

事件研究法的实现、问题和拓展

张子尧 黄 炜*

摘要:事件研究法是一种广泛应用于社会科学各领域的因果推断和政策评估工具。近年来事件研究法通常与双重差分法搭配使用,用于检验平行趋势假设和政策动态因果效应。本文结合前沿理论计量研究成果,试图对事件研究法进行系统性的归纳梳理,以厘清事件研究法在实践中需要注意的一些问题。本文梳理了事件研究法的历史发展脉络,详细展示事件研究法的实现方式、识别假设以及其与双重差分法之间的紧密联系。本文归纳分析了事件研究法实践中可能会面临的一些常见问题,包括基期选择、平行趋势检验、数据归并和截断、控制组间时间趋势差异等问题的影响和应对方法。本文还详细分析了异质性处理效应情形下的事件研究法估计偏误来源、修正思路、平行趋势检验等问题。

关键词:事件研究法 动态双重差分法 双向固定效应模型 异质性处理效应

中图分类号:F064.1 **文献标识码:**A **文章编号:**1000-3894(2023)09-0071-22

DOI:10.13653/j.cnki.jqte.20230725.003

一、引言

事件研究法(Event Study)又称事件分析法,是一种用于研究特定事件或政策冲击对个体行为影响的实证研究方法。事件研究法在早期主要应用于公司金融领域,例如分析股票拆分、信息披露、企业并购等事件对公司股票价格或超额收益率(Abnormal Return)的影响(Dolley, 1933; Ball和Brown, 1968; Fama等, 1969)。之后该方法逐渐向经济学和管理学等其他领域扩散,得到了更为广泛的应用(Ashenfelter, 1978; Laloned, 1986; Jacobson等, 1993)。随着计量经济学的“可信性革命”(Credibility Revelation)发展与政策评估研究的兴起(Angrist和Pischke, 2010),事件研究法在近年来的实证研究中占据着越来越重要的地位。Currie等(2020)统计了2010年后五本经济学国际权威期刊发表的论文中各类实证研究方法的使用情况,其中使用事件研究法的论文占比呈现出明显的增长趋势[见图1(a)]。在中文经济学权威期刊发表的论文中,使用事件研究法的实证研究类论文数量和占比同样开始快速增加,特别是从2017年起呈现爆发式增长[见图1(b)],事件研究法得到越来越多的社会科学领域研究者的广泛关注和应用。

事件研究法近年来的流行在很大程度上要归功于其与双重差分法(Difference-in-Differences, DID)之间的紧密联系。双重差分法已经成为经济学和管理学领域最为流行、应用最广泛的实证研究设计方法之一(Roth等, 2023; 黄炜等, 2022; 刘冲等, 2022)。双重差分法的核心识别假设是平行

* 张子尧,副教授,中南财经政法大学财政税务学院、收入分配与现代财政学科创新引智基地,电子邮箱: ziyao.zhang@zuel.edu.cn; 黄炜(通讯作者),副教授,北京大学国家发展研究院,电子邮箱: huangwei@nsd.pku.edu.cn。本文获得国家自然科学基金面上项目(72173131)和中南财经政法大学中央高校基本科研业务费专项资金(2722023BQ019)的资助。感谢匿名审稿专家的宝贵意见,文责自负。

趋势假设,即如果政策没有发生,处理组和控制组的结果变量具有相同的时间趋势。事件研究法能够直观地观察和检验政策发生前后个体行为的动态反应与组间差异,因此大量使用双重差分法的实证研究选择借助事件分析法来检验处理组和控制组在政策发生前是否具有相似的时间趋势,从而论证平行趋势假设是否成立。使用事件研究法来检验事前平行趋势已经成为目前双重差分法的标准步骤之一。伴随着双重差分法的大范围流行,事件研究法的身影也越来越多地出现。

