许在 TWFE 回归模型中添加额外协变量。当协变量为额外的副处理变量时，主处理变量未发生变化、副处理变量发生变化的组会被用作控制组，也不再以未受处理组为唯一控制组 (de Chaisemartin & D'Haultfoeuille, 2023)。在模型 (6) 中添加时变控制变量  $\mathbf{X}_{i,t}$  的普遍情形，也存在类似的模型误设 (Caetano & Callaway, 2023; Lin & Zhang, 2022)。

接下来，本文将对这三类推广情形下 TWFE 回归模型的误设逐一进行具体探讨。

### 3 第一类模型误设：二元交错处理变量情形

我国在探索一项新政策时，通常采用“试点先行，逐步推广”的策略。政府会先以部分地区为试点试行某项政策，然后根据试点效果和反馈，逐步将该政策改善并推广到更多地区。例如，我国在推行新型农村合作医疗保险、新型农村养老保险以及高速铁路建设等重大政策时，都采取了“试点先行，逐步推广”的策略。与  $2 \times 2$  DID 的数据结构相比，此时处理状态仍然为二元变量，但个体在不同时间点接受处理，被称为“二元交错处理变量”的情形。<sup>19</sup> 现有文献对这类政策进行评估时，往往采用模型 (1)，即对 TWFE 回归的推广 1。此时，未受处理组不是唯一控制组，仅依赖未受处理潜在结果的平行趋势假设进行 TWFE 回归就会导致模型误设，常系数不再是异质性处理效应的凸组合，从而无法得到“无伤大雅”的估计结果。

以评估新型农村合作医疗保险（下文简称“新农合”）的效果为例，该政策于 2003 年开始推行，到 2008 年几乎覆盖了中国所有县 (Gruber et al., 2023)。基于模型 (1) 评估该政策对居民健康的影响时，已经实施“新农合”的地区会被视作控制组；只要“新农合”的健康效应随实施年份

<sup>19</sup>本节的讨论聚焦于一旦接受处理，处理状态不再变化的情形，即处理变量为“吸收状态 (Absorbing State)”。

递增，TWFE 回归就会低估“新农合”对居民健康的改善作用；如果“新农合”的健康效应在地区间也存在异质性，TWFE 回归的模型误设会更严重。这类二元交错处理变量情形在中国经济研究中非常普遍。如图1所示，属于该情形的文章数量达 109 篇，占 TWFE 回归应用的 23% 左右。因此，以模型误设的统一框架来理解二元交错处理变量情形下的 TWFE 回归，对中国经济研究有着重大现实意义。

### 3.1 静态模型

二元交错处理变量情形下的 TWFE 回归，有静态模型 (如模型 (1)) 和动态模型两种形式。笔者首先分析静态模型。<sup>20</sup>在潜在结果框架下，笔者想要估计的目标参数—处理组的平均处理效应—可以表示为：

$$ATT = \sum_{(g,t) \in \Omega_1} \frac{N_{g,t}}{\sum_{(g,t) \in \Omega_1} N_{g,t}} ATT(g,t),$$

其中，其中  $\Omega_1 = \{(g,t) | D_{g,t} = 1\}$ ，是所有  $D_{g,t} = 1$  的观测值的集合。 $ATT(g,t) = E \left[ Y_{i,g,t}^1 - Y_{i,g,t}^0 \mid g, t \right]$  为组  $g$  在  $t$  期的平均处理效应，ATT 是  $ATT(g,t)$  根据样本量  $N_{g,t}$  加权平均的处理效应。

实证研究者通常希望通过回归模型 (1) 中  $\beta$  的估计量，来获得 ATT 的无偏估计。然而，回归模型 (1) 作为一个常系数 TWFE 回归模型，同一般的常系数回归模型 (2) 一样，需要满足处理效应同质性的假设。这意味着， $ATT(g,t)$  不随分组  $g$  或时间  $t$  发生改变。在现实情境中，处理效应同质性的假设往往难以得到满足。例如，考虑到“新农合”的实施过程中，相较于后来实施该制度的地区，最初试点的地区往往经济更发达、财政实力更雄厚。因此，可以合理推断“新农合”在不同地区的实施效果可能存在差异。在这种情况下， $\beta$  是否仍然能够无伤大雅地近似识别 ATT？答案是否定的。这一结论的根源在于 Borusyak et al. (2023) 提出的在回归模型 (1) 中存在“禁止的对比”：将先接受处理组在处理前作为后接受处理组的控制组。<sup>21</sup>

为了更清楚地说明这一点，笔者可以使用一个最简单的  $2 \times 3$  DID 为例来阐释，更为复杂的情况可以视为多个  $2 \times 3$  DID 的叠加。样本分为两个组 ( $G = 2$ )，有三个样本期 ( $T = 3$ )，其中组  $g = 1$  在  $t = 2$  期开始接受处理，组  $g = 2$  从  $t = 3$  期开始接受处理。两组在各期接受处理的情况如表2(a)所示。<sup>22</sup>

表 2:  $2 \times 3$  二元交错处理变量示例

	(a) $2 \times 3$ DID			(b) $2 \times 2$ DID <sub>1</sub>			(c) $2 \times 2$ DID <sub>2</sub>		
	$t = 1$	$t = 2$	$t = 3$	$t = 1$	$t = 2$		$t = 2$	$t = 3$	
$g = 1$	0	1	1	$g = 1$	0	1	$g = 1$	1	1
$g = 2$	0	0	1	$g = 2$	0	0	$g = 2$	0	1

这一  $2 \times 3$  DID 可以进一步分解为两个  $2 \times 2$  DID。首先是  $t = 1$  和  $t = 2$  期，如表2(b)所示，组  $g = 1$  在  $t = 2$  时接受处理，作为处理组， $g = 2$  两期均未接受处理，作为控制组，这是  $2 \times 2$

<sup>20</sup>静态模型隐含无预期效应假设，即结果变量与未来是否接受处理无关。

<sup>21</sup>这也包括样本期内始终接受处理组作为控制组的情况。此外，根据笔者后面的分析， $\beta$  不能够识别 ATT 的另一个原因是回归模型 (1) 中  $\beta$  的估计量并非  $ATT(g,t)$  的基于样本量的加权平均。

<sup>22</sup>为了表达简洁，笔者假设各组样本数相同，但这并不影响分析结论。

经典 DID 的形式。如果满足假设1，那么 DID 估计值  $DID_1$  为  $g = 1$  在  $t = 2$  时的平均处理效应  $ATT(1,2)$  的无偏估计。第二个  $2 \times 2$  DID 如表2(c) 所示， $g = 2$  从  $t = 3$  期开始接受处理，作为处理组， $g = 1$  两期均接受处理，因其处理状态没有发生改变，可以作为控制组。根据这一设定，笔者对表2(c) 进行如下 DID 估计：

$$\begin{aligned}
DID_2 &= E[Y_{i,2,3}] - E[Y_{i,2,2}] - (E[Y_{i,1,3}] - E[Y_{i,1,2}]) \\
&= E[Y_{i,2,3}^1] - E[Y_{i,2,2}^0] - (E[Y_{i,1,3}^1] - E[Y_{i,1,2}^1]) \\
&= E[Y_{i,2,3}^1] - E[Y_{i,2,2}^0] - (E[Y_{i,2,3}^0] - E[Y_{i,2,2}^0]) + (E[Y_{i,2,3}^0] - E[Y_{i,2,2}^0]) - (E[Y_{i,1,3}^1] - E[Y_{i,1,2}^1]) \\
&= E[Y_{i,2,3}^1] - E[Y_{i,2,3}^0] + (E[Y_{i,1,3}^0] - E[Y_{i,1,2}^0]) - (E[Y_{i,1,3}^1] - E[Y_{i,1,2}^1]) \\
&= E[Y_{i,2,3}^1] - E[Y_{i,2,3}^0] + (E[Y_{i,1,2}^1] - E[Y_{i,1,2}^0]) - (E[Y_{i,1,3}^1] - E[Y_{i,1,3}^0]) \\
&= ATT(2,3) + ATT(1,2) - ATT(1,3),
\end{aligned}$$

其中第四个等号运用了平行趋势假设。可见这一双重差分的估计值  $DID_2$  并非  $g = 2$  在  $t = 3$  的处理效应  $ATT(2,3)$ ，还混合了  $g = 1$  在  $t = 2$  和  $t = 3$  的处理效应  $ATT(1,2)$  和  $ATT(1,3)$ ；更重要的是  $ATT(1,3)$  是以其相反数进入  $DID_2$  的估计当中，带来了“负权重”问题。只有当  $ATT(1,2) = ATT(1,3)$  时，即  $g = 1$  在  $t = 2$  和  $t = 3$  的处理效应相等， $DID_2$  才可以识别  $ATT(2,3)$ 。这实际上隐含着一个时间维度上的处理效应同质性假设，即处理效应不随时间的推移而发生变化。

Goodman-Bacon (2021) 对这一问题进行详细阐述，将常系数 TWFE 回归模型 (1) 中  $\beta$  的 OLS 估计量分解为所有  $2 \times 2$  DID 组合的加权平均 (Bacon Decomposition, 培根分解法)。将这一分解法应用到表 2 的例子当中，回归模型 (1) 中的  $\beta$  可表达为：

$$\begin{aligned}
\beta &= \frac{1}{2}DID_1 + \frac{1}{2}DID_2 \\
&= \frac{1}{2}ATT(1,2) + \frac{1}{2}[ATT(2,3) + ATT(1,2) - ATT(1,3)] \\
&= ATT(1,2) + \frac{1}{2}ATT(2,3) - \frac{1}{2}ATT(1,3)
\end{aligned}$$

如果  $ATT(1,2) = ATT(2,3) = 1$ ， $ATT(1,3) = 4$ ，那么  $\beta = -\frac{1}{2}$ 。这种情况下，尽管所有的处理效应实际上都是正向的，常系数 TWFE 回归模型的估计结果却是负值。这表明， $\beta$  不仅无法准确估计处理效应，甚至还可能得出与实际效应相反的结论。因此，在二元交错数据结构下使用常系数 TWFE 回归模型可能导致严重偏误。

出现这一问题的根源是，在二元交错处理变量情形中，常系数 TWFE 回归模型存在模型误设。笔者接下来用模型误设的框架来进行分析。假设处理效应随时间发生改变，也可以理解为 TWFE 回归模型 (1) 遗漏了系列交互项  $\sum_{(g,t) \in \Omega} \beta_{g,t} I_{g,t} D_{g,t}$ ，其中  $I_{g,t}$  是个体  $i$  属于组  $g$  在时间  $t$  的虚拟变量， $\Omega$  代表所有  $(g,t)$  观测值的集合。那么相比于存在模型误设的常系数 TWFE 回归模型 (1)，

笔者应该估计的异质效应 TWFE 回归模型为：

$$Y_{i,g,t} = \sum_{(g,t) \in \Omega} \beta_{g,t} I_{g,t} D_{g,t} + \alpha_g + \lambda_t + e_{i,g,t}, \quad (7)$$

根据 OVB 公式和 FWL 定理，笔者可得到如下分解公式：

$$\beta = \sum_{(g,t) \in \Omega} \omega_{g,t} \beta_{g,t}, \quad (8)$$

其中  $\omega_{g,t}$  的表达式为

$$\omega_{g,t} = \frac{N_{g,t} \tilde{D}_{g,t}}{\sum_{(g,t) \in \Omega_1} N_{g,t} \tilde{D}_{g,t}} D_{g,t}, \quad (9)$$

其中， $\tilde{D}_{g,t} = D_{g,t} - \bar{D}_{\cdot,t} - \bar{D}_{g\cdot} + \bar{D}$ ， $\bar{D}_{\cdot,t} = \frac{1}{N_t} \sum_g D_{g,t}$ ， $\bar{D}_{g\cdot} = \frac{1}{N_g} \sum_t D_{g,t}$ ， $\bar{D} = \frac{1}{N} \sum_{g,t \in \Omega} D_{g,t}$ 。<sup>23</sup>

当满足假设1时，回归模型(7)中的  $\beta_{g,t}$  可以识别组  $g$  在时间  $t$  的平均处理效应  $ATT(g,t)$ 。依据分解公式(8)，回归模型(1)中  $\beta$  的估计量则是各组平均处理效应的加权平均。尽管这些权重的总和为1，但根据等式(9)，各组的平均处理效应权重取决于  $\tilde{D}_{g,t}$  的值，而  $\tilde{D}_{g,t} = D_{g,t} - \bar{D}_{\cdot,t} - \bar{D}_{g\cdot} + \bar{D}$  可能为负数，从而产生了“负权重”问题。根据 de Chaisemartin & D’Haultfoeuille (2020) 的研究，越早接受处理的组 ( $\bar{D}_{g\cdot}$  越大) 和在越多组受到处理的时期 ( $\bar{D}_{\cdot,t}$  越大)，其处理效应更可能被赋予“负权重”。<sup>24</sup>

尽管存在“负权重”的问题，但它并不一定导致灾难性后果。如果处理效应具有同质性，尤其是当处理效应不随时间变化时，由于权重之和为1，回归模型(1)中  $\beta$  仍然能够识别  $ATT(g,t)$ 。然而，当个体处理效应存在时间上的异质性时，“负权重”就会导致估计偏误；最严重的情况是，即使所有个体的处理效应都为正，由回归模型(1)可能得出负的估计值。这一现象表示，当存在处理效应异质性时，模型误设的常系数 TWFE 回归模型将不再给出“无伤大雅”的估计结果。

### 3.2 动态模型：事件研究法

时间维度上的处理效应异质性可能会导致静态模型在估计 ATT 时出现“负权重”问题。为了解决这一问题，一个直接有效的方法是在回归模型(1)的设定中直接纳入这种异质性并估计相关参数。这种做法就是如下所示的动态回归模型，文献当中通常称之为事件研究法：<sup>25</sup>

$$Y_{i,g,t} = \sum_{l \neq -1} \beta^l D_{g,t}^l + \alpha_g + \lambda_t + \epsilon_{i,g,t}, \quad (10)$$

<sup>23</sup>推导证明见附录B.2。

<sup>24</sup>值得注意的是，如果所有组在某个时期之后都接受了处理，那么在进行估计时应当排除这些时期，因为它们极有可能在静态模型的 DID 估计中引入“负权重”问题。此外，在后续讨论的动态模型中，由于缺乏控制组，这些时期的处理效应也无法识别。

<sup>25</sup>理论上来说，应该估计修正误设后的模型(7)。然而在实践中，考虑到该模型可能由于较低的自由度和有限的样本难以估计，研究者通常估计允许时间维度上的处理效应异质性的模型(10)，假设组别之间不存在处理效应异质性。关于事件分析法的应用指南详见黄炜、张子尧和刘安然(2021)。

其中  $D_{g,t}^l = I\{t - E_g = l\}$ ,  $E_g$  是  $g$  组开始接受处理的时间,  $D_{g,t}^l$  是代表首次接受处理之后  $l$  期的虚拟变量, 研究者们通常用  $\beta^l$  来识别首次接受处理之后  $l$  期的平均处理效应:

$$\text{ATT}(l) = \sum_e \frac{\sum_{(g,t):E_g=e,t=e+l} N_{g,t}}{\sum_{(g,t) \in \Omega_1} N_{g,t}} \text{ATT}(e,l),$$

其中,  $\text{ATT}(e,l) = E\left[Y_{i,g(e+l)}^1 - Y_{i,g(e+l)}^0 \mid E_g = e\right]$  为从  $e$  期开始接受处理的组经过  $l$  期之后的处理效应,  $\text{ATT}(l)$  是  $\text{ATT}(e,l)$  根据样本量加权平均的处理效应。

然而, Sun & Abraham (2021) 发现, 回归模型 (10) 中  $\beta^l$  的估计量, 不仅包括了距离处理开始时间  $l$  期的处理效应  $\text{ATT}(e,l)$ , 还包括了距离处理开始时间  $l'$  ( $l' \neq l$ ) 期的处理效应  $\text{ATT}(e,l')$ ; 当不同组之间处理效应存在差异时, 由于存在类似静态模型当中的“负权重”问题,  $\beta^l$  不能识别  $\text{ATT}(l)$ 。

笔者接下来使用模型误设的框架来分析  $\beta^l$  估计误差产生的来源。常系数 TWFE 回归模型 (10) 遗漏了两个重要的部分。首先, 由于不同组之间处理效应的时间路径可能存在差异, 即  $\text{ATT}(e,l)$  对于不同  $e$  可能不同, 由此导致的遗漏变量为  $\sum_{l \in L_{in}} \sum_e \beta_e^l I_e D_{g,t}^l$ , 其中  $L_{in}$  是包含在回归模型 (10) 中的相对时期的集合; 其次, 在估计回归模型 (10) 时需要剔除部分相对时间虚拟变量, 由此导致的遗漏变量为  $\sum_{l \in L_{ex}} \sum_e \beta_e^l I_e D_{g,t}^l$ , 其中  $L_{ex}$  是未被包含在回归模型 (10) 中的相对时期的集合。<sup>26</sup>如果将这两项都考虑在内, 那么真实的异质效应 TWFE 回归模型应该如下所示:

$$Y_{i,g,t} = \sum_{l \in L_{in}} \sum_e \beta_e^l I_e D_{g,t}^l + \sum_{l \in L_{ex}} \sum_e \beta_e^l I_e D_{g,t}^l + \alpha_i + \lambda_t + e_{i,g,t}, \quad (11)$$

根据 OVB 公式和 FWL 定理, 可得到如下分解公式:

$$\begin{aligned} \beta^l &= \sum_{l' \in L_{in} \cup L_{ex}} \sum_e w_{e,l'}^l \beta_e^{l'} \\ &= \sum_e w_{e,l}^l \beta_e^l + \sum_{l' \neq l, l' \in L_{in}} \sum_e w_{e,l'}^l \beta_e^{l'} + \sum_{l' \neq l, l' \in L_{ex}} \sum_e w_{e,l'}^l \beta_e^{l'}, \end{aligned} \quad (12)$$

其中  $w_{e,l'}^l$  来自于  $I_e D_{g,t}^{l'} = \tilde{\alpha}_i + \tilde{\lambda}_t + \sum_l w_{e,l'}^l + v_{i,g,t}$ 。<sup>27</sup>

分解公式 (12) 揭示了模型 (10) 中  $\beta^l$  的估计值可能存在误差的两个来源: 一是回归模型中其他时期的处理效应, 二是未被纳入回归模型的相对时期的处理效应。依据 FWL 定理, 可以明确  $\sum_{l'} \sum_e w_{e,l'}^l$  的三个关键性质:

- (1) 对于所有组的  $l$  期,  $\sum_e w_{e,l}^l = 1$ ;
- (2) 对于所有组的包括在模型 (10) 中任意  $l' \neq l$  且  $l' \in L_{in}$ ,  $\sum_e w_{e,l'}^l = 0$ ;
- (3) 对于所有组的排除在模型 (10) 的其他期,  $\sum_{l' \neq l, l' \in L_{ex}} \sum_e w_{e,l'}^l = -1$ 。

<sup>26</sup>实践当中通常将处理开始前一期作为基准期。此外, 当不存在始终未受处理组时, 为了去除多重共线性的问题, 笔者还需要排除一个相对时期的虚拟变量 (Borusyak et al., 2023)。

<sup>27</sup>推导证明见 Sun & Abraham (2021)。

性质 (2) 则表明, 如果不存在组间处理效应异质性, 那么分解公式 (12) 中的干扰项  $\sum_{l' \neq l, l' \in L_{in}} \sum_e w_{e,l'}^l \beta_e^{l'}$  将为 0。<sup>28</sup>性质 (3) 表明, 如果模型 (10) 中排除的时期的处理效应为 0, 那么分解公式 (12) 中的干扰项  $\sum_{l' \neq l, l' \in L_{ex}} \sum_e w_{e,l'}^l \beta_e^{l'}$  将为 0。在无预期效应假设下, 这通常可以通过排除  $l < 0$  的期数来实现。

总的来说, 对于任意  $l$ , 如果  $\sum_{l' \neq l} \sum_e w_{e,l'}^l$  的数值较大, 说明  $\beta^l$  受到其他期的处理效应的影响较大。在这种情况下, 对于回归模型 (10) 的估计结果应持谨慎态度, 并且需要怀疑基于该估计结果进行的事前趋势检验和动态效应估计的可靠性。<sup>29</sup>

### 3.3 “负权重”问题的识别诊断方法

在具体的实证研究过程中, 如何检验是否存在“负权重”问题? 首先, 研究者可以通过经济学理论结合具体情境来判断, 处理效应同质性的假设是否合理。此外, 现有文献提供了一些诊断方法, 可用于辅助判断处理效应异质性和“负权重”问题。

- (1) Goodman-Bacon (2021) 提出了培根分解法, 将回归模型 (1) 中  $\beta$  的 OLS 估计值分解为所有  $2 \times 2$  DID 处理效应的加权平均, 并给出每组 DID 处理效应的估计值和权重。据此, 研究者可直接观察前文所述“禁止的对比”(早处理组作为晚处理组的控制组、始终接受处理组作为控制组) 是否与其他 DID 组别(晚处理组作为早处理组的控制组、从未接受处理组作为控制组) 的估计值相反, 并且权重较大。该方法可以通过 STATA 命令 *bacondecomp* 来实现。
- (2) de Chaisemartin & D’Haultfœuille (2020) 提出直接计算  $\beta = \sum_{(g,t) \in \Omega} \omega_{g,t} \beta_{g,t}$  中的  $\omega_{g,t}$ , 并汇报其中“负权重”所占的比例。然而他们并未给出明确的阈值, 因此仅能作为“负权重”问题判断的参考。此外, 他们还基于 ATT(g,t) 标准差的最小值提出两个检验。这两个检验分别根据该数值判断“负权重”是否会导致 (1) 常系数 TWFE 回归模型的估计结果与实际处理效应方向相反或 (2) 与所有 ATT(g,t) 的处理效应方向相反。如果该值接近于零, 表明即使存在微小的处理效应异质性, 常系数 TWFE 回归模型的估计结果也可能出现偏误。相反, 如果该值较大, 说明只有在处理效应非常异质化的情况下, “负权重”才会使得常系数 TWFE 回归模型的估计结果失去意义, 此时回归模型 (1) 的估计结果在某种程度上对于真实的平均处理效应具有参考意义(至少在方向上)。然而, 这两个检验也并未明确提供阈值, 因此仅作为判断“负权重”问题严重程度的参考。这一方法可以通过 STATA 或 R 命令 *twowayfeweights* 来实现。
- (3) Sun & Abraham (2021) 提出, 对于每一个  $l$ , 可以计算出所有  $l' \neq l$  的  $w_{e,l'}^l$  的值, 从而获得回归模型 (10) 中  $\beta^l$  的估计值被  $l' \neq l$  时期处理效应混淆的程度, 进而可以判断模型 (10) 对于 ATT(l) 的识别受处理效应异质性的影响大小。这一方法都可以根据 STATA 命令 *eventstudyweights* 来实现。

<sup>28</sup>如果不存在组间处理效应异质性, 那么  $\beta_e^{l'} = \beta^{l'}$ 。干扰项  $\sum_{l' \neq l, l' \in L_{in}} \sum_e w_{e,l'}^l \beta_e^{l'} = \sum_{l' \neq l, l' \in L_{in}} \beta^{l'} \sum_e w_{e,l'}^l = 0$ , 是因为  $\sum_e w_{e,l'}^l = 0$ 。

<sup>29</sup>但笔者认为, 在实际应用中, 当期数较多时, 各期的异质处理效应在一定程度上会相互抵消, 使最终估计结果出现误差的可能较小。

以上检验方法当中，除了 Sun & Abraham (2021) 用于检验动态模型 (10) 以外，其他都适用于静态模型 (1)。其中，Goodman-Bacon (2021) 仅限于二元交错处理变量的情形，de Chaisemartin & D’Haultfœuille (2020) 可用于更加一般的情况。

### 3.4 处理效应异质性的稳健估计量

综合以上分析，在二元交错处理变量情形下，由于处理效应异质性，常系数 TWFE 回归模型可能产生估计偏误。为了应对这一挑战，学者们提出了若干新的估计方法，这些方法能够在存在处理效应异质性的情况下更准确地识别处理效应。这些方法的估计思路具有相似之处：首先，基于一定的平行趋势假设，选择合适的控制组，通常包括从未接受处理的组、尚未接受处理的组和最晚接受处理的组。这一选择旨在排除“禁止的对比”。<sup>30</sup> 通过这种方式，研究者能够构建干净、无偏的 DID 组合来识别  $ATT(g,t)$ 。其次，根据具体的实证研究需求，对估计得出的  $ATT(g,t)$  进行加权平均，以得到具有实际应用意义的处理效应。这包括处理组的平均处理效应  $ATT$ 、处理后某一特定时期的处理效应  $ATT(l)$  等。接下来，笔者将介绍新的估计量，并在附录 B.4 中讨论它们之间的差异和联系。

#### 3.4.1 队列平均处理效应估计量

Callaway & Sant’Anna (2021) 选取的控制组是在整个样本期未接受处理的样本 (Never-treated Group,  $E_g = \infty$ ) 或者到  $t$  期为止未接受处理的样本 (Not-yet-treated Group,  $E_g > t$ )，通过对比处理组和控制组的结果变量从  $E_g - 1$  期到  $t$  期变化的区别，来识别  $ATT(g,t)$ :

$$ATT^{CS}(g,t) = E \left[ Y_{i,g,t} - Y_{i,g,E_g-1} \right] - E \left[ Y_{i,g',t} - Y_{i,g',E_g-1} \right], \quad E_{g'} > t \text{ 或 } E_{g'} = \infty \quad (13)$$

其估计值  $\widehat{ATT}^{CS}(g,t)$  可以用样本均值来估计等式 (13) 中的期望得到。如果组数和期数较少，可以直接汇报所有的  $\widehat{ATT}^{CS}(g,t)$ ；当组数和期数较多时，Callaway & Sant’Anna (2021) 提供了一些将这些处理效应进行汇总的方式，例如常见的对应静态模型的平均处理效应  $ATT$ :

$$ATT^{CS} = \sum_{g,t} w_{g,t} ATT^{CS}(g,t),$$

<sup>30</sup>对于始终受处理组 ( $E_g = -\infty$ )，由于缺乏未受处理时的观察数据，其处理效果难以辨识。此外，为了规避“禁止的对比”，尽管这些组别的处理状态没有改变，他们一般不会被选作控制组，而是排除在研究样本之外。然而实践当中，假如始终受处理组首次接受处理的时间距离研究样本的时间点已经很远，有理由推断其处理效应已经稳定时，这样的组有潜力作为控制组参与分析，但需要满足如下平行趋势假设：始终受处理组在接受处理后的结果变量与其他处理组未接受处理时的潜在结果变量变化趋势相同。这意味着对于任何组别  $g$  和时间  $t$ ，有  $E \left[ Y_{i,g,t}^1 - Y_{i,g,t-1}^1 \mid E_g = -\infty \right] = E \left[ Y_{i,g,t}^0 - Y_{i,g,t-1}^0 \mid E_g \neq -\infty \right]$ 。验证这一假设是否成立，可以通过检查始终受处理组的结果变量与控制组的结果变量是否展现出相似的趋势来判断，具体过程可参考文档 *ATG* 是否适合作为控制组 *do*。此外，培根分解法可以分解出常系数 TWFE 回归模型中始终受处理组作为控制组时对应的处理效应和权重，如果分解结果体现出较强的“负权重”问题，则说明始终接受处理组并不适合作为控制组。

和对应动态模型“事件分析法”的距离处理开始  $l$  期的累积处理效应  $ATT(l)$ :

$$ATT^{CS}(l) = \sum_g w_g ATT^{CS}(g, E_g + l).$$

其中,  $w_{g,t}$  和  $w_g$  可以根据研究需求, 对所有  $(g, t)$  赋予相同的权重或根据样本量  $N_{g,t}$  来决定权重。对于  $l < 0$ ,  $ATT^{CS}(l)$  的估计量可以用于检验事前趋势。这一估计量可以通过 STATA 命令 *csdid* 来实现。

Callaway & Sant'Anna (2021) 的估计量是一种非参数估计方法, 而 Sun & Abraham (2021) 提出的 Interaction-weighted (下文简称 IW) 估计量则是基于回归模型的参数估计方法。这一方法考虑到不同组之间的处理效应异质性, 直接修正回归模型 (10) 中的模型误设, 估计如下异质效应 TWFE 回归模型:

$$Y_{i,g,t} = \alpha_i + \lambda_t + \sum_e \sum_{l \neq -1} \beta_e^l (\mathbf{1}\{E_g = e\} \times D_{g,t}^l) + \epsilon_{i,g,t} \quad (14)$$

在该模型中, 组别与处理变量的交互项  $\mathbf{1}\{E_g = e\} \times D_{g,t}^l$  取代了单一的处理变量  $D_{g,t}^l$ , 在满足平行趋势假设的前提下,  $\beta_e^l$  可以识别  $E_g = e$  的组  $g$  在处理开始之后  $l$  期的平均处理效应  $ATT(e, l)$ 。为了得到与回归模型 (10) 中  $\beta^l$  同样代表在处理之后  $l$  期 ( $l \geq 0$ ) 的平均处理效应的估计量, 他们提出根据处理之后  $l$  期时各组的样本量进行加权平均, 得到 IW 估计量  $ATT^{IW}(l)$ 。<sup>31</sup> 同样的, 对于  $l < 0$  的  $\beta^l$  的 IW 估计量可以用来检验事前趋势。

在 IW 估计中, 需要删除整个样本期均接受处理的个体以及所有个体都接受处理的样本期, 仅保留整个样本期均未接受处理的样本或最后接受处理的样本作为控制组。当由前者作为控制组时, IW 估计量  $ATT^{IW}(l)$  与 Callaway & Sant'Anna (2021) 的估计量  $ATT^{CS}(l)$  等价。<sup>32</sup> IW 估计量可以通过 STATA 命令 *eventstudyinteract* 来实现。

### 3.4.2 插补 (Imputation) 估计量

以上两种估计量都考虑了处理效应在处理开始时间和持续时间两个维度上的异质性, 分别采用了非参数估计和参数估计的思路。Borusyak et al. (2023) 则从半参数估计的思路出发, 提出了插补估计量。该方法通过单独对未接受处理的潜在结果  $Y_{i,t}^0$  进行建模, 规避了考虑异质性处理效应时模型自由度下降的问题。插补法还允许控制变量以灵活形式影响结果变量。

该方法包括三个步骤:

- 步骤 1 使用所有未接受处理的观测值估计一个 TWFE 回归模型:  $Y_{i,t}^0 = \alpha_i + \lambda_t + \epsilon_{i,t}$ , 作为反事实估计模型;
- 步骤 2 根据第一步估计的回归结果, 计算每个接受处理的个体在每个处理时间点上未接受处理的潜在结果  $\widehat{Y}_{i,t}^0$ , 并计算实际结果与潜在结果之间的差异  $Y_{i,t} - \widehat{Y}_{i,t}^0$ , 即每个  $(i, t)$  的处理效应;

<sup>31</sup>Callaway & Sant'Anna (2021) 中  $ATT^{CS}(g, t)$  的其他汇总方式, 对于  $\beta_e^l$  同样适用。

<sup>32</sup>这仅限于不存在控制变量的情况, 根据后文讨论二者对于控制变量的处理有所不同。

步骤 3 根据实证需求选择适当的权重，将第二步得到的处理效应进行汇总。

这一估计量可以通过 STATA 命令 *did\_imputation* 来实现。<sup>33</sup>

#### 4 第二类模型误设：非二元处理变量情形

我国在实施新政策时，地方政府的执行标准往往不同。例如，20 世纪 90 年代以来，部分省份逐步开征地方教育附加，但征收资金在省与地方之间的分配比例不同 (高跃光、范子英和冯晨, 2024)。类似地，某些看似“一刀切”的重大事件，例如中国加入 WTO 后进口关税的下降，对各行业的影响程度也不一样。此时个体的处理状态不再是非黑即白的二元变量，而是表征处理强度差异的多值离散变量或者连续变量，即“非二元处理变量”情形。根据处理时点是否一致，可进一步划分为单时点非二元处理变量与多时点非二元处理变量。现有文献的通常做法是将非二元处理变量  $D_{g,t}$  直接带入模型 (1) 中，即运用 TWFE 回归的推广 2，并将系数  $\beta$  解释为每单位处理强度变动对结果变量的影响。<sup>34</sup> 在推广 2 之下，未受处理组不再是唯一控制组，甚至可能不存在未受处理组；处理强度较低的组会被用作控制组。因此，仅依赖未受处理组潜在结果的平行趋势假设，难以进行有效的因果推断，更无法得到“无伤大雅”的处理效应的估计。

例如，中国加入 WTO 后，各行业的进口关税降幅有大有小，是典型的连续变量。理论上，进口关税的下降会带来进口竞争，促使本地行业生产率的提升 (Brandt et al., 2017)。研究者可以运用 TWFE 回归的推广 2，检验进口关税降幅对行业生产率的提升作用。此时，TWFE 回归会将进口关税降幅较小的行业作为控制组，并将进口关税降幅较大的行业作为处理组。只要单位降幅对行业生产率的影响不是常数，TWFE 回归就存在模型误设。如果在进口关税降幅较小的行业 (控制组)，每单位关税降幅对行业生产率的提升作用较强，就可能会出现 TWFE 回归系数与 ATT 符号相反的问题。从图 1 可以看到，推广 2 “非二元处理变量”情形占 TWFE 回归应用的 30.66%，占比甚至超过了推广 1 的 23%。因此，在模型误设统一框架下理解 TWFE 回归的推广 2 具有重要意义。

---

<sup>33</sup>需要注意的是，插补估计量在检验事前趋势时，并不是以处理前 1 期 (-1 期) 为基期进行估计。具体地，当要检验处理之前 1 期内的事前趋势时，基期是 1 期之前的所有时期。此外，同样使用插补估计量思路来估计 DID 方法的处理效应的还有 Liu et al. (2024)、Wooldridge (2021) 和 Gardner (2022)。

<sup>34</sup>估计非二元处理变量动态效应的一个常见做法是将回归模型 (10) 中的  $D_{g,t}^l = I[t - E_g = l]$  替换为  $I^s D_{g,t}^l$ ，其中  $I^s$  为组  $g$  的处理强度，这种做法适用于处理强度不随时间变化的情况。当处理强度随时间变化时，文献中常用的方法包括分布滞后回归 (Distributed-lag Regression) 和局部投影法 (Local-projection Method) (Jordà, 2005)。然而，de Chaisemartin & D'Haultfœuille (2024a) 发现在存在处理效应异质性时，这些方法可能会遇到“负权重”问题。为此，Dube et al. (2023) 对局部投影法进行了修正，并提出了适用于双重差分事件分析法的局部投影法估计量 (LP-DID)，具有处理效应异质性稳健的特点。这些讨论超出了本文的分析范畴，因此不在此详细展开。

#### 4.1 处理效应异质性和“负权重”问题

在处理效应异质性的背景下，使用常系数 TWFE 回归模型 (1) 来估计非二元处理变量的处理效应时，存在所谓的“负权重”问题。这个问题在本质上与二元交错处理变量中遇到的问题类似，根源在于模型误设引入了一种“禁止的对比”：将接受不同处理强度的组别进行比较，其中处理强度较高的组作为处理组，而处理强度较低的组作为控制组。为了更清晰地阐述这一问题，笔者将首先通过一个处理强度不同的双组双期案例进行展示，随后通过模型误设框架来说明。

如表 3 所示，样本被分为两组 ( $g = m$  和  $g = l$ )，包含两个样本期 ( $t = 1$  和  $t = 2$ )。两组从  $t = 2$  期开始接受不同强度的处理，分别为  $d_m$  和  $d_l$ ，且满足  $d_m > d_l > 0$ 。

表 3: 非二元处理变量的双组双期示例

	$t = 1$	$t = 2$
$g = m$	0	$d_m$
$g = l$	0	$d_l$

为分析方便，笔者进一步假设潜在结果变量  $Y_{i,g,t}^d$  与处理强度  $d$  之间存在以下线性关系：

$$E \left[ Y_{i,g,t}^d \right] = Y_{i,g,t}^0 + \delta_g d, \quad (15)$$

这意味着不同组对处理强度的敏感性各异。如果将处理强度较低的组作为控制组，可以得到的双重差分估计值为：

$$\begin{aligned} DID_d &= E[Y_{i,m,2}] - E[Y_{i,m,1}] - (E[Y_{i,l,2}] - E[Y_{i,l,1}]) \\ &= E \left[ Y_{i,m,2}^{d_m} \right] - E \left[ Y_{i,m,1}^0 \right] - \left( E \left[ Y_{i,l,2}^{d_l} \right] - E \left[ Y_{i,l,1}^0 \right] \right) \\ &= E \left[ Y_{i,m,2}^0 \right] + \delta_m d_m - E \left[ Y_{i,m,1}^0 \right] - \left( E \left[ Y_{i,l,2}^0 \right] + \delta_l d_l - E \left[ Y_{i,l,1}^0 \right] \right) \\ &= E \left[ Y_{i,m,2}^0 \right] - E \left[ Y_{i,m,1}^0 \right] - \left( E \left[ Y_{i,l,2}^0 \right] - E \left[ Y_{i,l,1}^0 \right] \right) + \delta_m d_m - \delta_l d_l \\ &= \delta_m d_m - \delta_l d_l, \end{aligned}$$

其中最后一个等式运用了平行趋势假设 (假设 1)，即  $E \left[ Y_{i,m,2}^0 \right] - E \left[ Y_{i,m,1}^0 \right] = E \left[ Y_{i,l,2}^0 \right] - E \left[ Y_{i,l,1}^0 \right]$ 。由此可见，作为控制组的  $l$  组的处理效应被赋予了“负权重”。如果  $\delta_m d_m < \delta_l d_l$ ，那么即使两组的处理效应  $\delta_m$  和  $\delta_l$  均为正，双重差分估计出的处理效应也可能为负。当  $\delta_m = \delta_l = \delta$  时， $DID_d = \delta(d_m - d_l)$ ，即不存在处理效应异质性时，双重差分估计值与处理效应的方向才始终一致。

根据以上分析，这种“负权重”问题在两组两期的情境中存在。进一步地，当模型扩展至多个时期，并且各组接受处理的时间点不一致时，类似二元交错处理变量中的“负权重”问题也会凸显出来，二者叠加使情况更加棘手。出现这一问题的本质原因是，在非二元处理变量情形中，常系数 TWFE 回归模型存在模型误设问题，引入了“禁止的对比”。接下来，笔者用模型误设的框架来进行分析。

延用等式 (15) 作为处理效应异质性的假设，那么常系数 TWFE 回归模型 (1) 的模型误设是遗漏了一系列交互项  $\sum_{(g,t) \in \Omega} \sum_{d=d_1}^{d_m} \beta_{g,t}^d I^d \mathbf{1}\{D_{g,t} > 0\}$ 。  $I^d$  为接受处理强度为  $d$  的虚拟变量， $d$  的取值范围为  $\{d_1, d_2, \dots, d_m\}$ ；而  $\mathbf{1}\{D_{g,t} > 0\}$  为接受任何非零处理强度的虚拟变量。因此，理论上应估计如下异质效应 TWFE 回归模型：

$$Y_{i,g,t} = \alpha_g + \lambda_t + \sum_{(g,t) \in \Omega} \sum_{d=d_1}^{d_m} \beta_{g,t}^d I^d \mathbf{1}\{D_{g,t} > 0\} + e_{i,g,t} \quad (16)$$

根据 OVB 公式和 FWL 定理，可以得到如下分解公式：

$$\beta = \sum_{(g,t) \in \Omega} \sum_{d=d_1}^{d_m} \omega_{g,t}^d \beta_{g,t}^d,$$

其中  $\omega_{g,t}^d$  的表达式为

$$\omega_{g,t}^d = \frac{N_{g,t}^d \tilde{D}_{g,t}}{\sum_{(g,t) \in \Omega_2} N_{g,t} D_{g,t} \tilde{D}_{g,t}} \mathbf{1}\{D_{g,t} > 0\},$$

$\tilde{D}_{g,t} = D_{g,t} - \bar{D}_{\cdot t} - \bar{D}_g + \bar{D}$ ， $\Omega_2 = \{(g,t) | D_{g,t} > 0\}$  是所有  $D_{g,t} > 0$  的观测值的集合； $N_{g,t}^d$  是  $D_{g,t} = d$  的样本数。<sup>35</sup>

如果满足平行趋势假设 (假设1)，即  $E[Y_{i,g,t}^0 - Y_{i,g,t-1}^0 | I^d] = E[Y_{i,g,t}^0 - Y_{i,g,t-1}^0]$ ，且  $E[Y_{i,g,t}^0 - Y_{i,g,t-1}^0]$  不随  $g$  改变，那么模型 (16) 中  $\beta_{g,t}^d$  将能识别接受处理强度为  $d$  的处理组处理效应  $E[Y_{i,g,t}^d - Y_{i,g,t}^0 | I^d = 1]$ 。而模型 (1) 中  $\beta$  的估计量则是各组平均处理效应的加权平均。虽然这些权重之和为 1，但是由于  $\tilde{D}_{g,t} = D_{g,t} - \bar{D}_{\cdot t} - \bar{D}_g + \bar{D}$  可能为负数，因此权重可能为负，从而导致“负权重”问题。<sup>36</sup>

此外，当观测期大于 2 时，相较于二元交错处理变量，非二元处理变量还存在处理轨迹多样化的特点。对于  $T$  期样本数据，前者最多只有  $T+1$  个处理轨迹，而对于具有  $J$  种可能取值的处理变量来说，至多存在  $J^T$  个处理轨迹。<sup>37</sup>理论上，只要样本中有足够处理强度变化的观测值，所有处理轨迹的处理效应都能够被估计出来。然而，在有限的样本中，估计结果可能不够精确，并且随着处理变量取值  $J$  和样本期  $T$  的增加，处理轨迹将呈指数级增长，导致估计进入维数灾难。此外，这种估计结果的实用意义不大，因此研究者可以对估计进行一定程度的简化，得到一个高度概括的处理效应估计。

<sup>35</sup>推导证明见附录B.3。

<sup>36</sup> Callaway et al. (2024a) 中也探讨了模型 (1) 中非二元处理变量系数的分解公式，构成要素 (Building Block) 包括接受处理强度为  $d$  的水平处理效应  $ATT(d|d') = E[Y_{i,g,t}^d - Y_{i,g,t}^0 | D_{g,t} = d']$  和斜率处理效应  $ACRT(d|d') = \frac{\partial ATT(l|d')}{\partial l} |_{l=d}$  等。构成要素不同，分解公式的形式也有所不同。详见 Callaway et al. (2024a) 中的定理 3.4。