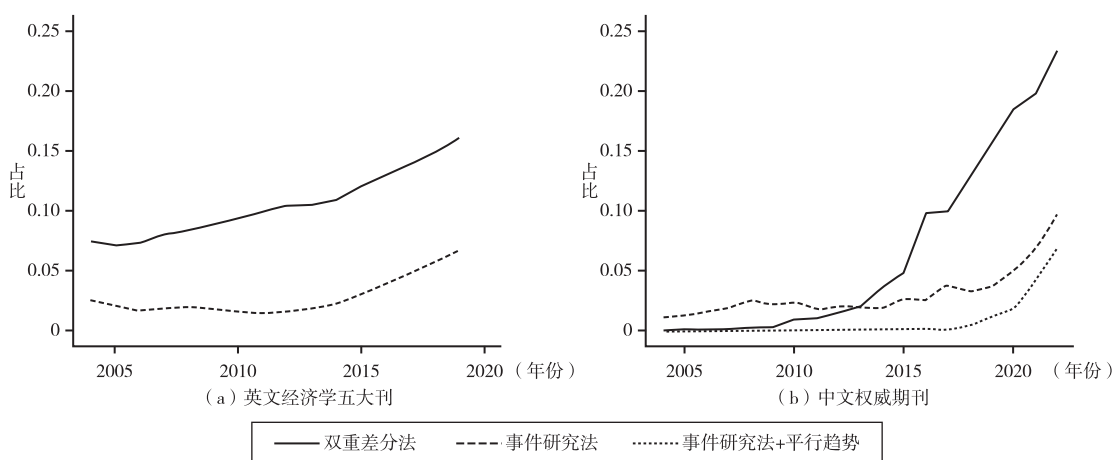


图1 双重差分法和事件研究法在中英文经济学权威期刊的发表占比

注:英文经济学五大刊分别为美国经济评论(American Economic Review)、经济学季刊(Quarterly Journal of Economics)、经济计量杂志(Econometrica)、政治经济学杂志(Journal of Political Economy)和经济研究评论(Review of Economic Studies);中文权威期刊包括《经济研究》《管理世界》《世界经济》《经济学(季刊)》《中国工业经济》《金融研究》《数量经济技术经济研究》《经济学动态》《财贸经济》9本经济学学术期刊。

数据来源:英文经济学五大刊数据来自Currie等(2020);中文权威期刊发表数据由笔者手工检索整理。

事实上,事件研究法的作用远不止检验事前平行趋势。相比于双重差分法,事件研究法具有很多有吸引力的性质。第一,事件研究法可以得到更加丰富的政策效应信息。双重差分法只能估计政策在样本期内的平均效应,但现实中一些政策效应存在滞后,需要施行一段时间才能发挥作用。并且,由于存在一般均衡效应,一项政策的短期和长期效果可能存在比较明显的差异。上述政策效应的动态特征是双重差分法无法说明的,而事件研究法则能以图形形式清晰地展示政策效应动态变化,这是事件研究法最吸引人的特征(Miller, 2023)。第二,事件研究法的识别假设相较于双重差分法更为宽松。近年来理论计量的新进展表明,对于政策交错发生的情形,若处理效应在组群或时间维度上存在异质性,双重差分法的估计结果存在偏误。事件研究法能够很好地应对时间维度上的异质性处理效应,比双重差分法的适用情景更为广泛。但是,目前国内的经济学实证研究中,大部分时候事件研究法仍然只是作为双重差分法的配套工具,处于一种辅助性的地位,没有完全发挥其本身的优势,不得不说是一种遗憾。这一方面可能是遵循传统双重差分法的实证研究步骤习惯,另一方面也可能是因为研究者没有深入、准确地理解事件研究法的内涵,忽略了事件研究法本身所独有的优势。此外,使用事件研究法的过程中有许多细节问题可能会引起研究者们的困扰,例如如何检验事前平行趋势、如何选择基期、对非平衡面板数据是否需要做归并或截断处理等。这些问题可能给研究者正确地使用事件研究法增添了障碍,但很少有文献专门讨论这些细节问题的影响和应对策略。本文试图对事件研究法的实现方式、识别假设和实践中的常见问题做系统性分析总结,

为社会科学各领域的研究者更好地发挥事件研究法在因果推断和政策评估方面的作用提供一些参考。

本文接下来的安排如下:第二部分对事件研究法的历史发展脉络进行梳理,明确早期公司金融领域使用的事件研究法和现代经济学实证研究使用的事件研究法之间的联系和区别。第三部分讨论事件研究法的实现方法和识别假设,并详细解释了双重差分法和事件研究法的联系、区别与应用场景。第四部分探讨了事件研究法在实践中的一些常见问题,包括如何选择基期、应该怎样检验和解读事前平行趋势、归并和截断数据对估计结果有何影响以及如何正确控制组间异质性时间趋势等。第五部分详细解释了异质性处理效应情形下的事件研究法估计偏误原因、各类稳健估计量的修正方法以及平行趋势检验等问题。第六部分是总结性评论。