<sup>37</sup>例如，当  $T=3$  时，二元交错处理变量仅有  $(0,0,0)$ ,  $(1,1,1)$ ,  $(0,1,1)$ ,  $(0,0,1)$  这四种处理轨迹，而对于  $D_{g,t} \in 1,2$ ，即具有  $J=2$  种可能取值的处理变量，就存在  $(1,1,1)$ ,  $(1,2,1)$ ,  $(1,1,2)$ ,  $(1,2,2)$ ,  $(2,1,1)$ ,  $(2,1,2)$ ,  $(2,2,1)$ ,  $(2,2,2)$  这八种处理轨迹。

## 4.2 处理效应异质性稳健的估计量

当前文献针对非二元处理变量估计的处置思路与二元交错处理变量类似，关键在于避免所谓的“禁止的对比”，寻找适合的控制组。de Chaisemartin & D’Haultfoeuille (2018, 2020, 2024a, 2024b) 在一系列文章中提出，以处理强度没有发生改变的组作为处理强度发生改变的组（处理组）的控制组，并借助处理强度是否发生变化的工具变量  $Z_{g,t} = \mathbf{I}\{D_{g,t} \neq D_{g,1}\}$  进行两步法分析：分别估计处理组在接受处理后结果变量的平均处理效应  $E\left[Y_{i,g,t}^{D_{g,t}} - Y_{i,g,t}^{D_{g,1}}\right]$  和处理强度的平均变化  $E\left[D_{g,t} - D_{g,1}\right]$ ，并基于这两者的估计值估计处理强度每单位变化带来的平均处理效应。接下来，本文将分别论述单时点非二元处理变量和多时点非二元处理变量两种情况下的具体分析方法。

### 4.2.1 单时点非二元处理变量

在单时点非二元处理变量情形中，所有组的处理时点相同，但受到的处理强度不同。如果仅考虑处理强度是否发生变化，即构成了一个简单的  $2 \times 2$  经典 DID 框架，无须考虑处理效应的异质性问题。此时用工具变量进行两步法估计有两个关键点。

首先，需要一个处理强度与处理组基期处理强度相同且保持不变的组（下文简称稳定组）作为控制组。如表 4 示例 1(a) 所示，处理组 ( $g = m$ ) 在  $t = 2$  时接受处理，使得处理强度从  $d_l$  变化为  $d_m$ 。对应的控制组是处理强度始终为  $d_l$  的  $g = l$  稳定组。

其次，必须满足一定的平行趋势假设，使得稳定组可以作为处理组的控制组。当处理组的基期处理强度为 0 时，只需要未经处理的潜在结果变量满足平行趋势假设（假设 1）即可；当基期处理强度不为 0 时，稳定组作为控制组需要满足以下形式的平行趋势假设：

假设 2 平行趋势（单时点非二元处理变量）：任意  $(g, g')$ ，如果  $D_{g,1} = D_{g',1}$ ，那么  $\forall t > 1$ ， $E\left[Y_{i,g,t}^{D_{g,t}} - Y_{i,g,t-1}^{D_{g,t}}\right] = E\left[Y_{i,g',t}^{D_{g',t}} - Y_{i,g',t-1}^{D_{g',t}}\right]$ 。

这一假设要求，基期处理强度相同的组，在处理强度保持在基期水平不变的情况下，潜在结果变量的变化趋势一致。如表 4 示例 1(a) 所示，如果满足  $E\left[Y_{i,m,2}^{d_l} - Y_{i,m,1}^{d_l}\right] = E\left[Y_{i,l,2}^{d_l} - Y_{i,l,1}^{d_l}\right]$ ，则  $g = l$  组可以作为  $g = m$  组的控制组。

工具变量两步法的估计思路如下。根据处理强度是否发生改变定义工具变量为  $Z_{g,t} = \mathbf{I}\{D_{g,t} \neq D_{g,1}\}$ 。表 4 示例 1(a) 对应的工具变量如示例 1(b) 所示， $g = m$  组在时间  $t = 2$  处理强度发生变化，因此  $Z_{m,2} = 1$ ，其他情况下  $Z_{g,t} = 0$ ，形式等价于  $2 \times 2$  经典 DID 框架。 $g = m$  组的处理强度从  $d_l$  变化至  $d_m$ ，每单位处理强度变化的平均处理效应可以通过以下估计量表示：

$$\text{DID}_m = \frac{\text{DID}_m^Y}{\text{DID}_m^D}, \quad (17)$$

表 4: 单时点非二元处理变量的双重差分方法示例

示例 1

1(a) 处理强度 $D_{g,t}$			1(b) 工具变量 $Z_{g,t}$		
	$t = 1$	$t = 2$		$t = 1$	$t = 2$
$g = m$	$d_l$	$d_m$	$g = m$	0	1
$g = l$	$d_l$	$d_l$	$g = l$	0	0

示例 2

2(a) 处理强度 $D_{g,t}$			2(b) 工具变量 $Z_{g,t}$		
	$t = 1$	$t = 2$		$t = 1$	$t = 2$
$g = h$	$d_l$	$d_h$	$g = h$	0	1
$g = m$	$d_l$	$d_m$	$g = m$	0	1
$g = l$	$d_l$	$d_l$	$g = l$	0	0

示例 3

3(a) 处理强度 $D_{g,t}$			3(b) 工具变量 $Z_{g,t}$		
	$t = 1$	$t = 2$		$t = 1$	$t = 2$
$g = m1$	$d_{l1}$	$d_{m1}$	$g = m1$	0	1
$g = l1$	$d_{l1}$	$d_{l1}$	$g = l1$	0	0
$g = m2$	$d_{l2}$	$d_{m2}$	$g = m2$	0	1
$g = l2$	$d_{l2}$	$d_{l2}$	$g = l2$	0	0

其中,  $DID_m^Y$  为通过 DID 方法估计处理组在接受处理后结果变量的平均处理效应:

$$\begin{aligned}
 DID_m^Y &= E[Y_{i,m,2}] - E[Y_{i,m,1}] - (E[Y_{i,l,2}] - E[Y_{i,l,1}]) \\
 &= E[Y_{i,m,2}^{d_m}] - E[Y_{i,m,1}^{d_l}] - (E[Y_{i,l,2}^{d_l}] - E[Y_{i,l,1}^{d_l}]) \\
 &= E[Y_{i,m,2}^{d_m}] - E[Y_{i,m,1}^{d_l}] - (E[Y_{i,m,2}^{d_l}] - E[Y_{i,m,1}^{d_l}]) \\
 &= E[Y_{i,m,2}^{d_m}] - E[Y_{i,m,2}^{d_l}]
 \end{aligned}$$

其中第三个等号利用了平行趋势假设 2。由于表 4 示例 1(b) 的形式等同于  $2 \times 2$  经典 DID,  $DID_m^Y$  可以通过结果变量  $Y_{i,g,t}$  对工具变量  $Z_{g,t}$  的 TWFE 回归得到, 这样的估计对处理效应异质性稳健。

$DID_m^D$  为通过 DID 方法估计处理组在接受处理后处理强度变量的平均处理效应:

$$\begin{aligned}
 DID_m^D &= E[D_{i,m,2}] - E[D_{i,m,1}] - (E[D_{i,l,2}] - E[D_{i,l,1}]) \\
 &= d_m - d_l - (d_l - d_l) \\
 &= d_m - d_l.
 \end{aligned}$$

该估计框架也适用于存在多个处理强度变化不同的处理组。如表 4 示例 2(a) 所示,  $g = m$  组的处理强度从  $d_l$  变化至  $d_m$ ,  $g = h$  组的处理强度从  $d_l$  变化至  $d_h$ 。根据等式 (17) 进行估计,

$DID_m^Y$  ( $DID_m^D$ ) 可以通过结果变量  $Y_{i,g,t}$  (处理变量  $D_{g,t}$ ) 对工具变量  $Z_{g,t}$  的 TWFE 回归得到。如果存在多期数据, 包括处理强度之后也发生改变的情况, 可以通过事件研究法分别估计  $DID_{m,l}^Y$  和  $DID_{m,l}^D$ 。同样地, 由于不存在多个处理时点,  $DID_{m,l}^Y$  和  $DID_{m,l}^D$  都可以通过事件研究法的 TWFE 回归模型 (10) 得到。<sup>38</sup>

进一步地, 如果存在基期强度不同的组, 如表 4 示例 3(a) 所示,  $g = m1$  组的处理强度从  $d_{l1}$  变化至  $d_{m1}$ ,  $g = m2$  组的处理强度从  $d_{l2}$  变化至  $d_{m2}$ , 存在  $g = l1$  组, 其处理强度为  $d_{l1}$  并保持不变, 可以作为  $g = m1$  组的控制组来估计  $DID_{m1}^Y$  和  $DID_{m1}^D$ ;  $g = l2$  组处理强度为  $d_{l2}$  并保持不变, 可以作为  $g = m2$  组的控制组来估计  $DID_{m2}^Y$  和  $DID_{m2}^D$ 。随后, 将  $DID_{m2}^Y$  ( $DID_{m2}^D$ ) 和  $DID_{m1}^Y$  ( $DID_{m1}^D$ ) 进行加权平均后估计等式 (17) 中的估计量。<sup>39</sup>

#### 4.2.2 多时点非二元处理变量

当非二元处理变量结合多处理时点时, 处理效应异质性导致的分析复杂性不仅体现在前述的“禁止的对比”中, 还涉及多元的处理轨迹。此时,  $Y_{i,g,t}$  取决于个体  $i$  的处理路径  $(D_{g,1}, D_{g,2}, \dots, D_{g,T})$ 。笔者对前文的符号进行调整,  $Y_{i,g,t} = Y_{i,g,t}^{D_{g,1}, D_{g,2}, \dots, D_{g,T}}$  表示组  $g$  中的个体  $i$  在  $t$  期的实际观测结果。 $Y_{i,g,t}^{d_1, d_2, \dots, d_T}$  表示当处理变量  $(D_{g,1}, D_{g,2}, \dots, D_{g,T}) = (d_1, d_2, \dots, d_T)$  时组  $g$  中的个体在  $t$  期的潜在结果。假设无预期处理效应, 则  $Y_{i,g,t}^{d_1, d_2, \dots, d_T} = Y_{i,g,t}^{d_1, d_2, \dots, d_t}$ , 即  $t$  期之后的处理状态不影响  $t$  期的结果变量取值。

de Chaisemartin & D’Haultfoeuille (2024a) 针对这一复杂情况提出了一种简化方法, 即只考虑处理变量首次发生变化的时间点来定义目标估计量。为此, 需要定义处理变量首次变化的时间点  $F_g$ , 并且与单时点处理的情形相同, 要求存在稳定组并满足平行趋势假设。具体来说, 需要存在  $(g, g')$ , 使得  $D_{g,1} = D_{g',1}$  并且  $F_g \neq F_{g'}$ , 即必须存在一个组  $g'$ , 其开始接受处理的时期与  $g$  组不同, 但初始处理变量取值与  $g$  组相同, 称为稳定组。稳定组可以作为控制组需要满足以下平行趋势假设:

假设 3 平行趋势 (多时点非二元处理变量): 任意  $(g, g')$ , 如果  $D_{g,1} = D_{g',1}$ , 那么  $\forall t > 1$

$$E \left[ Y_{i,g,t}^{(D_{g,1}, \dots, D_{g,1})} - Y_{i,g,t-1}^{(D_{g,1}, \dots, D_{g,1})} \right] = E \left[ Y_{i,g',t}^{(D_{g,1}, \dots, D_{g,1})} - Y_{i,g',t-1}^{(D_{g,1}, \dots, D_{g,1})} \right].$$

这一假设要求, 如果两组在基期的处理变量相同, 并在后续时期保持不变, 则这两组的潜在结果变化趋势应当一致。如表 5(a) 所示,  $g = h$  组的处理强度从  $t = 1$  时的  $d_l$  变化为  $t = 2$  时的  $d_{h1}$ , 处理强度在前两期保持在  $d_l$  的  $g = l$  组和  $g = m$  组均可以作为估计  $g = h$  组接受处理一期的处理效应  $ATT_{h,2} = E \left[ Y_{i,h,2}^{(d_l, d_{h1})} - Y_{i,h,2}^{(d_l, d_l)} \right]$  的控制组;  $g = h$  组的处理强度从  $t = 2$  时的  $d_{h1}$  进一步变化为  $t = 3$  时的  $d_{h2}$ , 只有处理强度始终为  $d_l$  的  $g = l$  组可以作为  $g = h$  组接受处理持续两

<sup>38</sup>这一分析思路由 de Chaisemartin & D’Haultfoeuille (2018) 首次提出, 适用于模糊 DID (Fuzzy DID) 框架的 Time-corrected Wald 估计量; 随后在 de Chaisemartin & D’Haultfoeuille (2020) 中进一步推广至多期多时点 DID 框架下的即期处理效应估计。

<sup>39</sup>这种情况下不能通过一个统一的 TWFE 回归来估计  $DID_m^Y$  和  $DID_m^D$ , 因为处理组分别有各自的控制组, 混淆会出现“负权重”问题。实践中, 这一过程可以视为多处理时点的特殊情况, 通过 4.2.2 节中介绍的 de Chaisemartin & D’Haultfoeuille (2024a) 提出的方法实现。

期的处理效应  $ATT_{h,3} = E \left[ Y_{i,h,3}^{(d_l, d_{h1}, d_{h2})} - Y_{i,h,3}^{(d_l, d_l, d_l)} \right]$  的控制组； $g = m$  组的处理强度从  $t = 2$  时的  $d_l$  变化为  $t = 3$  时的  $d_{m1}$ ，依旧只有处理强度始终为  $d_l$  的  $g = l$  组可以作为  $g = m$  组接受处理一期的处理效应  $ATT_{m,3} = E \left[ Y_{i,m,3}^{(d_l, d_l, d_{m1})} - Y_{i,m,3}^{(d_l, d_l, d_l)} \right]$  的控制组。

表 5: 多时点非二元处理变量的双重差分方法示例

	(a) 处理强度 $D_{g,t}$			(b) 工具变量 $Z_{g,t}$		
	$t = 1$	$t = 2$	$t = 3$	$t = 1$	$t = 2$	$t = 3$
$g = h$	$d_l$	$d_{h1}$	$d_{h2}$	$g = h$	0	1
$g = m$	$d_l$	$d_l$	$d_{m1}$	$g = m$	0	0
$g = l$	$d_l$	$d_l$	$d_l$	$g = l$	0	0

接下来，笔者根据等式 (17) 所示的工具变量两步法思路进行估计，定义工具变量  $Z_{g,t} = \mathbf{I}\{D_{g,t} \neq D_{g,1}\}$ ，表示组  $g$  在时间  $t$  时处理强度是否相对于基期发生改变，如表 5(b) 所示。然而，考虑到表 5(b) 中表示的是一个二元交错处理变量框架，根据前文讨论，TWFE 回归估计的结果由于处理效应异质性的存在会出现偏差。因此，等式 (17) 中的  $DID^Y$  和  $DID^D$  需要使用处理效应异质性稳健的估计量，如 Callaway & Sant'Anna (2021)，Sun & Abraham (2021) 和 Borusyak et al. (2023)。<sup>40</sup>

de Chaisemartin & D'Haultfoeuille (2024a) 的估计方法则内置了工具变量两步法的思路。在估计结果变量的处理效应  $DID^Y$  时，估计目标是在特定时间点  $F_g$  处理状态首次发生变化的组  $g$ ，在接受处理  $l$  期内累积的处理效应：

$$\delta_{g,l}^Y = E \left[ Y_{i,g,F_g-1+l} - Y_{i,g,F_g-1+l}^{(D_{g,1}, \dots, D_{g,1})} \right].$$

令  $T_g = \max_{g': D_{g',1} = D_{g,1}} F_{g'} - 1$  表示存在与组  $g$  具有相同基期处理效应且保持不变的控制组的最大期限。对于所有满足  $F_g < T_g$  以及  $l \in \{1, \dots, T_g - F_g + 1\}$  的情况， $\delta_{g,l}^Y$  可通过以下公式估计：

$$DID_{g,l}^Y = E \left[ Y_{i,g,F_g-1+l} \right] - E \left[ Y_{i,g,F_g-1} \right] - \frac{1}{N_{F_g-1+l}^g} \sum_{g': D_{g',1} = D_{g,1}, F_{g'} > F_g-1+l} \left( E \left[ Y_{i,g',F_g-1+l} \right] - E \left[ Y_{i,g',F_g-1} \right] \right),$$

其中  $N_{F_g-1+l}^g$  为直至  $F_g - 1 + l$  期间处理效应未发生变化的组数。此估计方法本质上是将具有与处理组  $g$  相同初始处理状态且直到  $F_g - 1 + l$  期未发生变化的样本作为控制组，通过双重差分法估计组  $g$  在  $l$  期内累积的处理效应。在此基础上，将逐期的处理效应  $DID_{g,l}^Y$  依据样本量权重进行加权平均，从而估计累积接受处理  $l$  期的平均处理效应：

$$\delta_l^Y = \frac{1}{N_l} \sum_{g: F_g-1+l \leq T_g} S_g \delta_{g,l}^Y$$

<sup>40</sup>由于 Callaway & Sant'Anna (2021)，Sun & Abraham (2021) 和 Borusyak et al. (2023) 等方法对应的 STATA 命令并没有内置这种工具变量两步法的思路，不能直接适用于如表 5(a) 中所示的非二元处理变量情形，需要手动生成如表 5(b) 所示的工具变量后进行估计和汇总。详情请见 do 文档非二元处理变量-多时点.do。

可以通过下式来估计：

$$DID_l^Y = \frac{1}{N_l} \sum_{g:F_g-1+l \leq T_g} S_g DID_{g,l}^Y$$

其中， $N_l$  是可以计算出  $DID_{g,l}^Y$  的组数，即在样本期内接受处理持续  $l+1$  期且存在与其基期处理强度相同并持续至  $l+1$  期的控制组的处理组样本数。 $S_g = 1 \{D_{g,F_g} > D_{g,1}\} - 1 \{D_{g,F_g} < D_{g,1}\}$ ，是一个符号指标，表示在  $F_g$  时刻处理变量增加则为 1，减少则为 -1。de Chaisemartin & D'Haultfœuille (2024a) 建议剔除处理变量先增加后减少至基期水平以下的情况，只保留  $D_{g,t} \geq D_{g,1}$  或  $D_{g,t} \leq D_{g,1}$  的情况，这样可以将  $DID_l^Y$  估计量解释为接受较高处理强度累计  $l$  期的平均处理效应。<sup>41</sup>

工具变量两步法的第二步是估计处理强度的变化  $DID_l^D$ ，可以通过将以上分析当中的结果变量  $Y_{i,g,t}$  替换为处理强度  $D_{g,t}$  得到。<sup>42</sup>在实证分析当中，笔者建议分别展示  $DID_l^Y$  和  $DID_l^D$  的估计结果以及对应的事件研究图。<sup>43</sup>如图 2 所示，图 2a 为工具变量  $Z_{g,t}$  对结果变量  $Y_{g,t}$  的动态影响，图 2b 为工具变量  $Z_{g,t}$  对处理变量  $D_{g,t}$  的动态影响。<sup>44</sup>

最后，结合前述分析，de Chaisemartin & D'Haultfœuille (2024a) 提出一个适用于全局成本-收益分析的估计量，和等式 (17) 类似。目标参数如下：

$$\delta = \frac{\sum_{g:F_g \leq T_g} \sum_{l=1}^{T_g-F_g+1} \delta_{g,l}}{\sum_{g:F_g \leq T_g} \sum_{l=1}^{T_g-F_g+1} (D_{g,F_g-1+l} - D_{g,1})}$$

这一参数可以解读为每一单位处理强度的平均处理效应，即  $\delta_l^Y$  的加权平均，具体形式为： $\delta = \sum_{l=1}^L w_l \delta_l^Y$ ，其中权重的定义为  $w_l = \frac{N_l}{\sum_{g:F_g \leq T_g} \sum_{l'=1}^{T_g-F_g+1} (D_{g,F_g-1+l'} - D_{g,1})}$ 。

<sup>41</sup>当  $l = 1$  时， $DID_l^Y$  等价于 de Chaisemartin & D'Haultfœuille (2020) 中的估计量  $DID_m$ 。

de Chaisemartin & D'Haultfœuille (2024a) 中还提出了检验平行趋势假设的安慰剂估计量 (Placebo Estimator)，即  $DID_{-l}$ ：对于任意  $g: 3 \leq F_g \leq T_g$  并且  $l \in \{1, \dots, \min(T_g - F_g + 1, F_g - 2)\}$ ，

$$DID_{g,l}^{Y,pl} = E[Y_{i,g,F_g-1-l}] - E[Y_{i,g,F_g-1}] - \frac{1}{N_{F_g-1+l}^g} \sum_{g': D_{g',1}=D_{g,1}, F_{g'} > F_g-1+l} (E[Y_{i,g',F_g-1-l}] - E[Y_{i,g',F_g-1}])$$

在此基础上， $DID_l^{Y,pl}$  的估计方法与  $DID_l^Y$  类似，结合这两者，可以绘制出事件分析法的动态效应图。

$DID_l^Y$  可以直接应用于二元交错处理变量的估计，解释为累积接受处理  $l$  期的处理效应。当不存在控制变量时，这一估计量与 Callaway & Sant'Anna (2021) 中使用尚未接受处理组作为控制组来估计的  $ATT_l^{CS}$  一致。