二、事件研究法的历史发展脉络

(一)事件研究法的早期发展

事件研究法最早可以追溯到1933年Dolley关于股份分拆对股票价格影响的研究。Dolley(1933)通过比较股份拆分前后的股价变动发现,大部分股份拆分事件发生后股票价格会上升。之后许多研究者继承了Dolley(1933)的思想,对事件研究法的实现方法和技术细节进行改进升级。得益于Ball和Brown(1968)以及Fama等(1969)在资本市场研究中对事件研究法的出色应用,早期的事件研究法主要被应用于金融学领域关于股票超额回报率的研究(Peterson, 1989; Armitage, 1995; MacKinlay, 1997)。这类事件研究法的研究步骤通常为:首先定义事件发生前后可能对股票价格产生影响的时间段(事件窗口),然后计算事件窗口内的股票超额回报率,最后比较事件发生前后的股票超额回报率变动来识别事件本身对股票价格的影响。^①

早期的事件研究法没有引入未受处理的样本作为对照,仅比较个体在事件发生前后的自身差异来识别因果效应。从研究设计的角度看,这种做法非常类似于以时间作为驱动变量的断点回归设计(Imbens和Lemieux, 2008; Lee和Lemieux, 2010):以事件发生时点为断点,通过比较断点两侧结果变量的变化来识别事件本身对结果变量的因果效应。一些经济学领域的研究同样延续了这种思路,例如黄炜等(2023)通过比较个人在法定退休年龄前后的行为动态变化来识别退休行为对劳动供给、收入和消费的影响。因此,这类事件研究法的识别假设与断点回归设计的要求相同,即在给定相关控制变量的条件下事件发生时点与结果变量是不相关的(Dobkin等, 2018)。该识别假设可以直观地理解为在给定控制变量的条件下,事件没有发生时的不可观测反事实潜在结果与可观测的现实潜在结果相同。有许多情形可能违反上述识别假设。最常见的情形是存在时变混杂因素,例如同期发生的其他冲击或存在时间趋势。对于可观测的时变混杂因素,可以进行类似于断点回归设计的平衡性检验,观察相关协变量在事件时点发生前后是否保持连续。如果相关协变量在事件发生前后确实发生了跳跃,可以通过控制相关协变量的手段剔除时变混杂因素的影响。不过,这种处理方式对不可观测的时变混杂因素无能为力。另一种常用的手段是选择一个较为狭窄的事件窗口以确保时变混杂因素在事件窗口内保持稳定从而满足识别假设。但是,缩短事件窗口带来的一个挑战是对数据颗粒度有更高的要求,很多情形下需要日度甚至分钟级别的高频数据。这无疑大大限制了该方法的适用范围。

(二)事件研究法在现代经济学实证研究中的应用

随着事件研究法的广泛应用,该方法的使用范围逐步开始突破金融学领域,开始被其他领域的

^① MacKinlay(1997)是关于事件研究法早期应用的杰出综述。

研究者重视和利用,而这些应用又促进了事件研究法本身的进步和突破。从20世纪80年代开始,经济学实证研究中的“可信性革命”逐渐兴起,基于Rubin潜在结果框架的因果推断方法成为经济学实证研究的主流范式。研究者们从研究设计的视角重新审视和理解事件研究法,特别是将事件研究法和双重差分法联系了起来。现代事件研究法一个重要的标志是引入未受处理的控制组(Ashenfelter, 1978)。在新的研究设计中,事件研究法不再是简单比较个体行为在事件发生前后的变化,而是比较处理组和控制组个体在事件发生前后行为变化的差异。通过引入未受处理的控制组以剔除时间趋势和宏观经济周期变动的影响,从而克服早期事件研究法的潜在缺陷。显然,这种思路非常类似于双重差分法(Angrist和Pischke, 2009),近年来的一系列理论计量研究表明,基于多期面板数据和双向固定效应模型的事件研究法和双重差分法之间确实存在着紧密的内在联系,事件研究法的估计结果可以表述为一系列基本的“两组一两期”双重差分法结果的加权平均(Sun和Abraham, 2021; Borusyak等, 2023)。因此,现代的事件研究法本质上是对于传统双重差分法在多期面板数据结构和多个处理时点的研究情景下的一种拓展。正是由于这种紧密的联系,事件研究法有时也会被称为动态双重差分法(Dynamic DID)。

事件研究法近年来的广泛流行在很大程度上要归功于它与双重差分法之间的紧密联系。双重差分法已经成为目前最为流行的因果推断和政策评估工具(Goodman-Bacon, 2021; 黄炜等, 2022)。双重差分法的有效性建立在平行趋势(Parallel Trends)和无预期效应(No Anticipatory Effects)两个假设成立的基础上。平行趋势假设是指若政策没有发生,处理组和控制组的潜在结果应具有相同的时间趋势。由于政策未发生的反事实潜在结果不可观测,无法直接检验平行趋势假设是否成立,研究者们通常的做法是退而求其次,检验在事件发生前处理组和控制组的时间趋势是否平行。无预期效应假设是指处理组不会在事件发生之前提前做出反应。无预期效应假设可以通过观察事件发生前处理组相对于控制组是否发生了显著的行为变化进行检验。事件研究法可以检验处理组和控制组一段时间内的组间差异动态变化,恰好可以用来检验事前趋势是否平行以及是否存在预期效应。在当前的双重差分法实证研究中,使用事件研究法检验事前平行趋势假设和无预期效应假设已经成为一种“标准化”操作。随着研究者们对事件研究法的理解越来越深入,其在实证研究中的作用开始发生变化。事件研究法不仅能够用于检验事前平行趋势是否满足,还能够评估处理效应随时间推移产生的变化,帮助理解政策在中长期的影响。近几年许多发表在顶级期刊的论文中,事件研究法不再仅是作为一种辅助工具,而是成为主要的因果推断和政策评估工具之一(Dobkin等, 2018; Dustmann等, 2022)。

三、事件研究法的实现方法和识别假设

(一)基于双向固定效应模型的事件研究法实现

假设数据结构为平衡面板数据,其中包含个体 $i \in \{1, \dots, N\}$,每个个体可以观测到 $t \in \{1, \dots, T\}$ 共 T 期,样本量为 $N \times T$ 。假设每个个体接受处理的时间点为 T_i^* ,可以定义每个个体相对于处理时点 T_i^* 的相对时间变量 $D_{it}^l = \mathbb{I}[t - T_i^* = l]$,其中 \mathbb{I} 为示性函数,满足条件时取1;否则取0。例如个体 i 在 t 期距接受处理还有两期($t - T_i^* = -2$)时 D_{it}^{-2} 取1,其余相对时间变量 $D_{it}^{k \neq -2}$ 取0。^①

在实践中研究者们通常采用如下形式的计量方程设定,使用双向固定效应模型(Two-Way Fixed Effects, TWFE)实现事件研究法:

^① 从未接受处理的控制组个体(Never-taker)的处理时点定义为 $T_i^* = \infty$,其相对时间变量 D_{it}^l 恒取0值。

$$y_{it} = \alpha_i + \lambda_t + \sum_{l=-K}^{-2} \mu_l D_{it}^l + \sum_{l=0}^L \beta_l D_{it}^l + \epsilon_{it} \quad (1)$$

式(1)中 α_i 为个体固定效应, λ_t 为时间固定效应。相对时间变量 D_{it}^l 被分为了两组,第一组代表处理发生前($l \leq -2$),第二组代表处理发生后($l \geq 0$)。有两点需要额外说明:第一,为了避免与固定效应完全共线,不能将全部的相对时间变量 D_{it}^l 加入到回归方程中,需要手动丢弃一个变量作为基期。^①第二,对于 T 期的面板数据,假设最早和最晚的处理时点分别为 T_i^{*min} 和 T_i^{*max} ,则理论上最长的事件窗口为 $[-K, L]$,其中 $K = T_i^{*max} - 1, L = T - T_i^{*min}$ 。如果式(1)中包含全部的处理前和处理后相对时期变量,这种模型设定被称为“完全动态设定”(Fully-dynamic Specification)。在实际研究中,研究者们通常更为关心在事件发生前后一段时期内的动态效应,此时可能对相对处理时点过早或过晚的样本进行归并(Binning)或截断(Trimming)处理。