<sup>42</sup>de Chaisemartin & D'Haultfœuille (2024a) 提出也可以通过  $\delta_{g,l}^D = \sum_{k=1}^l |D_{g,F_g-1+k} - D_{g,1}|$  ( $\delta_{g,l}^D$  的均值) 来对  $DID_{g,l}^Y$  ( $DID_l^Y$ ) 进行标准化，得到  $DID_{g,l}^{Yn}$  ( $DID_l^{Yn}$ )。 $\delta_{g,l}^D$  表示的是产生  $\delta_{g,l}$  的从  $F_g$  到  $F_g-1+l$  的处理强度增量的汇总。相比于  $DID_l^Y$  是一种基于样本量的加权平均，标准化的  $DID_l^{Yn}$ ，是基于处理强度增量的加权平均。本文附录 B.5 对这些估计量进行了详细的总结和解释。这一估计量可用于检验当前和滞后的处理是否对结果变量有相同的影响。

<sup>43</sup>Callaway et al. (2024b) 提出另外一种思路：将处理组样本根据处理强度进行分组，分别展示不同处理强度的处理组的  $DID_l^Y$ 。

<sup>44</sup>图 2 数据来源于 Favara & Imbs (2015)，该文处理变量是取值范围为 0-4 的非二元处理变量，详细背景可见附录 A。

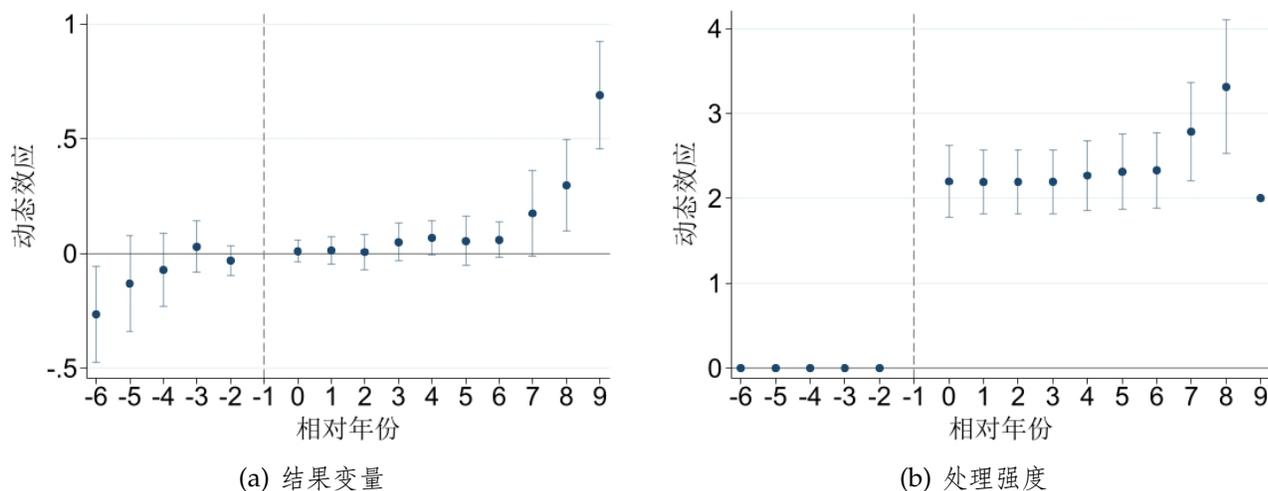


图 2:  $DID_t^Y$  和  $DID_t^D$  的事件研究图

这一参数可以通过以下估计量进行估计：

$$\hat{\delta} = \frac{\sum_{g:F_g \leq T_g} \sum_{l=1}^{T_g - F_g + 1} DID_{g,l}}{\sum_{g:F_g \leq T_g} \sum_{l=1}^{T_g - F_g + 1} (D_{g,F_g - 1 + l} - D_{g,1})}$$

de Chaisemartin & D’Haultfoeuille (2018, 2020, 2024a) 的估计量可通过 STATA 命令 `did_multiptegt_dyn` 实现。<sup>45</sup>

## 5 第三类模型误设：协变量误用

在我国“试点先行，逐步推广”的政策探索中，试点地区的选择通常与地理位置、社会经济发展水平等因素密切相关。例如，在 Gruber et al. (2023) 评估“新农合”健康效果的案例中，优先开展试点的地区往往财政实力较强。即使未实施“新农合”，这些财政实力较强的地区也可能通过增加医疗卫生投入，更好地改善居民的健康状况。也就是说，由于处理组的分配并非随机，处理组与控制组之间的特征差异可能会影响未处理时潜在结果的变化趋势，从而导致平行趋势假设（假设1）不成立。现有文献中常见的做法是直接在 TWFE 回归模型中加入控制变量，例如各地区的医疗卫生财政支出，即运用 TWFE 回归的推广3。然而，如果控制变量直接受到处理变量的影响，或控制变量对结果变量的影响随时间发生变化，TWFE 回归的推广3就存在模型误设，无法估计出 ATT 的合理近似。

以各类环保政策的环保效果评估为例。通常来说，优先进入试点名单的城市通常碳排放较高，同时经济较为发达，拥有更多资源和能力投资环保基础设施。因此，即使未被纳入试点名单，这些城市的空气污染水平也可能因经济转型和技术进步而显著下降。鉴于此，研究者往往在 TWFE

<sup>45</sup>此外，实证研究当中还有一种常见的情形是同一个体存在多次进入或退出。用回归模型 (1) 来估计这种情形的处理效应同样存在模型误设，其原因与二元交错处理变量情形类似，是因为把较早接受处理组作为控制组进行估计。这一情形适用 4.2.2 节中多时点非二元处理变量的估计思路，详见 de Chaisemartin & D’Haultfoeuille (2024a) 的研究设计 2。

回归中加入城市特征变量，如产业结构和经济发展水平等。然而，这两种情况都会导致 TWFE 回归模型的误设，从而影响处理效应的估计。原因在于：第一，环保政策可能加速试点城市的产业结构转型，这使得产业结构是“坏控制变量 (Bad Control)”，不能在回归中直接控制。第二，经济发展水平对空气污染的影响可能随时间变化，呈现“倒 U 型”的环境库兹涅茨曲线，不适用于控制变量具有常系数的模型。协变量的模型误设问题在实证研究中普遍存在，尤其是在中国经济近几十年高速发展与结构转型的现实背景之下（张征宇、林丽花、曹思力和周亚虹，2024）。因此，从模型误设的统一框架出发理解加入协变量的 TWFE 回归，对于中国经济研究具有重要的现实意义。

## 5.1 时变的控制变量

在推广 3 的情形下，研究者通常会直接在 TWFE 回归中加入时变控制变量  $\mathbf{X}_{i,t}$ 。为简化讨论，假设  $g = i$ ，扩展的 TWFE 回归模型 (6) 如下：

$$Y_{i,t} = \beta D_{i,t} + \mathbf{X}'_{i,t} \delta + \alpha_i + \lambda_t + \epsilon_{i,t}.$$

这一做法延续了回归分析的传统思路，因其简单易行而被应用研究者广泛采用。de Chaisemartin & D'Haultfœuille (2020) 指出，在 TWFE 回归中直接控制  $\mathbf{X}_{i,t}$  有两个隐含假设。第一， $\mathbf{X}_{i,t}$  是外生变量。第二， $\mathbf{X}_{i,t}$  对解释变量的影响没有异质性。

细心的应用研究者注意到了  $\mathbf{X}_{i,t}$  外生的隐含假设，尤其是  $\mathbf{X}_{i,t}$  不能受到处理变量  $D_{g,t}$  的直接影响。否则， $\mathbf{X}_{i,t}$  就会成为所谓的“坏控制变量”，这是回归分析的经典问题 (Angrist & Pischke, 2009)。Caetano et al. (2024) 在理论上详细分析了 DID 方法中的“坏控制变量”问题。应用层面，在选择控制变量时，应结合现实背景和经济理论，避免使用存在明显内生性的控制变量。

与应用研究者不同，理论计量经济学家在研究 DID 方法时很少讨论时变控制变量  $\mathbf{X}_{i,t}$ 。也许是出于对控制变量内生性的敏感，理论计量经济学家聚焦于非时变控制变量  $\mathbf{X}_i$ ，以非时变控制变量来界定处理组与控制组潜在结果变量的趋势差异，做出了条件平行趋势假设：

假设 4 条件平行趋势：  $E \left[ Y_{i,t}^0 - Y_{i,t-1}^0 \mid \mathbf{X}_i \right]$  不随组别  $i$  变化。

基于条件平行趋势假设(4)，理论计量经济学家开发了结果变量回归调整法 (Heckman et al., 1997, 1998)、逆概率加权法 (Abadie, 2005) 以及双重稳健估计量 (Sant'Anna & Zhao, 2020) 等方法。Callaway & Sant'Anna (2021) 进一步将这些方法运用于二元交错处理变量的情形。在条件平行趋势假设下，回归模型应当设定为：

$$Y_{i,t} = \beta D_{i,t} + \mathbf{X}'_i \delta_t + \alpha_i + \lambda_t + \epsilon_{i,t}.$$

此时，非时变控制变量  $\mathbf{X}_i$  前的系数向量  $\delta_t$  必须随时间变化，否则非时变控制变量  $\mathbf{X}_i$  会被固定效应  $\alpha_i$  吸收。与  $\mathbf{X}_i$  不同，时变控制变量  $\mathbf{X}_{i,t}$  前的系数可以是非时变的，只要在 TWFE 回归代码中直接纳入  $\mathbf{X}_{i,t}$  的水平项即可完成估计。无意之间，应用研究者假设了  $\mathbf{X}_{i,t}$  前系数不随时变，这就是笔者下面讨论的第二个隐含假设。

假设之二， $\mathbf{X}_{i,t}$  对潜在结果变量  $E\left[Y_{i,t}^0 \mid \mathbf{X}_{i,t}\right]$  的影响不随时间而变化，即参数  $\delta$  是常数向量。相较于控制变量外生性的假设，控制变量的常系数假设长久以来被忽视。近期研究发现，即使在  $2 \times 2$  经典 DID 的情形下，控制变量的常系数假设也会带来 TWFE 回归的估计偏误 (Lin & Zhang, 2022; Caetano et al, 2022; Caetano & Callaway, 2023; 张征宇等, 2024)。<sup>46</sup> Lin & Zhang (2022) 进一步研究了事件研究法的动态模型 (10) 中引入  $\mathbf{X}_{i,t}$  的情形，发现包含  $\mathbf{X}_{i,t}$  的影响随时变会导致  $\beta^l$  的估计量有偏。张征宇等 (2024) 强调了控制变量的常系数假设往往不适用于中国经济研究，他们在文中第 203 页写道：“研究者往往使用时间跨度较长的省级、地级市和县级面板数据，同时将地区特征  $\mathbf{X}'_{i,t}$  作为控制变量。然而，由于中国经济在过去几十年间保持长期快速增长的态势，这些地区特征（诸如产业结构、人口特征、劳动力受教育水平）对地区经济发展的贡献可能存在阶段性变革”。以上洞见为中国经济研究的模型设定起到了重要的指引作用。

基于以上讨论，常系数 TWFE 回归模型 (6) 存在针对控制变量的模型误设问题。在实证研究中，为了允许控制变量的系数随时间变化，一种比较常见的有效做法是，在回归模型中加入非时变控制变量与时间虚拟变量的交互项  $\mathbf{X}'_i \delta_t$ ，构建以下模型：

$$Y_{i,t} = \beta D_{i,t} + \mathbf{X}'_{i,t} \delta + \mathbf{X}'_i \delta_t + \alpha_i + \lambda_t + \epsilon_{i,t}. \quad (18)$$

例如，宋弘、孙雅洁和陈登科 (2019) 在研究“低碳城市”建设试点政策的空气污染治理效应时，考虑到试点和非试点城市的特征差异对城市环境可能具有时变的影响，在回归中加入了城市初始产业结构等因素与时间趋势的交互项。Gruber et al. (2023) 在评估我国“新农合”对居民健康的影响时，控制了各县农业户籍人口比重等因素与时间虚拟变量的交互项。Guo et al. (2024) 在研究出生队列规模对我国国内人口迁移的推挤和拉引效应时，控制了各地区初始人口规模等因素与时间虚拟变量的交互项。<sup>47</sup>

Caetano & Callaway (2023) 和张征宇等 (2024) 进一步提出， $\mathbf{X}_{i,t}$  对结果变量的影响不限于其当期取值，而有可能取决于  $\mathbf{X}_{i,t}$  的整条变动路径。此时，可以将时变控制变量关键期的取值单独提出，作为非时变控制变量加入模型。例如，Caetano & Callaway (2023) 考虑到潜在结果的变化可能与非时变控制变量、时变控制变量的基期取值及其相对于基期的变化相关，提出通过回归调整法和逆概率加权法同时控制这些变量。张征宇等 (2024) 提出一个基于 OLS 的纠偏估计量，将样本分为多个双期子样本，并对每个子样本进行双重差分估计。每个子样本包含控制组和处理组在  $t = 1$  期和  $t = e + l$  期（即距离处理组在  $e$  期初次接受处理  $l$  期）的观测值，同时控制  $\mathbf{X}_{i,1}$  和  $\mathbf{X}_{i,e+l}$ 。这一估计量能够有效避免政策效应异质性和控制变量的干扰。他们将这一估计量应用到国家高新技术开发区和个税减免的政策效果评估中，发现并纠正了已有文献中的估计偏误，是中

<sup>46</sup>Caetano & Callaway (2023) 推导出模型 (6) 中  $\beta$  的估计量的分解公式，发现可能存在与二元处理变量情况下类似的“负权重”问题和“反权重”问题 (Słoczyński, 2022)。其中，“反权重”问题指的是处理组中那些控制变量取值与控制组有较大差异的成员，其处理效应会被赋予更高的权重。

<sup>47</sup>模型 (18) 的思路同样适用于 de Chaisemartin & D'Haultfoeuille (2020, 2024a)、Sun & Abraham (2021) 以及 Borusyak et al. (2023) 提出的异质性稳健估计量。在 de Chaisemartin & D'Haultfoeuille (2020, 2024a) 和 Sun & Abraham (2021) 提出的估计方法中，可以先将  $Y_{i,g,t}$  回归于  $\mathbf{X}'_i \delta_t$  等，而后取回归残差来作为结果变量进行估计。在 Borusyak et al. (2023) 的估计量中，可以将  $\mathbf{X}'_i \delta_t$  等纳入步骤 1 的反事实估计模型中，随后在步骤 2 应用于反事实结果的估计中。基于上述新估计量添加非时变控制变量与时间虚拟变量交互项的相关代码可见 do 文档二元交错处理变量.do。

国经济实证研究方法创新的典型案例。

## 5.2 多个处理变量

在模型 (1) 中引入控制变量的一个特殊情况是存在多个处理变量。这种情况常见于评估具有相同政策目标的组合政策，研究者会通过以下回归模型来估计不同政策的影响：

$$Y_{i,g,t} = \beta_1 D_{g,t}^1 + \beta D_{g,t}^{-1} + \alpha_g + \lambda_t + \epsilon_{i,g,t}, \quad (19)$$

其中  $D_{g,t}^{-1}$  代表除  $D_{g,t}^1$  外的其他处理变量。

然而，de Chaisemartin & D’Haultfœuille (2023) 发现， $\beta_1$  的估计值不仅包含了主处理变量  $D_{g,t}^1$  的处理效应的加权平均值，还包括了副处理变量  $D_{g,t}^{-1}$  的处理效应的加权平均值。这种估计误差出现的主要原因是存在两类“禁止的对比”。接下来，笔者用一个有两个处理变量 ( $D_{g,t}^1$  和  $D_{g,t}^2$ ) 的  $2 \times 2$  经典 DID 来举例。

表 6: 多个处理变量的双重差分方法示例

	(a) 第一类“禁止的对比”		(b) 第二类“禁止的对比”		
	$t = 1$	$t = 2$	$t = 1$	$t = 2$	
$g = 1$	(0, 0)	(1, 1)	$g = 1$	(0, 0)	(1, 0)
$g = 2$	(0, 0)	(0, 1)	$g = 2$	(0, 1)	(0, 1)

表 6 展示了各组在不同时间点接受两个处理变量的情况。假设研究目标是估计  $D_{g,t}^1$  的处理效应，即  $E \left[ Y_{i,g,t}^{(1, D_{g,t}^2)} \right] - E \left[ Y_{i,g,t}^{(0, D_{g,t}^2)} \right]$ 。如果只考虑处理变量  $D_{g,t}^1$ ，那么表 6(a) 和表 6(b) 都符合  $2 \times 2$  经典 DID 的形式：在处理组 ( $g = 1$ ) 中， $D_{g,t}^1$  从 0 变为 1，而在控制组 ( $g = 2$ ) 中， $D_{g,t}^1$  始终为 0。但是，当同时考虑  $D_{g,t}^2$  时，表 6(a) 和表 6(b) 揭示了两种“禁止的对比”。

表 6(a) 所示的处理组  $g = 1$  和控制组  $g = 2$ ，均在  $t = 2$  时同时接受了  $D_{g,t}^2$  的处理。如果  $D_{g,t}^2$  的处理效应在两个组间没有异质性，即  $E \left[ Y_{i,g,t}^{(0,1)} \right] - E \left[ Y_{i,g,t}^{(0,0)} \right]$  不随  $g$  变化，那么组  $g = 1$  接受的  $D_{g,t}^2$  的处理效应可以通过双重差分方法与组  $g = 2$  的  $D_{g,t}^2$  的处理效应相抵消。否则， $D_{g,t}^2$  的处理效应则会对  $D_{g,t}^1$  的处理效应估计造成干扰。这是第一类“禁止的对比”。

表 6(b) 所示的处理组  $g = 1$  和控制组  $g = 2$  在  $t = 1$  和  $t = 2$  均受到  $D_{g,t}^2$  的处理。如果  $D_{g,t}^2$  的处理效应在时间上没有异质性，即组  $g = 2$  的  $E \left[ Y_{i,g,t}^{(0,1)} \right] - E \left[ Y_{i,g,t}^{(0,0)} \right]$  在  $t = 1$  和  $t = 2$  都相同，那么双重差分方法可以将组  $g = 2$  的  $D_{g,t}^2$  在  $t = 1$  和  $t = 2$  的处理效应相互抵消。否则， $D_{g,t}^2$  的处理效应也会干扰  $D_{g,t}^1$  的处理效应估计。这是第二类“禁止的对比”。<sup>48</sup>

这一发现与 Sun & Abraham (2021) 在回归模型 (10) 的分析中呈现的观点具有相关性。回归模型 (10) 中的多个处理变量代表了距事件发生  $l$  期的不同阶段。然而，与 Sun & Abraham (2021) 所发现的来自非  $l$  期处理效应的权重总和为 0 的结论不同，de Chaisemartin & D’Haultfœuille (2023) 指出，只有在特定情形下， $D_{g,t}^{-1}$  的处理效应的权重总和才为零：这种情况仅出现在只有两

<sup>48</sup>以上分析基于两个处理变量均为 0 时的平行趋势假设： $E \left[ Y_{i,g,t}^{(0,0)} - Y_{i,g,t-1}^{(0,0)} \mid g = 1 \right] = E \left[ Y_{i,g,t}^{(0,0)} - Y_{i,g,t-1}^{(0,0)} \mid g = 2 \right]$ 。

个处理变量，或这些处理变量是相互排斥的情况下。若  $D_{gt}^{-1}$  的处理效应权重总和不为 0，即使  $D_{gt}^{-1}$  的处理效应不存在异质性，使用回归模型 (19) 来估计  $D_{g,t}^1$  的处理效应仍可能受到其他处理变量的影响。

de Chaisemartin & D’Haultfœuille (2023) 提出了一种适用于多个处理变量的处理效应异质性稳健估计方法。该方法的核心思想是，在估计  $g = 1$  组从  $D_{g=1,t-1}^1 = 0$  变更为  $D_{g=1,t}^1 = 1$  的处理效应时，只选择那些满足条件  $D_{g',t-1}^{-1} = D_{g',t}^{-1} = D_{g=1,t-1}^{-1}$  的组  $g'$  作为控制组，即在基期所有变量的取值均与处理组相同且始终保持不变的组。这种估计方法建立在以下平行趋势假设上：

假设 5 平行趋势（多个处理变量）：对于  $t \geq 2$ ， $E \left[ Y_{i,g,t}^d - Y_{i,g,t-1}^d \mid D_{g,t-1} = d \right]$  不随  $g$  发生改变，其中  $D_{g,t-1} = (D_{g,t}, D_{g,t}^{-1})$ 。

这一假设要求，如果两组在  $t-1$  期所有处理变量相同且保持不变，那么这两组的潜在结果从  $t-1$  期到  $t$  期的变化趋势也将保持一致。

## 6 结语

现有文献针对 TWFE 回归三类推广下的模型误设问题，开发了一系列新的估计方法。这些新方法背后的分析思路是一致的：既然 TWFE 回归的估计偏误来源于模型误设，那就在模型设定上充分考虑处理效应的异质性，通过引入更强的识别假设，直接估计包含异质性处理效应的复杂模型。正如 de Chaisemartin and D’Haultfœuille (2022) 所述，某些新方法不仅估计效率低，对数据结构要求高，还做出了一些特殊的假设，在应用研究中不一定适用。以非二元处理变量为例，为了估计任意处理轨迹所对应的处理效应，de Chaisemartin and D’Haultfœuille (2021) 提出可以假设跨期处理效应不随时间改变且线性可加。<sup>49</sup>这类假设暗含了对研究对象行为模式的限制，而这些限制又缺乏明确的理论基础。潜在的异质性维度越多，就需要更强的假设、更丰富的数据，以及更复杂的模型。总的来说，现有的解决方案通常基于统计意义上的因果推断，旨在准确识别具有异质性的子效应，并通过某种形式的加总和平均来得到处理效应。这些方法依赖于先验设定的异质性处理效应并进行加权平均。然而，这种研究方法缺乏明确的经济理论基础。通过这种方法得到的处理效应往往与具体的经济学理论不相对应，因此难以为政策制定提供科学依据。

在中国经济学研究中，笔者主张经验分析应当依托明确的经济理论。正如 Wolpin (2013) 所指出的，“无理论 (atheoretical)”的经验分析往往缺乏明确的经济学问题定义。即使是政策评估类的因果推断研究，也应当明确定义政策效应背后的经济学理论，以便通过理论的支持增强研究结论的外部有效性 (Heckman, 2010)，从而为经济政策的制定提供更加科学的基础。经验分析与经济学理论之间存在相辅相成的关系：(1) 基于中国经济现实背景的理论分析能够指导经验分析模型的构建，避免研究者陷入“起手就跑回归”的窠臼；(2) 经验分析可用于检验理论模型的有效性；(3) 经验分析所提供的典型事实能够为新理论模型的构建奠定基础；(4) 经验分析还可量化理论模型参数，从而进行反事实模拟，为政策制定提供科学依据。近年来，国际学术界兴

<sup>49</sup>这一假设可以表示为  $Y_{i,g,t}(d) = Y_{i,g,t}(0) + \sum_{l=0}^{t-1} \alpha_{g,l} d_{t-l}$ ， $\alpha_{g,l}$  表示  $l$  期之前的处理强度  $d_{t-l}$  对当前的结果变量  $Y_{i,g,t}$  的影响。也许是意识到该假设不切合实际且难以验证，该文章正式发表的版本 (de Chaisemartin & D’Haultfœuille, 2024a) 已删除基于这一假设进行估计的讨论。

起了一种将因果推断与结构模型相结合的新研究范式 (Todd & Wolpin, 2023)。这一范式吸收了 Angrist 与 Pischke (2009) “简化式” 因果推断方法的核心，将社会经济现实、经验分析、经济学理论与政策研究有机结合，对于中国的经济研究具有重要的借鉴意义。

## References

- Abadie, A. (2005). Semiparametric Difference-in-Differences Estimators. *Review of Economic Studies*.
- Angrist, J. (1998). Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants. *Econometrica*.
- Angrist, J., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Angrist, J. D. (1991). Grouped-data estimation and testing in simple labor-supply models. *Journal of Econometrics*, 47(2-3), 243–266.
- Angrist, J. D., & Pischke, J.-S. (2014). *Mastering'metrics: The path from cause to effect*. Princeton university press.
- Bilinski, A., & Hatfield, L. A. (2018). Nothing to see here? Non-inferiority approaches to parallel trends and other model assumptions. *arXiv preprint arXiv:1805.03273*.
- Blandhol, C., Bonney, J., Mogstad, M., & Torgovitsky, A. (2022). *When is tsls actually late?* (Tech. rep.). National Bureau of Economic Research.
- Borusyak, K., Jaravel, X., & Spiess, J. (2023). Revisiting Event Study Designs: Robust and Efficient Estimation. *arXiv*.
- Brandt, L., Van Biesebroeck, J., Wang, L., & Zhang, Y. (2017). WTO Accession and Performance of Chinese Manufacturing Firms. *American Economic Review*, 107(9), 2784–2820.
- Caetano, C., & Callaway, B. (2023). Difference-in-differences with time-varying covariates in the parallel trends assumption. *arXiv preprint arXiv:2202.02903*.
- Caetano, C., Callaway, B., Payne, S., & Rodrigues, H. S. (2022). Difference in differences with time-varying covariates. *arXiv preprint arXiv:2202.02903*.
- Caetano, C., Payne, S., & Sant'Anna, H. (2024). Difference-in-differences with bad controls. *Working paper*.
- Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. (2024a). *Difference-in-differences with a continuous treatment* (tech. rep.). National Bureau of Economic Research.
- Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. (2024b). Event studies with a continuous treatment. *AEA Papers and Proceedings*, 114, 601–605.
- Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230.
- Chiu, A., Lan, X., Liu, Z., & Xu, Y. (2023). What to do (and not to do) with causal panel analysis under parallel trends: Lessons from a large reanalysis study. *Available at SSRN 4490035*.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge.
- de Chaisemartin, C., Ciccia, D., & D'Haultfœuille, X. (2024). Event-study estimators and variance estimators computed by the `did_multipligt_dyn` command.
- de Chaisemartin, C., & D'Haultfœuille, X. (2018). Fuzzy Differences-in-Differences. *The Review of Economic Studies*, 85(2), 999–1028.
- de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9), 2964–2996.
- de Chaisemartin, C., & D'Haultfœuille, X. (2021). Difference-in-Differences Estimators of Intertemporal Treatment Effects. *SSRN Electronic Journal*.
- de Chaisemartin, C., & D'Haultfœuille, X. (2022). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey. *The Econometrics Journal*.
- de Chaisemartin, C., & D'Haultfœuille, X. (2023). Two-way fixed effects and differences-in-differences estimators with several treatments. *Journal of Econometrics*, 236(2), 105480.
- de Chaisemartin, C., & D'Haultfœuille, X. (2024a). Difference-in-Differences Estimators of Intertemporal Treatment Effects. *Review of Economics and Statistics*, 1–45.
- de Chaisemartin, C., & D'Haultfœuille, X. (2024b). Two-way Fixed Effects and Difference-in-Difference Estimators in Heterogeneous Adoption Designs.
- Dube, A., Girardi, D., Jordà, Ò., & Taylor, A. M. (2023). A Local Projections Approach to Difference-in-Differences Event Studies. *NBER Working Paper No. 31184*.
- Favara, G., & Imbs, J. (2015). Credit supply and the price of housing. *American economic review*, 105(3), 958–992.
- Freyaldenhoven, S., Hansen, C., & Shapiro, J. M. (2019). Pre-Event Trends in the Panel Event-Study Design. *American Economic Review*, 109(9), 3307–3338.
- Gardner, J. (2022). Two-stage differences in differences. *arXiv preprint arXiv:2207.05943*.
- Goodman-Bacon, A. (2021). Difference-in-Differences with variations in treatment timing. *Journal of Econometrics*, 225(2), 254–277.
- Gruber, J., Lin, M., & Yi, J. (2023). The largest insurance program in history: Saving one million lives per year in china. *Journal of Public Economics*, 226, 104999.
- Guo, R., Zhang, J., & Zhou, M. (2024). The Demography of the Great Migration in China. *Journal of Development Economics*, 167, 103235.
- Heckman, J. (2010). Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy. *Journal of Economic Literature*, 48(2), 356–98.
- Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica*, 66(5), 1017.
- Heckman, J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The review of economic studies*, 64(4), 605–654.

- Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American economic review*, 95(1), 161–182.
- Kahn-Lang, A., & Lang, K. (2020). The Promise and Pitfalls of Differences-in-Differences: Reflections on 16 and Pregnant and Other Applications. *Journal of Business & Economic Statistics*, 38(3), 1–14.
- Lin, L., & Zhang, Z. (2022). Interpreting the coefficients in dynamic two-way fixed effects regressions with time-varying covariates. *Economics Letters*, 216, 110604.
- Liu, L., Wang, Y., & Xu, Y. (2024). A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data. *American Journal of Political Science*, 68(1), 160–176.
- Lu, Y., & Yu, L. (2015). Trade liberalization and markup dispersion: Evidence from china’s wto accession. *American Economic Journal: Applied Economics*, 7(4), 221–253.
- Rambachan, A., & Roth, J. (2023). A More Credible Approach to Parallel Trends. *Review of Economic Studies*.
- Roth, J. (2022). Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends. *American Economic Review: Insights*, 4(3), 305–322.
- Roth, J., & Sant’Anna, P. H. C. (2023). When Is Parallel Trends Sensitive to Functional Form? *Econometrica*, 91(2), 737–747.
- Roth, J., Sant’Anna, P. H., Bilinski, A., & Poe, J. (2023). What’s trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2), 2218–2244.
- Sant’Anna, P. H., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1), 101–122.
- Słoczyński, T. (2022). Interpreting ols estimands when treatment effects are heterogeneous: Smaller groups get larger weights. *Review of Economics and Statistics*, 104(3), 501–509.
- Stevenson, B., & Wolfers, J. (2006). Bargaining in the shadow of the law: Divorce laws and family distress. *The Quarterly Journal of Economics*, 121(1), 267–288.
- Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2), 175–199.
- Todd, P. E., & Wolpin, K. I. (2023). The Best of Both Worlds: Combining Randomized Controlled Trials with Structural Modeling. *Journal of Economic Literature*, 61(1), 41–85.
- Wolpin, K. I. (2013). *The Limits of Inference without Theory*. The MIT Press.
- Wooldridge, J. (2021). Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators. *SSRN Electronic Journal*.
- 刘冲, 沙学康, & 张妍. (2022). 交错双重差分: 处理效应异质性与估计方法选择. *数量经济技术经济研究*, 39(9), 177–204.
- 宋弘, 孙雅洁, & 陈登科. (2019). 政府空气污染治理效应评估——来自中国“低碳城市”建设的经验研究. *管理世界*, (6), 95–108.

- 张子尧 & 黄炜. (2023). 事件研究法的实现, 问题和拓展. 数量经济技术经济研究, 40(9), 71-92.
- 张征宇, 林丽花, 曹思力, & 周亚虹. (2024). 双重差分设计下固定效应估计量何时可信?. 管理世界, (1), 196-211.
- 许文立. (2023). 双重差分法的最新理论进展与经验研究新趋势. 广东社会科学, (5), 51-62.
- 高跃光, 范子英, & 冯晨. (2024). 义务教育专项融资与教育投入: 基于开征地方教育附加的研究. 管理世界, (02), 72-82+123+83-87.
- 黄炜, 张子尧, & 刘安然. (2021). 从双重差分法到事件研究法. 产业经济评论, (2), 17-36.

# 在模型误设的统一框架下理解双重差分方法的最新发展—附录\*

林梦芸 徐阳 郭汝飞 易君健

## A 实例和代码

本部分基于三个实例对正文与附录提及的相关命令进行展示与梳理。<sup>1</sup>第一个实例来源于 Stevenson & Wolfers (2006)，该文基于美国各州实施单边离婚法案的时间差异构建二元交错处理变量进行分析。研究发现，单边离婚法案实施对家庭暴力、女性自杀率和被配偶谋杀具有负面影响。<sup>2</sup>第二个实例来源于 Lu & Yu (2015)，该文利用中国加入 WTO 后进口关税下降这一冲击衡量贸易自由化，考察贸易自由化对行业内加成率分布的影响。考虑到 2001 年进口关税越高的行业在中国加入 WTO 后进口关税下降的幅度越大，该文基于 2001 年行业层面进口关税与进口关税下降时间节点构建单时点非二元处理变量进行分析。研究发现，贸易自由化能够显著降低行业内加成率分布。第三个实例来源于 Favara & Imbs (2015)，该文先考察美国各州银行管制放松对信贷供给的影响，然后进一步考察银行管制放松对房价的影响。各州具有 4 项措施遏制其他州的银行在本州设立分行，取消 1 项措施那么该州银行管制放松指数为 1，数值越大代表放松程度越高（取值范围为 0-4）。每个州银行管制放松指数可以随时间发生变化，因此处理变量为多时点非二元处理变量。研究发现，银行管制放松增加了信贷供给、提升了房价。主要内容总结如表 A1 所示。

---

\*林梦芸，厦门大学经济学院与王亚南经济研究院；徐阳，湖北经济学院财政与公共管理学院；郭汝飞，武汉大学经济与管理学院健康经济与管理研究中心；易君健，北京大学中国经济研究中心，北京大学国家发展研究院。

<sup>1</sup>除特殊说明外，本部分均是未添加控制变量时的核心命令与主要结果展示，添加控制变量各种形式的命令可见 do 文档二元交错处理变量.do。

<sup>2</sup>需要注意的是，本文所使用的数据并非是来自 Stevenson & Wolfers (2006) 的原始数据，而是来自 Goodman-Bacon (2021) 的整理数据，故结果与原文存在一定差异。此外，本文重点使用女性自杀率作为结果变量进行展示，未使用其他结果变量。

表 A1: 主要内容总结

STATA 命令与相关文献	适用范围	文章内容	do 文档
<b>Panel A</b> “负权重”问题的识别诊断方法			
<i>bacondecomp</i>	1. 检验静态模型	正文 3.3 “负权重”问题的识别诊断方法	1. 负权重诊断.do
Goodman-Bacon (2021)	2. 用于二元交错处理变量		
<i>twowayfeweights</i>			
de Chaisemartin & D’Haultfœuille (2020)	1. 检验静态模型		
<i>eventstudyweights</i>	1. 检验动态模型		
Sun & Abraham (2021)	2. 用于二元交错处理变量		
<b>Panel B</b> 处理效应异质性稳健的估计量			
<i>csdid</i>			
Callaway & Sant’Anna (2021)	1. 估计静态与动态模型 2. 用于二元交错处理变量	正文 3.4.1 队列平均处理效应估计量	1. 二元交错处理变量.do
<i>eventstudyinteract</i>	1. 估计静态与动态模型		2. 非二元处理变量-单时点.do
Sun & Abraham (2021)	2. 用于二元交错处理变量		
<i>did_imputation</i>	1. 估计静态与动态模型	正文 3.4.2 插补估计量	3. 非二元处理变量-多时点.do
Borusyak et al. (2023)	2. 用于二元交错处理变量		
<i>did_multiplengt_dyn</i>			
de Chaisemartin & D’Haultfœuille (2020, 2024a)	1. 估计静态与动态模型 2. 用于二元交错处理变量与非二元处理变量	正文 4.2 第二类模型误设：非二元处理变量情形	
<b>Panel C</b> 多个处理变量			
<i>did_multiplengt_old</i>			
de Chaisemartin & D’Haultfœuille (2023)	1. 用于二元交错处理变量与非二元处理变量	正文 5.2 多个处理变量	1. 多个处理变量.do
<b>Panel D</b> 事前趋势检验			
<i>pretrends</i>			
Roth (2022)	1. 用于二元交错处理变量与非二元处理变量	附录 C 平行趋势假设的检验	1. 平行趋势假设的检验.do
<i>honestdid</i>			
Rambachan & Roth (2023)			

## A.1 “负权重”问题识别诊断

目前,关于“负权重”问题识别诊断涉及的 STATA 命令主要包括 *bacondecomp*、*twowayfweights* 与 *eventstudyweights*。

### (1) *bacondecomp*

该命令的常用格式是:

`bacondecomp Y D, ddetail`

其中, Y 代表结果变量; D 代表处理变量, 比如一个州实施了单边离婚法案时为 1, 否则为 0; `ddetail` 代表展示更加详细的分解结果, 如果未添加该选项, 仅展示 `Timing_groups` (包括早处理组作为晚处理组的控制组与晚处理组作为早处理组的控制组)、`Always_v_timing` (始终接受处理组作为控制组) 与 `Never_v_timing` (从未接受处理组作为控制组) 三种情况下的估计系数与权重。如果添加该选项, `Timing_groups` 会进一步明确划分。以 [Stevenson & Wolfers \(2006\)](#) 数据为例, 使用 *bacondecomp* 进行估计的命令与 STATA 结果窗口显示如下:<sup>3</sup>

```
. bacondecomp asmrs post, ddetail
Computing decomposition across 14 timing groups
including an always-treated group and a never-treated group
```

asmrs	Coefficient	Std. err.	z	P> z	[95% conf. interval]
post	-3.079926	1.111656	-2.77	0.006	-5.258732 - .9011193

### Bacon Decomposition

	Beta	TotalWeight
Early_v_Late	3.089867353	.0001259938
Late_v_Early	-23.83420181	.0006803664
Early_v_Late	5.63313818	.0001259938
Late_v_Early	16.86439133	.0000755963
Always_v_timing	-7.043688925	.3844322134
Never_v_timing	-5.330906825	.2402701287

其中, `asmrs` 代表女性自杀率; `post` 代表一个州是否实施单边离婚法案。该命令一方面展示了回归估计结果, 即实施单边离婚法案显著降低了女性自杀率, 另一方面展示了四种 DID 组合的分解结果, 包括 `Early_v_Late` (晚处理组作为早处理组的控制组)、`Late_v_Early` (早处理组作为晚处理组的控制组)、`Always_v_timing` (始终接受处理组作为控制组) 与 `Never_v_timing` (从未接受处理组作为控制组)。将四种组合加总整理后可得:

	group	BetaGroup	TotalW~t
1.	Always_v_timing	-7.043689	.3844322
2.	Early_v_Late	-.18677	.110654
3.	Late_v_Early	3.511967	.2646436
4.	Never_v_timing	-5.330907	.2402701

<sup>3</sup>由于结果过长, 故展示结果窗口部分内容。

可以发现，Late\_v\_Early（早处理组作为晚处理组的控制组）与其他 DID 组合以及回归估计结果均相反，是导致结果出现偏误的来源，表明该例子存在“负权重”问题。另外，值得注意的是，另一组“禁止的对比”，即 Always\_v\_timing（始终接受处理组作为控制组）的估计系数也为负。因此，该例子中始终接受处理组一定程度上可能也可以视作从未接受处理组，进而作为控制组进行估计。do 文档 ATG 是否适合作为控制组.do 对此进行了深入分析，主要基于始终接受处理组与从未接受处理组结果变量的时间趋势折线图、线性时间趋势回归估计两种措施进行判断，发现该例子中的始终接受处理组可以作为控制组进行估计。如果使用者遇到类似情况，也可参考处理。

## (2) twowayfeweights

该命令的常用格式是：

```
twowayfeweights Y G T D, type(string) summary_measures
```

其中，Y 与 D 与上述定义一致；G 代表组别变量，比如美国单边离婚法案发生在州层面，则组别变量为州；T 代表时间变量，比如年份等；type(string) 代表在哪种情况下进行估计，包括 feTR（平行趋势假设下的固定效应回归）、feS（平行趋势假设与处理效应不随时间变化两个假设下的固定效应回归）、fdTR（平行趋势假设下的一阶差分回归）与 fdS（平行趋势假设与处理效应不随时间变化两个假设下的一阶差分回归）；summary\_measures 代表展示补充结果，如果未添加该选项，仅展示“正负权重”ATT 的数量与对应权重。以 Stevenson & Wolfers (2006) 数据为例，使用 twowayfeweights 进行估计的命令与 STATA 结果窗口显示如下：

```
. twowayfeweights asrms stfips year post, type(feTR) summary_measures
```

```
Under the common trends assumption,
the TWFE coefficient beta, equal to -3.0799, estimates a weighted sum of 1107 ATTs.
773 ATTs receive a positive weight, and 334 receive a negative weight.
```

Treat. var: post	# ATTs	$\Sigma$ weights
Positive weights	773	1.3601
Negative weights	334	-0.3601
-----		
Total	1107	1.0000

### Summary Measures:

```
TWFE coefficient ( $\beta_{fe}$ ) = -3.0799
```

```
min  $\sigma(\Delta)$  compatible with  $\beta_{fe}$  and  $\Delta_{TR} = 0$ : 1.4875
```

```
min  $\sigma(\Delta)$  compatible with treatment effect of opposite sign than  $\beta_{fe}$  in all (g,t) cells: 3.2006
```

```
Reference: Corollary 1, de Chaisemartin, C and D'Haultfoeuille, X (2020a)
```

同名变量与前文定义一致。上述结果共三点需要注意：（1）“负权重”（Negative weights）ATT 的数量与对应权重；（2）当 TWFE 回归估计结果与 ATT 实际结果相反时，不同组与不同时间 ATT 标准差的最小值是 1.4875，该数值较大意味着处理效应异质性较大才可能导致 TWFE 回归估计结果与 ATT 实际结果相反，该数值较小则意味着即使处理效应异质性不大也可能导致 TWFE 回归估计结果与 ATT 实际结果相反；（3）当 TWFE 回归估计结果与所有 DID 组合的 ATT 实际结果相反时，不同组与不同时间 ATT 标准差的最小值是 3.2006，该数值含义与（2）种的数值含义类似。然而，目前并没有明确的阈值判断“负权重”ATT 的数量与对应权重是否够大、以

及不同组与不同时间 ATT 标准差的最小值是否够大，因此该建议仅作为判断“负权重”问题的参考。

### (3) *eventstudyweights*

该命令的常用格式是：

```
eventstudyinteract rel_time_list, cohort(varname) absorb(varlist) rel_time(variable) saveweights(filename)
```

其中，*rel\_time\_list* 代表相对时间的虚拟变量；*cohort(varname)* 代表初次接受处理的具体时间，样本期内从未接受处理的组定义为缺失值，*absorb(varlist)* 代表个体和时间固定效应；*rel\_time(variable)* 代表相对时间变量；*saveweights(filename)* 代表保存计算出的权重情况。该命令主要用于检验动态模型，计算出权重情况后，如果需要查看某一期处理效应是否受其他期处理效应影响，则针对相关数据进行画图分析。

以 Stevenson & Wolfers (2006) 数据为例，使用 *eventstudyweights* 进行估计的命令与 STATA 结果窗口显示如下：

```
. eventstudyweights pre* npost*, absorb(stfips year) cohort(_nfd) rel_time(rt) sa  
> veweights("weights")  
file weights.xlsx saved
```

其中，*pre\**、*npost\** 代表每个相对时期的虚拟变量，以-1期为基期；*stfips* 代表州；*year* 代表年份；*\_nfd* 代表各州单边离婚法案实施的具体年份，*rt* 代表实际年份减去各州单边离婚法案实施的具体年份，*weights* 代表文件名。以事后第2期为例，根据图 A1 可以看到事后第2期估计系数的构成情况。可以发现，事后第2期加总估计系数主要由不同组事后第2期的估计系数构成，然而也受其他期估计系数的部分影响，如事后第4期、事前第3期等。如果其他期估计系数所占权重较大，那么说明该期估计系数受“负权重”问题的影响较大。如果期数很多，建议使用者选择政策发生后较近或者估计结果显著性更强的时期进行展示。

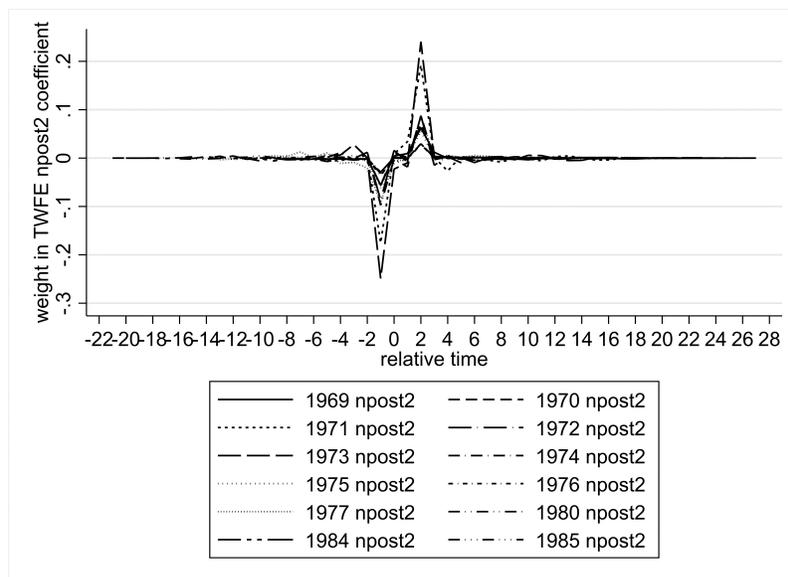


图 A1: 事后第2期估计系数的构成情况 (*eventstudyweights*)

## A.2 处理效应异质性的稳健估计量（二元交错处理变量）

关于二元交错处理变量（对应正文表2），主要涉及到 *csdid*、*eventstudyinteract*、*did\_imputation* 与 *did\_multiplegt\_dyn* 四个命令。<sup>4</sup>

### (1) *csdid*

该命令的常用格式是：

```
csdid Y controls, ivar(G) time(T) gvar(varname) agg(aggtype) cluster(varname) notyet long2
```

其中，Y、G、T与上述定义一致；controls代表控制变量；gvar(varname)代表初次接受处理的具体时间。样本期内从未接受处理的组定义为0，一直接受处理的组需从样本中剔除；<sup>5</sup>agg(aggtype)代表不同情况下的ATT 加总估计结果，包括 simple（所有组所有时期）、group（每个组所有时期）、calendar（每个时期所有组）与 event（每个相对时期所有组）；cluster(varname)代表聚类标准误，当使用面板数据时，标准误已聚类在面板层面。因此，使用该选项时实际是使用双向聚类，如果继续聚类在面板层面，STATA 会报错；notyet代表使用样本期内从未接受处理的样本与尚未接受处理的样本作为控制组，如果未添加 notyet 则代表只使用样本期内从未接受处理的样本作为控制组；<sup>6</sup>long2代表进行动态模型估计时，事前各期系数估计均以政策前1期为基期，如果不添加 long2，事前各期系数估计则以当期前1期为基期。

当进行静态模型估计时，agg(aggtype)需设置为 agg(simple)。当进行动态模型估计时，agg(aggtype)需设置为 agg(event)。需要注意的是，为了使做图时与传统意义上的 TWFE 动态估计具有可比性，需添加选项 long2。

以 Stevenson & Wolfers (2006) 数据为例，使用 *csdid* 进行静态模型估计的命令与 STATA 结果窗口显示如下：

---

<sup>4</sup>此前更早开发的命令 *did\_multiplegt*（目前命名为 *did\_multiplegt\_old*）主要用于估计 de Chaisemartin & D'Haultfœuille (2020) 提出的 DID\_M 估计量，即处理状态发生变化时的处理效应。而 *did\_multiplegt\_dyn* 主要用于估计 de Chaisemartin & D'Haultfœuille (2024a) 提出的 DID\_L 估计量，即处理变量每一期的动态处理效应。两个命令的全面比较可见 de Chaisemartin et al. (2024)。

<sup>5</sup>实际使用过程中，*csdid* 与 *did\_multiplegt\_dyn* 会自动忽略一直接受处理的组进行估计，是否删除不会影响估计结果。但是不删除一直接受处理的组会影响 *eventstudyinteract* 与 *did\_imputation* 的估计结果。为避免使用时出现混淆，建议使用者删除一直接受处理的组。

<sup>6</sup>未添加 notyet 且样本中没有从未接受处理的样本时，该命令会自动使用 notyet 进行估计。



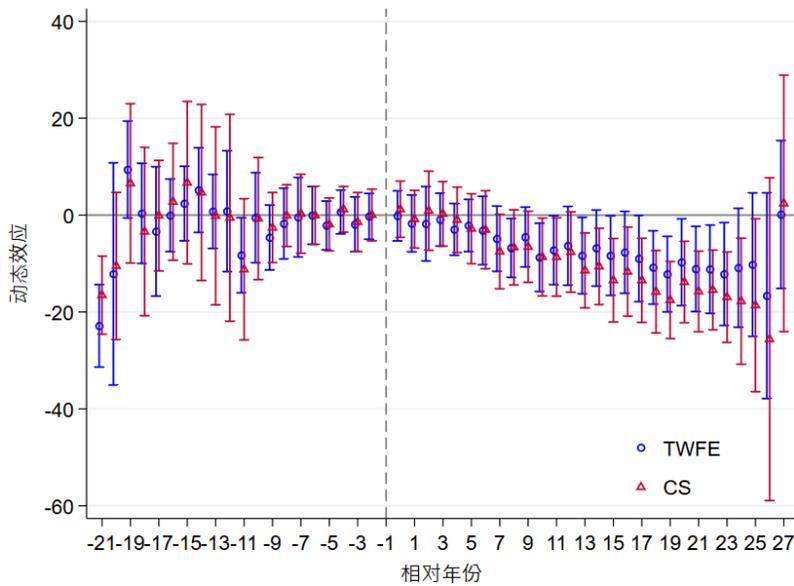


图 A2: 二元交错处理变量动态模型估计 (*csdid*)

## (2) *eventstudyinteract*

该命令的常用格式是:

```
eventstudyinteract Y D, cohort(varname) control_cohort(varname) absorb(varlist) vce(cluster varname) covariates(varlist)
```

其中,  $Y$ 、 $D$  与上述定义一致;  $\text{cohort}(\text{varname})$  代表初次接受处理的具体时间, 样本期内从未接受处理的组定义为缺失值, 一直接受处理的组需从样本中剔除;  $\text{control\_cohort}(\text{varname})$  代表控制组虚拟变量, 通常将从未接受处理的组或最后接受处理的组设置为 1, 否则为 0;  $\text{absorb}(\text{varlist})$  代表个体和时间固定效应;  $\text{vce}(\text{cluster varname})$  代表聚类标准误;  $\text{covariates}(\text{varlist})$  代表控制变量。

当进行静态模型估计时, 使用上述常用格式即可。当进行动态模型估计时, 需将  $D$  替换为一系列相对时间虚拟变量, 即实际时间与接受处理时间相减的虚拟变量。并且将 -1 期剔除作为基期。以 Stevenson & Wolfers (2006) 数据为例, 使用 *eventstudyinteract* 进行静态模型估计的命令与 STATA 结果窗口显示如下:

```
. eventstudyinteract asmrs post, cohort(_nfd) control_cohort(never) absorb(stfips
> year) vce(cluster stfips)
(obs=1,452)
```

```
IW estimates for dynamic effects
Absorbing 2 HDFE groups
Number of obs = 1,617
F(12, 48) = 636.51
Prob > F = 0.0000
R-squared = 0.7146
Adj R-squared = 0.6974
Root MSE = 10.7939
(Std. err. adjusted for 49 clusters in stfips)
```

asmrs	Robust		t	P> t	[95% conf. interval]	
	Coefficient	std. err.				
post	-3.452297	1.829451	-1.89	0.065	-7.130654	.2260612

同名各变量定义与前文一致，变量 `_nfd` 代表各州单边离婚法案实施的具体年份，样本期内从未实施单边离婚法案的州定义为缺失值，一直实施单边离婚法案的州从样本中剔除；变量 `never` 为虚拟变量，从未实施单边离婚法案的州为 1，否则为 0。当将处理变量 `post` 替换为每个相对时期的虚拟变量时可得到动态模型估计，具体命令以及与传统意义上 TWFE 的动态估计系数对比图 (A3) 如下。可以发现，与图 A2 类似，在政策发生后并且距离政策前 1 期越来越远时，“负权重”问题使 TWFE 估计上偏。

```
. eventstudyinteract asmrs pre* npost*, cohort(_nfd) control_cohort(never) absorb(stfips
> year) vce(cluster stfips)
```

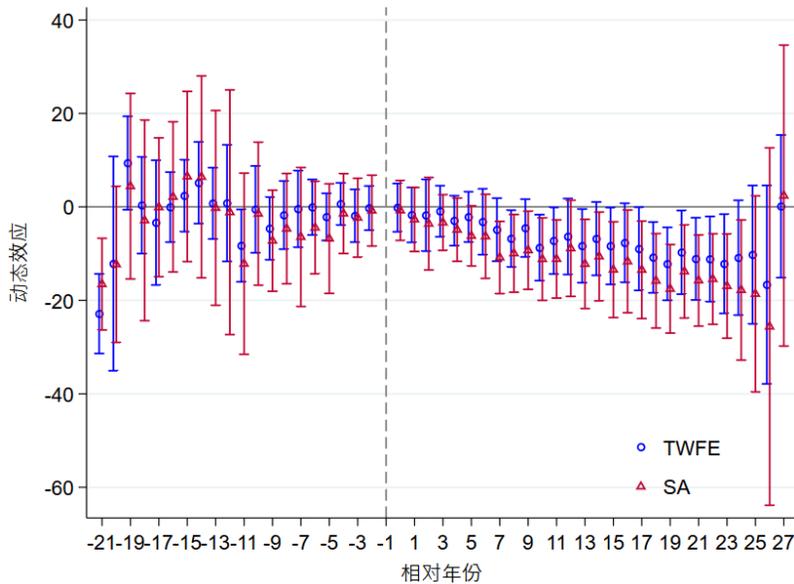


图 A3: 二元交错处理变量动态模型估计 (`eventstudyinteract`)

### (3) `did_imputation`

该命令的常用格式是：

```
did_imputation Y G T Ei, fe(FElist) cluster(varname) controls(varlist) pretrends(#) allhorizons
autosample minn(#) tol(real) maxit(integer)
```

其中，Y、G、T 与上述定义一致；Ei 代表初次接受处理的具体时间，样本期内从未接受处理的组定义为缺失值，一直接受处理的组需从样本中剔除；`fe(FElist)` 代表需要添加的固定效应；`cluster(varname)` 代表聚类标准误；`controls(varlist)` 代表控制变量；`pretrends(#)` 代表与政策发生 # 期前的样本相比，政策发生前 # 期内的处理组与控制组的结果变量差异（即事前趋势检验）；`allhorizons` 代表展示政策发生后所有可估计的处理效应；`autosample` 代表不能应用插补法的观测值会自动从样本中剔除然后进行估计，不添加该选项时若存在不能应用插补法的观测值则会报错；`minn(#)` 代表受处理观测值的数量低于 # 时，相关系数无法估计。默认值为 30，为了估计出相关系数可以灵活调整。需要注意的是，受处理观测值的数量较少时，相关系数推断可能不可靠；`tol(real)` 与 `maxit(integer)` 代表容忍度和迭代最大次数，当标准误计算无法收敛时可以尝试添加这两个选项。两个选项的默认值分别是  $10^{-6}$  和 100。另外，这两个选项可以单独使用，也可以一起使用。

当进行静态模型估计时，基于 *did\_imputation* 上述常用格式即可，不需添加 `pretrends(#)` 和 `allhorizons`。当进行动态模型估计时，需添加 `pretrends(#)` 和 `allhorizons`。需要注意的是，该估计量的事前系数估计并非以政策前 1 期为基期，而是以 # 前的样本为基期。因此，当 `pretrends(#)` 中的 # 变化时，政策发生前各期系数也会发生变化。并且基于该估计量进行动态模型估计时，政策前 1 期的系数也能够估计。另外，如果使用者不需展示政策发生后所有可估计的处理效应，可将 `allhorizons` 调整为 `horizons(numlist)`，`numlist` 可根据需要设置为政策某一期或者某几期。由于该估计量的事后系数是基于尚未接受处理的样本进行插补估计，`numlist` 的数值变化不会导致事后系数发生变化。

以 Stevenson & Wolfers (2006) 数据为例，使用 *did\_imputation* 进行静态模型估计的命令与 STATA 结果窗口显示如下：

```
. did_imputation asmrs stfips year _nfd, fe(stfips year) cluster(stfips) autosample
```

Number of obs = 1,353					
asmrs	Coefficient	Std. err.	z	P> z	[95% conf. interval]
tau	-5.530065	2.978178	-1.86	0.063	-11.36719 .3070564

同名各变量定义与前文一致。如果要进行动态模型估计，需要添加 `pretrends(#)` 和 `allhorizons`，具体命令以及与传统意义上 TWFE 的动态估计系数对比图（图 A4a）如下。可以发现，二者事后估计较为接近，但事前估计差别很大，*did\_imputation* 产生的估计系数明显更大并且均显著为正。可能的原因在于该估计量事前估计的基期不是 -1 期，而是取决于 `pretrends(#)`。此时 `pretrends(#)` 设置为 `pretrends(20)`，那么基期是 -21 期及之前的所有时期。由于距离较远的相对时期通常观测值较少，进而可能影响估计结果。笔者将 `pretrends(20)` 替换为 `pretrends(15)`，以 -16 期及之前的所有时期为基期进行估计，结果如图 A4b 所示。可以发现，此时新估计结果（BJS2）与 TWFE 的事前估计结果非常接近，以及不同 `pretrends(#)` 的事后估计结果完全一致。这些结果说明：第一，`pretrends(#)` 需要进行合理设置，否则可能影响事前估计结果进而干扰对平行趋势检验的判断；第二，这一估计量事前与事后的估计结果互不影响，事前估计受 `pretrends(#)` 的影响，事后估计不会受该选项的影响发生变化。

```
. did_imputation asmrs stfips year _nfd, fe(stfips year) cluster(stfips) autosample pretrend(20)
> allhorizons minn(0)
```

#### (4) *did\_multiplengt\_dyn*

该命令的常用格式是：

```
did_multiplengt_dyn Y G T D, effects(#) placebo(#) controls(varlist) cluster(varname)
```

其中，Y、G、T、D 与上述定义一致；`effects(#)` 代表与政策发生前 1 期相比，政策发生后 # 期内处理组与控制组的结果变量差异（即处理效应）；`placebo(#)` 代表与政策发生前 1 期相比，政策发生前 # 期内的处理组与控制组的结果变量差异（即事前趋势检验）；`controls(varlist)` 代表添加控制变量；`cluster(varname)` 代表聚类标准误。

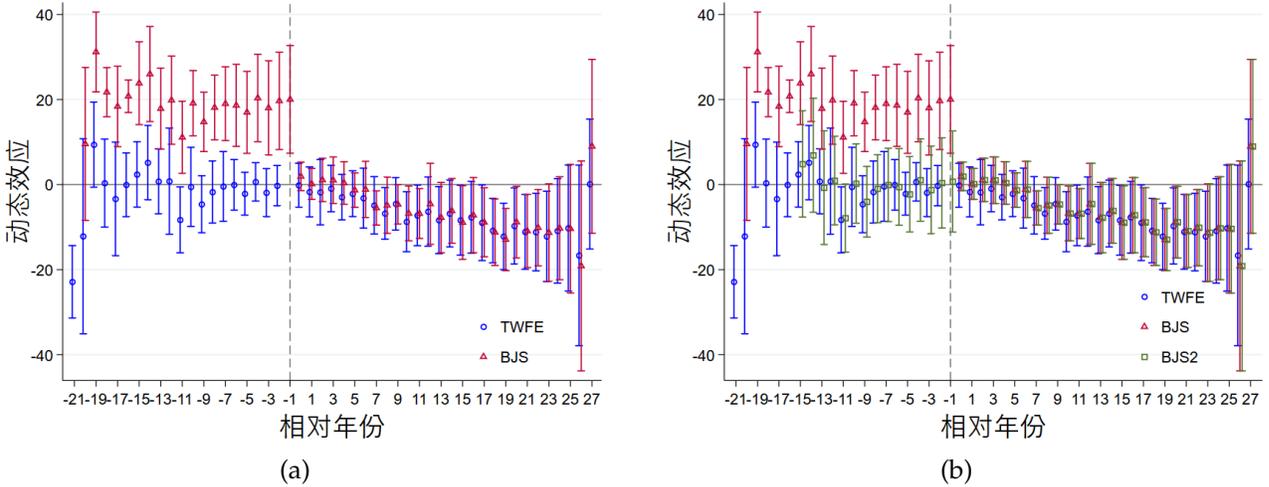


图 A4: 二元交错处理变量动态模型估计 (*did\_imputation*)

当进行静态模型估计时，平均处理效应估计值取决于 `effects(#)` 的数值，与 `placebo(#)` 的数值无关。当进行动态模型估计时，每一期的处理效应估计值与 `effects(#)` 的数值、`placebo(#)` 的数值均无关，不受数值 # 的影响。使用者可根据具体需求，选择合适的数值。需要注意的是，`Effect_1` 代表与政策发生前 1 期相比，政策发生当期（即传统意义上 TWFE 动态估计时的第 0 期）处理组与控制组的结果变量差异，以此类推；`Placebo_1` 代表与政策发生前 1 期相比，政策发生前第 2 期（即传统意义上 TWFE 动态估计时的 -2 期）处理组与控制组的结果变量差异，以此类推。为了使做图时与传统意义上的 TWFE 动态估计具有可比性，笔者在 `do` 文档中对相关定义进行了修改（如二元交错处理变量 `.do` 的第 765-784 行）。以 Stevenson & Wolfers (2006) 数据为例，使用 `did_multiplegt_dyn` 进行估计的命令与 STATA 结果窗口显示如下：

```
. did_multiplegt_dyn asrms stfips year post, cluster(stfips) graph_off
```

Estimation of treatment effects: Event-study effects						
	Estimate	SE	LB CI	UB CI	N	Switchers
Effect_1	1.209345	2.954957	-4.582263	7.000954	249	36

Average cumulative (total) effect per treatment unit						
	Estimate	SE	LB CI	UB CI	N	Switch x Periods
Av_tot_eff	1.209345	2.954957	-4.582263	7.000954	249	36

Average number of time periods over which a treatment's effect is accumulated = 1

其中，同名变量与前文定义一致；`graph_off` 代表不生成图片；<sup>8</sup> `Swithers` 代表处理状态发生变化的州（处理组），即从未实施单边离婚法案转变为实施单边离婚法案的州。由于未设置 `effects(#)`，默认估计政策当期的处理效应 (`Effect_1`)，而且与 ATT 一致 (`Av_tot_eff`)。如果使用者想要估计政策发生后 # 期内的 ATT 可添加 `effects(#)`。以估计政策发生后 3 期内的 ATT 为例，具体命令与 STATA 结果窗口显示如下。可以看到，此时可估计出政策发生后 3 期内的每一期系数，基于这 3 期的处理效应可得到静态模型估计的 ATT。

```
. did_multiplengt_dyn asmrs stfips year post, effects(3) cluster(stfips) graph_off
```

Estimation of treatment effects: Event-study effects						
	Estimate	SE	LB CI	UB CI	N	Switchers
Effect_1	1.209345	2.954957	-4.582263	7.000954	249	36
Effect_2	-.8466385	3.035562	-6.79623	5.102953	217	36
Effect_3	.9020601	4.322185	-7.569266	9.373387	188	36

Test of joint nullity of the effects : p-value = .61026132

Average cumulative (total) effect per treatment unit						
	Estimate	SE	LB CI	UB CI	N	Switch x Periods
Av_tot_eff	.421589	3.251223	-5.950691	6.793869	361	108

Average number of time periods over which a treatment's effect is accumulated = 2

如果使用者想要得到所有相对时期的估计系数，可以基于样本内相对时期的具体情况设置 `effects(#)` 与 `placebo(#)`。比如，该实例中的实际年份与各州单边离婚法案实施年份可得到的相对时期为 -21 至 27，因此可添加选项 `effects(28)` 与 `placebo(21)`。具体命令与 STATA 结果窗口显示如下。其中，有以下三点需要注意：第一，并非政策前所有相对时期均可估计，本例中政策前最多只能估计 `Placebo_15`。另外，不论 `effects(#)` 与 `placebo(#)` 的数值设置为多少，如果超过可以估计的上限，那么该命令会自动给出可以估计的最大期数。第二，如前文所述，`Effect_1` 代表传统意义上 TWFE 动态估计时的第 0 期，`Placebo_1` 代表传统意义上 TWFE 动态估计时的 -2 期，因此作图时需重新定义。第三，最下方提供了政策前估计系数的联合无效性检验 (`Test of joint nullity of the placebos`)，可以作为参考。

<sup>8</sup>如果不添加选项 `graph_off`，该命令会基于 `effects(#)` 与 `placebo(#)` 的设置自动生成一副动态效应图。为了方便与传统意义上的 TWFE 动态估计、以及其他估计量对比，建议添加该选项，然后手动提取相关数据作图。具体命令见 `do` 文档二元交错处理变量 `do`。

```
. did_multiplengt_dyn asms stfips year post,effects(28) placebo(21) cluster(stfips) graph_off
```

The number of placebos which can be estimated is at most 15.  
The command will therefore try to estimate 15 placebo(s).

Estimation of treatment effects: Event-study effects						
	Estimate	SE	LB CI	UB CI	N	Switchers
Effect_1	1.209345	2.954957	-4.582263	7.000954	249	36
Effect_2	-.8466385	3.035562	-6.79623	5.102953	217	36
Effect_28	2.41363	17.37399	-31.63877	36.46603	7	2

Test of joint nullity of the effects : p-value = 0

Average cumulative (total) effect per treatment unit							
	Estimate	SE	LB CI	UB CI	N	Switch	x Periods
Av_tot_eff	-8.563262	3.403609	-15.23421	-1.892311	1148	843	

Average number of time periods over which a treatment's effect is accumulated = 12.475682

Testing the parallel trends and no anticipation assumptions						
	Estimate	SE	LB CI	UB CI	N	Switchers
Placebo_1	.0265667	2.942508	-5.740644	5.793777	249	36
Placebo_2	-1.158382	3.106005	-7.24604	4.929276	217	36
Placebo_15	.4273383	7.542735	-14.35615	15.21083	6	1

Test of joint nullity of the placebos : p-value = .00027144

*did\_multiplengt\_dyn* 得到的动态估计系数与传统意义上 TWFE 的动态估计系数的对比如图 A5 所示。Effect\_# 与 Placebo\_# 已重新定义，与传统意义上的 TWFE 动态估计一致，每期系数估计均以政策前 1 期为基期。可以发现，与图 A2、图 A3 类似，在政策发生后并且距离政策前 1 期越来越远时，TWFE 估计在“负权重”影响下开始出现上偏。

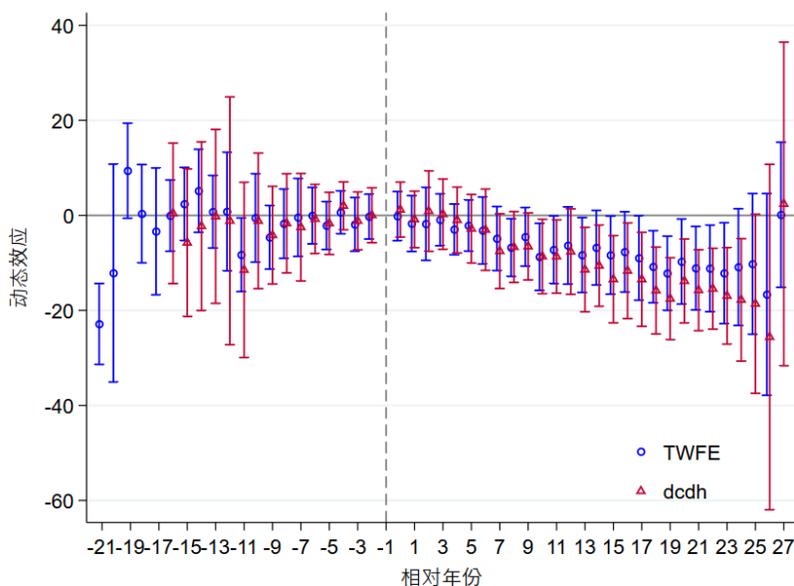


图 A5: 二元交错处理变量动态模型估计 (*did\_multiplegt\_dyn*)

### A.3 处理效应异质性稳健的估计量 (非二元处理变量)

关于非二元处理变量，包括单时点非二元处理变量（对应正文表 4）与多时点非二元处理变量（对应正文表 5）。其中，单时点非二元处理变量的估计通过 Lu & Yu (2015) 为例来展示，多时点非二元处理变量的估计通过 Favara & Imbs (2015) 来展示，这两种情形均是基于工具变量两步法思路计算 ATT (de Chaisemartin & D’Haultfoeuille, 2024a)，主要涉及 *did\_multiplegt\_dyn* 这一命令，常用格式、静态模型与动态模型估计的相关内容 with A.2 第 (4) 部分一致。笔者在 do 文档（见非二元处理变量-单时点.do 的第 94-162 行和 非二元处理变量-多时点.do 的第 181-414 行）呈现了该估计量的工具变量两步法思路。以 Favara & Imbs (2015) 为例，该文处理变量为 0-4 的多时点非二元处理变量。基于该处理变量生成 0-1 工具变量之后，先估计工具变量对结果变量的影响（图 A6a），再估计工具变量对处理变量的影响（图 A6b），最后根据每个相对时期处理组观测值数量占全部时期处理组观测值数量的比例进行加权，得到 ATT。<sup>9</sup>

### A.4 协变量

现有文献关于控制变量的添加形式主要包括直接添加随时间变化的控制变量、添加不随时间变化控制变量或前定变量（处理发生前的控制变量值，如研究样本期初值），与时间虚拟变量或线性时间趋势的交互项。可以基于研究主题，通过 *did\_multiplegt\_dyn*、*csdid*、*eventstudyinteract*、*did\_imputation* 四个命令的相关选项添加各种形式的控制变量避免模型误设。需要注意的是，部分命令（如 *did\_multiplegt\_dyn*、*csdid*）不支持通过 *c.varname # i.year* 添加控制变量，因此需要手动生成相关变量进行添加。

<sup>9</sup>图 A6b 第 9 期没有置信区间的原因可能是此时共有 130 个观测值，只有 2 个观测值属于处理组。

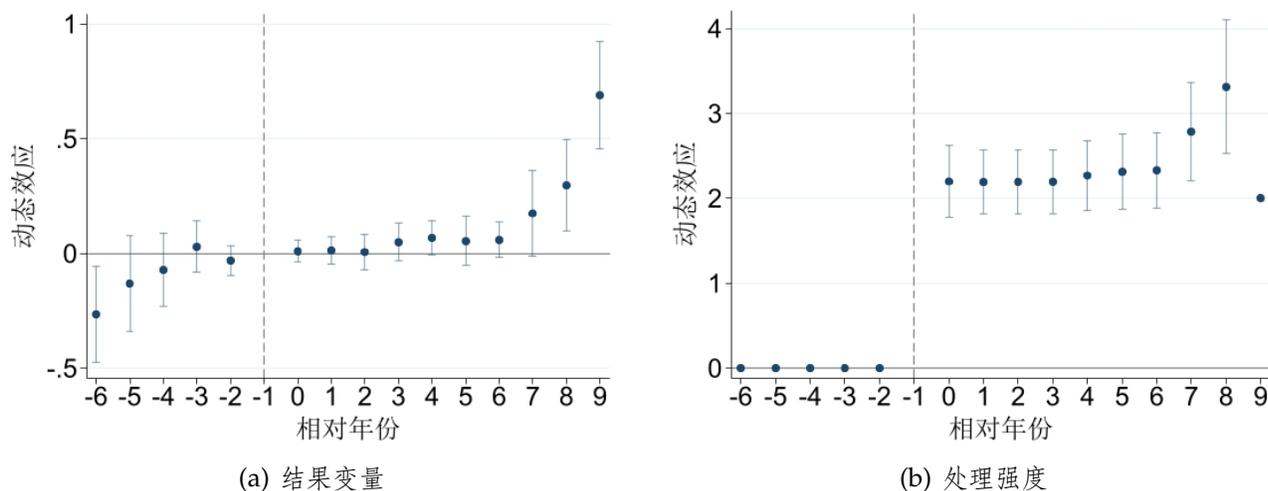


图 A6: 非二元处理变量工具变量两步法思路 (*did\_multiplegt\_dyn*)

另外，当存在多个处理变量时（对应正文表6），为避免可能导致的“负权重”问题，可以基于 STATA 命令 *did\_multiplegt\_old* 进行估计。<sup>10</sup> 该命令的常用格式是：

```
did_multiplegt_old Y G T D, robust_dynamic dynamic(#) placebo(#) controls(varlist) cluster(varname) breps(#) seed(#)
```

与 *did\_multiplegt\_dyn* 的格式十分类似。区别之处在于当进行动态模型估计时，需要添加选项 *robust\_dynamic*，以及 *effects(#)* 此时由 *dynamic(#)* 代替。另外，*breps(#)* 代表计算标准误时 bootstrap 的重复次数，默认为 50 次；*seed(#)* 代表设置 bootstrap 估计时使用的种子，以方便结果复现。

当需要控制其他处理变量 (*othertreat*) 时，需要添加 *if\_first\_diff(fd\_othertreat==0)*、*trends\_nonparam(othertreat)*、*always\_trends\_nonparam* 这三个选项（可见 do 文档多个处理变量.do）。其中，*fd\_othertreat* 为其他处理变量的一阶差分；*if\_first\_diff(fd\_othertreat==0)* 代表估计时剔除其他处理变量一阶差分不为 0 的观测值；*trends\_nonparam(othertreat)* 代表估计量是基于其他处理变量 (*othertreat*) 相同的样本进行估计的加权平均值；*trends\_nonparam(othertreat)* 和 *always\_trends\_nonparam* 同时使用代表剔除其他处理变量 (*othertreat*) 相同的样本中，处理状态发生变化 (*switchers*) 且没有对应控制组的观测值进行估计。如果需要控制多个处理变量，可以先通过 *egen othertreats = group(othertreat1 othertreat2)* 生成一个新变量，然后沿用上述思路进行估计。

由于 21 世纪早期中国实施了国有企业改革与外资管制放松，属于中国加入 WTO 的同期冲击。笔者以 Lu & Yu (2015) 为例进行呈现使用 *did\_multiplegt\_old* 控制处理变量的估计结果，并与 TWFE 估计结果进行对比（图 A7）。<sup>11</sup>

<sup>10</sup>*did\_multiplegt\_old* 是之前的命令 *did\_multiplegt* 的重新命名，该命令是控制其他处理变量的初始命令。由于目前 *did\_multiplegt\_dyn* 的选项与 *did\_multiplegt\_old* 存在一定区别，故仍使用后者。

<sup>11</sup>为简化分析，该部分将处理变量、刻画国有企业改革与外资管制放松的三个连续型变量均设置为二元变量进行估计。

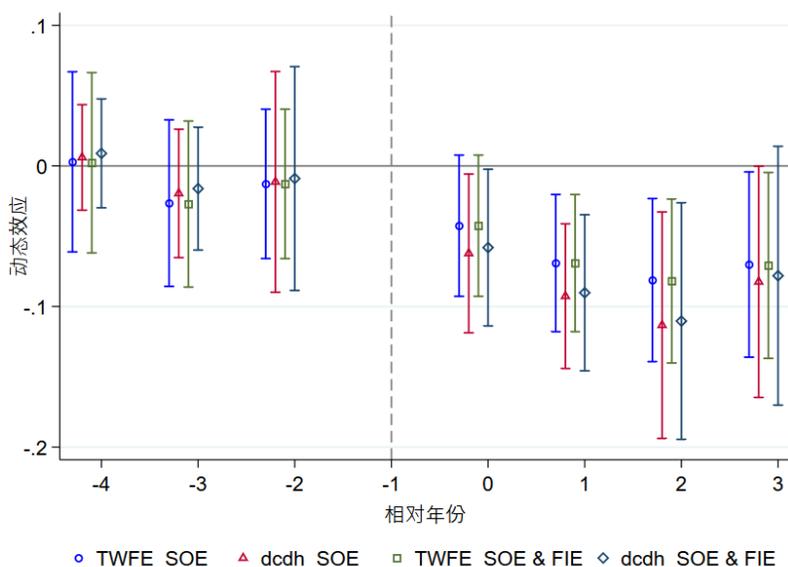


图 A7: 多个处理变量

## A.5 平行趋势假设的检验

### (1) 事前趋势检验可靠性的诊断

这一诊断主要用到 STATA 命令 *pretrends*, 该命令常用的常用格式是:

```
pretrends power ##, pre(##/##) post(##/##) pretrends, numpre(#) b(str) v(str) slope(real)
```

第一行命令代表基于事前期间 *pre(##/##)* 与事后期间 *post(##/##)* 计算出有多大功效能发现的线性事前趋势。通常将 *power* 设置为 0.8, 即 80% (Cohen, 1988; Roth, 2022)。第二行命令代表基于回归估计结果与第一行命令得到的线性事前趋势斜率拟合出线性事前趋势存在情况下的动态效应估计。*numpre(#)* 代表事前期数的数量, *b(str)* 与 *v(str)* 分别代表回归估计的系数矩阵与方差-协方差矩阵, *slope(real)* 代表第一行命令得到的线性事前趋势斜率。

以 Stevenson & Wolfers (2006) 数据为例, 基于传统 TWFE 回归, 使用 *pretrends* 进行诊断的相关命令与 STATA 结果窗口主要内容显示如下:

```
. reghdfe asmrs pre* npost*, a(stfips year) cluster(stfips)

. pretrends power 0.8, pre(16/20) post(21/25)
Slope for 80% power = 1.52451

. matrix beta = e(b)

. matrix sigma = e(V)

. matrix beta = beta[., 16..25]

. matrix sigma = sigma[16..25, 16..25]

. pretrends, numpre(5) b(beta) v(sigma) slope(`r(slope)') ///
> graphregion(color(white))
```

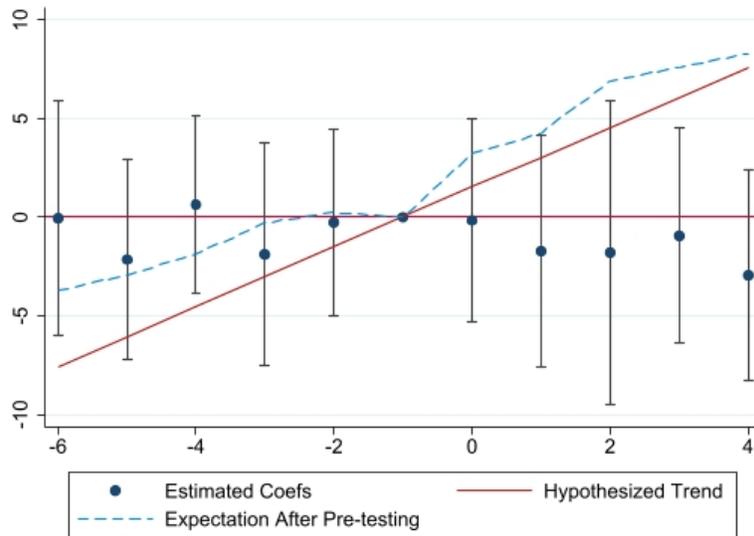


图 A8: 事前趋势检验可靠性的诊断 (*pretrends*)

其中，同名变量与前文定义一致。诊断顺序为：第一，先进行回归估计。第二，根据动态估计系数次序选择需要检测的时期，笔者选择事前第 6 期至第 2 期（次序为 16-20），事后当期至第 4 期（次序为 21-25）为例，并选择功效为 80%。可以得到 80% 功效能够发现的线性事前趋势斜率是 1.52451。第三，根据该斜率，结合第一步回归估计的系数矩阵与方差-协方差矩阵得到图 A8。图中带有置信区间与深蓝色点的竖线为原始回归估计系数，红色实线为斜率是 1.52451 的直线，蓝色虚线为假如不满足事前趋势，即存在斜率是 1.52451 的线性事前趋势时预计的回归估计系数。如果预计的回归估计系数大部分落在原始回归估计系数的置信区间内，那么表明可能存在线性事前趋势，即不满足平行趋势假设。另外，*pretrends* 除了可以针对命令 *reghdfe* 的回归估计结果进行检测外，针对渐进正态估计量的回归估计结果也可以检测，比如 Callaway & Sant’Anna (2021) 与 Sun & Abraham (2021) 估计量。

## (2) 稳健性和敏感性分析

这一分析主要用到 STATA 命令 *honestdid*，该命令常用的常用格式是：

`honestdid, pre(#) post(#) mvec(str) delta(str) l_vec(str) coefplot`

该命令主要基于事前估计，判断当平行趋势假设不满足时，事后估计结果是否保持稳健。其中，*pre(#)* 与 *post(#)* 分别代表事前与事后期间，均可以设置 1 个时期或多个时期。如果要对多个事后时期的整体处理效应进行分析，需添加选项 *l\_vec(str)*。比如对两个时期的平均处理效应进行分析，设置 *l\_vec(0.5 \0.5)*。当假设事后趋势是事前趋势的倍数时（相对幅度限制），设置 *mvec(str)*；当假设连续时期趋势差异的斜率变化不超过某个百分比时（平滑性限制），进一步添加 *delta(sd)*。需要做图时，添加选项 *coefplot*。

以 Stevenson & Wolfers (2006) 数据为例，基于传统 TWFE 回归，假设事后趋势是事前趋势的倍数变化时，使用 *honestdid* 进行诊断的相关命令与 STATA 结果窗口主要内容显示如下：

```
. reghdfe asmr pre* npost*, a(stfips year) cluster(stfips)
```

```
. honestdid,pre(16/20) post(29) mvec(0.5(0.5)2) coefplot xtitle(Mbar) ytitle(95%置信区间) ///
> graphregion(color(white))
```

M	lb	ub	
-----	-----	-----	
.	-12.849	-0.711	(Original)
0.5000	-14.450	0.382	
1.0000	-16.690	1.964	
1.5000	-19.417	4.005	
2.0000	-22.250	6.258	

(method = C-LF, Delta = DeltaRM, alpha = 0.050)

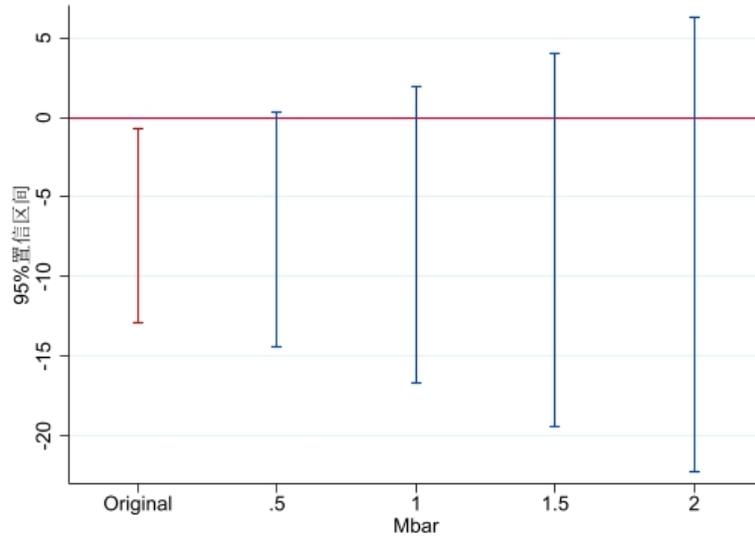


图 A9: 相对幅度限制下的稳健性和敏感性分析 (*honestdid*)

同名变量与前文定义一致。笔者基于事前第 6 期至第 2 期（次序为 16-20）判断事后第 8 期（次序为 29）系数估计的稳健性和敏感性。图 A9 中红线为事后第 8 期估计系数的原始置信区间，后续蓝线分别为当事后趋势是事前趋势的 0.5 倍、1 倍、1.5 倍与 2 倍时，事后第 8 期估计系数的置信区间。可以发现，当事后趋势是事前趋势 0.5 倍时，事后第 8 期系数就开始变为不显著。

当假设连续时期趋势差异的斜率变化不超过某个百分比时，使用 *honestdid* 进行诊断的相关命令与 STATA 结果窗口主要内容显示如下。基于图 A10 可以发现，即使当连续时期间趋势差异的斜率变化为 0.1% 时，事后第 8 期系数依然显著。

```
. honestdid,pre(16/20) post(29) mvec(0(0.02)0.1) delta(sd) coefplot xtitle(Mbar) ///
> ytitle(95%置信区间) graphregion(color(white))
```

M	lb	ub	
-----	-----	-----	
.	-12.849	-0.711	(Original)
0.0000	-10.829	-0.291	
0.0200	-10.832	-0.289	
0.0400	-10.843	-0.282	
0.0600	-10.860	-0.271	
0.0800	-10.883	-0.257	
0.1000	-10.911	-0.239	

(method = FLCI, Delta = DeltaSD, alpha = 0.050)

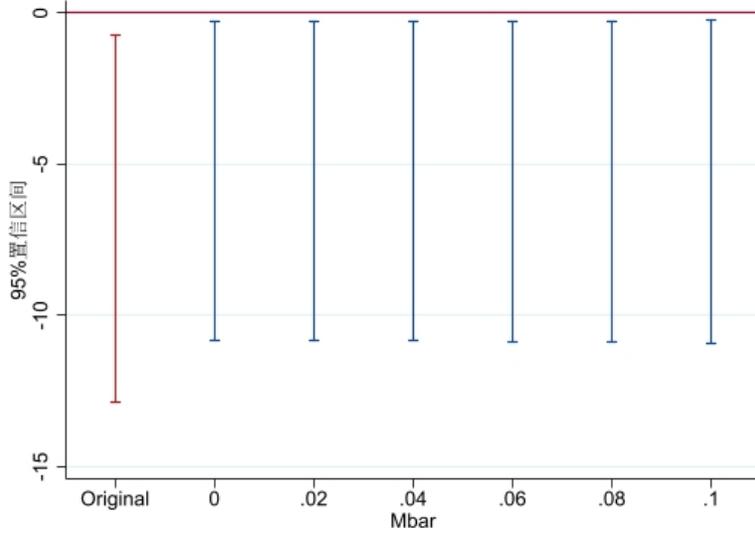


图 A10: 平滑性幅度限制下的稳健性和敏感性分析 (*honestdid*)

## B 证明

### B.1 权重：常系数回归模型

$\omega_x$  来自于回归模型  $D_i I_i^x = \omega_x D_i + \sum_{x=x_1}^{x_m} \tilde{\lambda}_x I_i^x + v_i, \forall x \in x_1, x_2, \dots, x_m$ 。<sup>12</sup>

根据 Frisch—Waugh—Lovell (下文简称 FWL) 定理，可得到  $\omega_x = \frac{\text{Cov}(D_i I_i^x, \tilde{D}_i)}{\text{Var}(\tilde{D}_i)}$ ，其中  $\tilde{D}_i$  来自于  $D_i = \sum_{x=x_1}^{x_m} \tilde{\lambda}_x I_i^x + \tilde{D}_i$ ，即  $D_i$  对  $\sum_{x=x_1}^{x_m} \tilde{\lambda}_x I_i^x$  进行回归得到残差项的估计值。根据 OLS 估计公式，可知  $\tilde{\lambda}_x = E(D_i | X_i = x)$ 。据此，可推导出  $\omega_x$  的表达式为

$$\omega_x = \frac{\text{Cov}(I_i^x D_i, \tilde{D}_i)}{\text{Var}(\tilde{D}_i)} = \frac{\frac{N_x}{N} \text{Var}(D_i | X_i = x)}{\sum_{x=x_1}^{x_m} \frac{N_x}{N} \text{Var}(D_i | X_i = x)}$$

其中， $N$  为总样本数， $N_x$  是  $X$  取值为  $x$  的样本数。

<sup>12</sup>这个模型从遗漏变量偏误 (Omitted Variable Bias, 下文简称 OVB) 公式得到。TWFE 回归模型 (1) 遗漏了系列交互项  $\sum_{(g,t) \in \Omega} \beta_{g,t} I_{g,t} D_{g,t}$ ，将其补足则为模型 (7)，此时模型 (1) 中原有  $D_{g,t}$  的系数  $\beta = 0$ 。后文 B.2 中的  $\omega_{g,t}$  和 B.3 中的  $\omega_{g,t}^d$  证明类似。

## B.2 权重：二元交错处理变量（静态模型）

$\omega_{g,t}$  来自回归模型  $I_{g,t}D_{g,t} = \tilde{\alpha}_g + \tilde{\lambda}_t + \omega_{g,t}D_{g,t} + v_{i,g,t}$ ,  $\forall (g,t) \in \Omega$ .<sup>13</sup>

根据 FWL 定理，可推导出  $\omega_{g,t} = \frac{\text{Cov}(I_{g,t}D_{g,t}, \tilde{D}_{g,t})}{\text{Var}(\tilde{D}_{g,t})}$ ,  $\tilde{D}_{g,t}$  来自  $D_{g,t} = \hat{\alpha}_g + \hat{\lambda}_t + \tilde{D}_{g,t}$ ; 则  $\tilde{D}_{g,t} = D_{g,t} - \bar{D}_{\cdot t} - \bar{D}_{g\cdot} + \bar{D}$ , 其中  $\bar{D}_{\cdot t} = \frac{1}{N_t} \sum_g D_{g,t}$ ,  $\bar{D}_{g\cdot} = \frac{1}{N_g} \sum_t D_{g,t}$ ,  $\bar{D} = \frac{1}{N} \sum_{g,t \in \Omega} D_{g,t}$ . 进一步推导可得：

$$\begin{aligned} \omega_{g,t} &= \frac{\text{Cov}(I_{g,t}D_{g,t}, \tilde{D}_{g,t})}{\text{Var}(\tilde{D}_{g,t})} = \frac{\frac{1}{N} \sum_{(g,t) \in \Omega} I_{g,t}D_{g,t}\tilde{D}_{g,t}}{\frac{1}{N} \sum_{(g,t) \in \Omega} \tilde{D}_{g,t}^2} \\ &= \frac{N_{g,t}D_{g,t}\tilde{D}_{g,t}}{\sum_{(g,t) \in \Omega} N_{g,t}D_{g,t}\tilde{D}_{g,t}} \\ &= \frac{N_{g,t}\tilde{D}_{g,t}}{\sum_{(g,t) \in \Omega_1} N_{g,t}\tilde{D}_{g,t}} D_{g,t}, \end{aligned}$$

其中， $\Omega_1 = \{(g,t) | D_{g,t} = 1\}$ ，是所有  $D_{g,t} = 1$  的观测值的集合； $N_{g,t}$  为  $(g,t)$  当中的样本数。

## B.3 权重：非二元处理变量

$\omega_{g,t}^d$  来自回归模型  $I^d \mathbf{1}\{D_{g,t} > 0\} = \tilde{\alpha}_g + \tilde{\lambda}_t + \omega_{g,t}^d D_{g,t} + v_{g,t}^d$ ,  $\forall (g,t) \in \Omega$ .

根据 FWL 定理，可推导出  $\omega_{g,t}^d = \frac{\text{Cov}(I^d \mathbf{1}\{D_{g,t} > 0\}, \tilde{D}_{g,t})}{\text{Var}(\tilde{D}_{g,t})}$ , 其中  $\tilde{D}_{g,t}$  来自  $D_{g,t} = \hat{\alpha}_g + \hat{\lambda}_t + \tilde{D}_{g,t}$ , 则  $\tilde{D}_{g,t} = D_{g,t} - \bar{D}_{\cdot t} - \bar{D}_{g\cdot} - \bar{D}$ , 其中  $\bar{D}_{\cdot t} = \frac{1}{N_t} \sum_g D_{g,t}$ ,  $\bar{D}_{g\cdot} = \frac{1}{N_g} \sum_t D_{g,t}$ ,  $\bar{D} = \frac{1}{N} \sum_{g,t \in \Omega} D_{g,t}$ . 进一步推导，可得：

$$\begin{aligned} \omega_{g,t}^d &= \frac{\text{Cov}(I^d \mathbf{1}\{D_{g,t} > 0\}, \tilde{D}_{g,t})}{\text{Var}(\tilde{D}_{g,t})} \\ &= \frac{\frac{1}{N} \sum_{(g,t) \in \Omega} \sum_{d=d_1}^{d_m} I^d \mathbf{1}\{D_{g,t} > 0\} \tilde{D}_{g,t}}{\frac{1}{N} \sum_{(g,t) \in \Omega} \tilde{D}_{g,t}^2} \\ &= \frac{N_{g,t}^d \mathbf{1}\{D_{g,t} > 0\} \tilde{D}_{g,t}}{\sum_{(g,t) \in \Omega} N_{g,t} D_{g,t} \tilde{D}_{g,t}} \\ &= \frac{N_{g,t}^d \tilde{D}_{g,t}}{\sum_{(g,t) \in \Omega_2} N_{g,t} D_{g,t} \tilde{D}_{g,t}} \mathbf{1}\{D_{g,t} > 0\}, \end{aligned}$$

<sup>13</sup>这一分解结果是 de Chaisemartin and D'Haultfoeuille (2020) 的重要发现之一，并且在 Borusyak et al. (2023) 的研究中也得到了体现。

其中,  $\Omega_2 = \{(g, t) | D_{g,t} > 0\}$ , 是所有  $D_{g,t} > 0$  的观测值的集合;  $N_{g,t}^d$  为  $D_{g,t} = d$  的样本数。

#### B.4 讨论: 对比处理效应异质性稳健的二元交错处理变量估计量

插补估计量和 Sum & Abraham (2021) 的 IW 估计量依赖于平行趋势假设对于所有组在所有样本期都成立 (假设1)。而 Callaway & Sant'Anna (2021) 的估计量所要求的平行趋势假设相对较弱:

假设 6 对于任意  $g$  和  $t \geq E_g$ ,  $E \left[ Y_{i,g,t}^0 \right] - E \left[ Y_{i,g,t-1}^0 \right] = E \left[ Y_{ic,t}^0 \right] - E \left[ Y_{ic,t-1}^0 \right]$ , 其中  $c$  表示的是控制组, 包括从未接受处理组或尚未接受处理组。

这一假设只要求处理组  $g$  在处理开始前期  $E_g - 1$  起, 结果变量的变化趋势与控制组保持一致即可。因此, 如果研究者对于整个样本期都满足平行趋势假设这一点不够自信, 例如处理组和控制组有不同的事前趋势, 那么 Callaway & Sant'Anna (2021) 的估计量带来的潜在估计偏误会更小。

如果存在预期效应, Callaway & Sant'Anna (2021) 和 Sum & Abraham (2021) 的估计量都是将处理发生前期作为基期来估计 DID 的处理效应, 因此如果存在预期效应, 即存在  $t < E_g$ ,  $E \left[ Y_{i,g,t}^1 \right] - E \left[ Y_{i,g,t}^0 \right] \neq 0$ , 这两个估计量的估计误差较大;<sup>14</sup>但这一问题对于 Borusyak et al. (2023) 的插补估计量影响较小, 因为他将所有处理发生前的观测值均纳入了反事实结果的估计当中。Borusyak et al. (2023) 进一步证明得到, 在高斯-马尔可夫定理的假设下, 插补估计量是最有效的。

此外, 在二元交错处理变量的情境中, de Chaisemartin & D'Haultfœuille (2020) 和 de Chaisemartin & D'Haultfœuille (2024a) 提出的估计方法同样适用。其中, 前者专注于评估处理组在处理当期的处理效应, 而后者在不引入控制变量的条件下, 与 Callaway and Sant'Anna (2021) 的估计方法相等。不过, 由于 de Chaisemartin & D'Haultfœuille (2020) 和 de Chaisemartin & D'Haultfœuille (2024a) 的方法还能适用于非二元处理变量的场景, 笔者将在后续内容中对此进行更为详尽的探讨。<sup>15</sup>

#### B.5 de Chaisemartin & D'Haultfœuille (2024a) 中的标准化估计量 $DID_{g,l}^{Yn}$ 和 $DID_l^{Yn}$

考虑到广义 DID 中处理强度变化的一般化, de Chaisemartin & D'Haultfœuille (2024a) 提出用处理强度变化来对  $\delta_{g,l}^Y$  进行标准化, 得到标准化的处理效应: 对于任意  $F_g < T_g$ , 并且  $l \in \{1, \dots, T_g - F_g + 1\}$ ,

$$\delta_{g,l}^{Yn} = \frac{\delta_{g,l}^Y}{\delta_{g,l}^D},$$

<sup>14</sup>Callaway & Sant'Anna (2021) 和 Sum & Abraham (2021) 都对处理存在预期效应的方法进行了初步讨论。然而, 实际操作中, 关键在于识别预期效应持续的具体时期, 即需要明确处理组预期效应出现的具体时间段。随后, 以这个时间段作为开始接受处理的实际起点。

<sup>15</sup>此外, 在附录 A 部分, 笔者展示了如何将这此估计方法应用于二元交错处理变量的估计当中。

其中  $\delta_{g,l}^D = \sum_{k=0}^{l-1} |D_{g,F_g+k} - D_{g,1}|$  为  $g$  组从  $F_g$  期到  $F_g - 1 + l$  期累计接受的处理强度相较于基期的变化。由于  $\text{DID}_{g,l}$  可以识别  $\delta_{g,l}$ ,  $\delta_{g,l}^{Yn}$  可以通过  $\text{DID}_{g,l}/\delta_{g,l}^D$  得到。

$\delta_{g,l}^{Yn}$  实际上是  $g$  组在  $F_g - 1 + l$  期和前  $l - 1$  期处理效应的加权平均:

$$\delta_{g,l}^{Yn} = \sum_{k=0}^{l-1} w_{g,l,k} s_{g,l,k},$$

其中  $k$  期前的处理效应指的是  $t \leq F_g - 1 + l - k$  期的处理变量带来的处理效应, 具体可以表示为

$$s_{g,l,k} = \frac{E \left[ Y_{i,g,F_g-1+l}^{\mathbf{D}_{g,F_g-1}, D_{g,F_g}, \dots, D_{g,F_g-1+l-k-1}, D_{g,F_g-1+l-k}, \mathbf{D}_{g,1,k}} - Y_{i,g,F_g-1+l}^{\mathbf{D}_{g,F_g-1}, D_{g,F_g}, \dots, D_{g,F_g-1+l-k-1}, D_{g,1}, \mathbf{D}_{g,1,k}} \right]}{D_{g,F_g-1+l-k} - D_{g,1}},$$

即组  $g$  在  $F_g - 1 + l - k$  期的处理强度从  $D_{g,1}$  变为  $D_{g,F_g-1+l-k}$ , 之前的处理变量为实际值, 之后的处理变量为基期值的情况下, 在  $F_g - 1 + l$  时的潜在结果的变化与  $F_g - 1 + l - k$  期的处理强度的变化的比值。 $\mathbf{D}_{g,1,k}$  为一个每个元素均为  $D_{g,1}$  的  $k$  维向量。

而权重表示为

$$w_{g,l,k} = \frac{D_{g,F_g-1+l-k} - D_{g,1}}{\delta_{g,l}^D}.$$

据此, 权重与处理强度与基期处理强度的差值成正比, 处理强度变化越大则权重越大。

de Chaisemartin & D'Haultfoeuille (2024a) 同样对事件分析估计量  $\delta_l$  进行了如下标准化:

$$\delta_l^{Yn} = \frac{1}{N_l} \sum_{g:F_g-1+l \leq T_g} \frac{|\delta_{g,l}^D|}{\delta_l^D} \delta_{g,l}^{Yn} = \frac{\delta_l^Y}{\delta_l^D},$$

其中,  $\delta_l^D = \frac{1}{N_l} \sum_{g:F_g-1+l \leq T_g} |\delta_{g,l}^D|$ , 是处理强度增量的均值。

$\delta_l^{Yn}$  可以通过下式进行估计:

$$\text{DID}_l^{Yn} = \frac{1}{N_l} \sum_{g:F_g-1+l \leq T_g} \frac{|\delta_{g,l}^D| \text{DID}_{g,l}^Y}{\delta_l^D \delta_{g,l}^D} = \frac{\text{DID}_l^Y}{\delta_l^D}.$$

同  $\delta_{g,l}^{Yn}$  类似,  $\delta_l^{Yn}$  也可以理解为各组在  $F_g - 1 + l$  期和前  $l - 1$  期处理效应的加权平均, 其赋予  $k$  期前的处理效应的权重为:

$$\omega_{l,k} = \frac{1}{N_l} \sum_{g:F_g-1+l \leq T_g} \frac{|D_{g,F_g-1+l-k} - D_{g,1}|}{\delta_l^D}$$

当各组处理强度只变化一次的时候, 其他时候的处理变量强度均与基期相同, 只有  $w_{l,l-1} = 1$ , 那么  $\delta_l^{Yn}$  反应的就是第  $l - 1$  个处理变量的滞后期的处理效应, 这个解释是比较清晰的。但是在

其他情况下， $\delta_i^{Y^n}$  可能将不同的滞后期合并在一起，解释不再直观。

## C 平行趋势假设的检验

在实证分析中，常常通过动态回归模型 (10) 的估计结果来检验事前趋势，即对事前趋势估计参数  $\beta^l = 0$  ( $l < 0$ ) 的原假设进行假设检验。通常情况下，如果在回归模型 (10) 中，所有的  $\beta^l$  ( $l < 0$ ) 估计值均不显著不同于 0，则研究者认为平行趋势假设是成立的。根据 Sum & Abraham (2021) 的研究，二元交错处理变量中如果处理效应存在组间异质性，那么基于 TWFE 回归模型得出的事前趋势参数将是有偏的。然而，即使在处理效应具有同质性的条件下，依赖事前趋势进行平行趋势检验也存在问题。

已有文献对此问题有了丰富讨论。简单来说，Roth (2022) 提出目前事件分析法存在检验功效的不足 (Low Power) 等问题，提出了新的检测方法；Roth & Sant'Anna (2023) 指出，平行趋势假设对于变量形式的具有很高的敏感度；Rambachan & Roth (2023) 提出一些在平行趋势假设可能不满足的情况下可以进行的稳健性和敏感性的分析。接下来笔者将逐一介绍文献当中对事前趋势检验的局限性的讨论、检验可靠性的诊断以及稳健性和敏感性分析方法，结合实例和代码，以期为实证研究者提供指南。

### C.1 事前趋势检验的局限性

首先，Roth (2022) 认为，即使检验结果显示处理组和控制组在处理前的结果变量趋势是平行的，也不能保证处理后两者的趋势仍然平行。这意味着，仅通过检验事前趋势来推断事后趋势的方法本身存在一定的问题。Kahn-Lang & Lang (2020) 中提到一个例子：13 岁之前，男生和女生的身高变化趋势相对平行，但之后两者的差距逐渐扩大。如果要分析一个针对初中女生的健康政策的效果，使用男生作为控制组显然会导致估计结果有偏。这一点构成了对常用平行趋势检验的根本性挑战。之所以利用事前趋势来推断事后趋势，是因为后者在现实中往往无法直接观测；同时，这种方法的隐含假设是，不存在随时间变化并与接受处理条件相关的不可观测的混淆因素。对此，实证研究者不仅可以检验其他可能的相关变量在接受处理时是否同样发生变化，更应结合研究背景和经济学理论深入讨论这一假设的合理性。毕竟，相较于仅依赖事前趋势检验的统计显著性，将经济学理论应用于具体情境以深入分析平行趋势假设可能遇到的问题，更有说服力。

其次，文献当中指出，原假设为事前趋势估计参数等于 0 的假设检验具有较低的检验功效 (Bilinski & Hatfield, 2018; Freyaldenhoven et al., 2019; Kahn-Lang & Lang, 2020; Roth, 2022)。换言之，即使事前趋势存在差异，检验结果也可能无法检测出这一点。这是因为此类假设检验的原假设是不存在事前趋势差异，即平行趋势假设得到满足，因此只有在有显著证据时，平行趋势假设才会被认为不成立。这种检验将误判平行趋势假设不成立的概率 (第一类错误) 控制在 5% 以内 (假设显著性水平为 5%)，但允许将不成立的平行趋势假设误判为成立的概率 (第二类错误) 显著增加，从而降低了检测事前趋势存在的可能性。

再次，以通过事前趋势检验为前提来进行双重差分估计可能引入样本选择偏误。简而言之，如果全样本存在事前趋势差异，但研究样本恰好满足事前平行趋势假设，则估计结果可能存在偏差。Roth (2022) 发现，这种样本选择偏误可能比平行趋势假设不满足导致的偏误更为严重。

最后，如果通过事前趋势检验发现平行趋势假设不成立，是否就应该停止研究？理论上，只要提高检验的精度（降低假设检验的显著性水平），总能拒绝平行趋势假设成立的原假设。然而，在给定的事前趋势估计和事后处理效应估计的情况下，特别是在事前趋势差异虽然显著，但其参数的估计量与事后处理效应的估计量相比差异较大时，研究者仍希望能对处理效应提出一些建议性的结论。

当前文献对这些问题进行了探讨，提出了新的检测方法来应对检测功效低的问题，并在平行趋势假设可能不满足的情况下提供了进行稳健性和敏感性分析的方法。出于实用性的角度考虑，笔者接下来详细分析 Roth (2022) 和 Rambachan & Roth (2023) 中的方法，因为他们能够很好地与前文提到的那些处理效应异质性稳健的估计量进行结合。

## C.2 事前趋势检验可靠性的诊断

Roth (2022) 提出了两种基于事件分析法的“功效检验” (Power Test) 来估计结果。第一种方法是在给定一定的检验功效（通常为 80%）的条件下，计算一个线性事前趋势的斜率。这个斜率代表了在存在此类线性事前趋势的情况下，事前趋势检验有 80% 的概率能够成功检测到它，即检验的功效为 80%。第二种方法是，在已知某种事前趋势存在的情况下，计算事前趋势检验的功效，并在事件分析图中绘制假设的事前趋势以及虽存在但未被当前事前趋势检验发现显著的事前趋势。如果这些假定的事前趋势能解释事件分析图中观察到的事前趋势（即落在事前趋势检验的置信区间内），则表明当前的事前趋势检验的功效较弱。这些方法帮助研究者评估当前的事前趋势检验是否有足够的检测功效。Roth (2022) 的方法可以通过 STATA 命令和 R 包 *pretrends* 来实现。

## C.3 稳健性和敏感性分析

然而，Roth (2022) 提到的方法主要是对事前趋势检验本身进行检验和改进，并没有直接对处理效应的估计提供较大的启发。接下来，笔者将探讨在事前趋势检验不支持平行趋势假设的情况下，如何对处理效应进行估计。

Rambachan & Roth (2023) 提出可以通过事前趋势为处理效应提供边界估计。具体来说，这一方法的目的是，假设存在事后趋势但不超过事前趋势的  $\bar{M}$  倍时，计算出处理效应的置信区间，并由此来判断处理效应对于违反平行趋势假设的稳健性。<sup>16</sup> 例如，如果研究者报告，处理效应为正的结论在  $\bar{M} = 2$  的情况下依然稳健，就意味着要推翻非零处理效应的结论，事后趋势需要达到事前趋势的两倍以上。这种表述清楚地说明了，为了得到当前的结论，对平行趋势假设的违反需要达到何种程度。Rambachan & Roth (2023) 建议实证研究者应根据具体情境选择适当的回归模型形式（线性或非线性），根据不同的  $\bar{M}$  值报告稳健的置信区间，以及使处理效应评估不显著的  $\bar{M}$  的最小值或临界值。Rambachan & Roth (2023) 的方法可以通过 STATA 命令和 R 包 *honestdid* 来实现。

---

<sup>16</sup>如果  $\bar{M} = 1$ ，则要求事后趋势不能超过事前趋势本身。Rambachan & Roth (2023) 的方法也允许事前与事后趋势存在非线性差异。