接下来本文使用模拟数据演示如何使用回归方法实现事件研究法。^②模拟样本中共有400个个体,每个个体有20期观测值。样本中共包含4个组群(Group),分别为3组在不同时点接受处理的实验组和1组未受处理的控制组,每组均包含100个个体。3组实验组分别在第5、10、15期接受处理。实验组个体的处理效应随时间增长,受处理期限每增加1期处理效应随之增加1单位。图2描述了模拟数据中的结果变量 Y 的组群均值变化[见图2(a)]和每个组群的真实平均处理效应[见图2(b)]。从图中可以看到样本数据的两个重要特征,一是处理交错发生(Staggered Adoption),二是处理效应具有动态性,但各组群拥有相同的处理效应路径(Treatment Effects Path)。

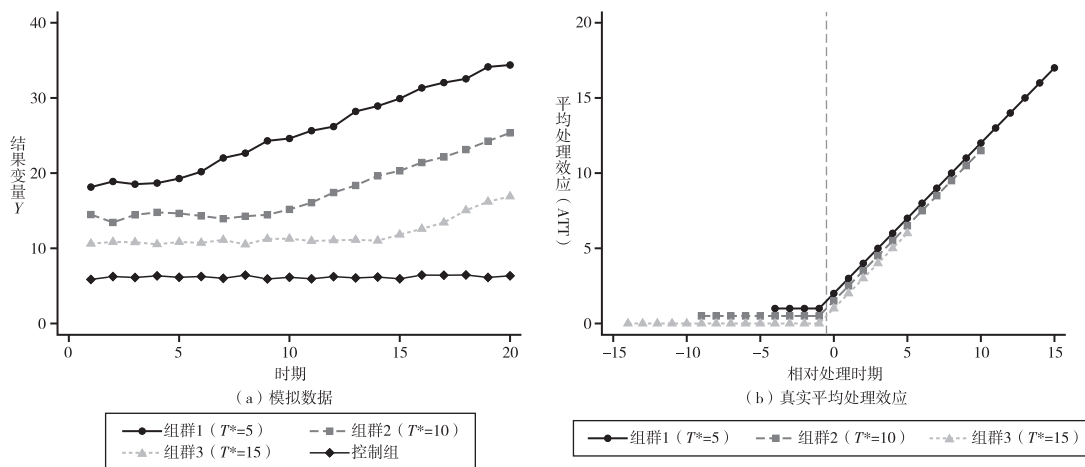


图2 模拟数据与真实处理效应

注:图b中为了区分不同的组群,分别将组群1和组群2的曲线向上平移了1个和0.5个单位。

根据模拟数据的基本结构,可知最早的相对处理时期是-14期,最晚的相对处理时期是15期,可以构建完全动态回归方程:

① 若样本中不存在从未接受处理的控制组个体,由于存在其他的共线性约束,还需要额外丢弃一个相对时间变量。感兴趣的读者可阅读 Borusyak 等(2023)的3.2节 Proposition 1 部分。

② Baker 等(2022)以及 Sant'Anna 和 Callaway 的个人网站(见 <https://psantanna.com/posts/twfe>)同样使用数据模拟方法分析了异质性处理效应情形下交错型 DID 的估计偏误。本文部分参考了他们的数据模拟方法思路,但分析重点侧重于事件研究法在各类情形下的适用性以及实践过程中需要注意的一系列特定问题。

$$y_{it} = \alpha_i + \lambda_t + \sum_{l=-14}^{-2} \mu_l D_{it}^l + \sum_{l=0}^{15} \beta_l D_{it}^l + \epsilon_{it} \quad (2)$$

图3描述了式(2)中 μ_l 和 β_l 的估计结果。可以看到, 处理发生前的估计系数 μ_l 均在0值附近且联合检验不显著, 表明在处理发生前处理组和控制组的组间均值差异没有发生变化, 满足事前平行趋势。对比处理发生后的估计系数 β_l 和图2中的真实平均处理效应, 可以发现 β_l 能够准确识别每个相对时期的平均处理效应 ATT_l 。

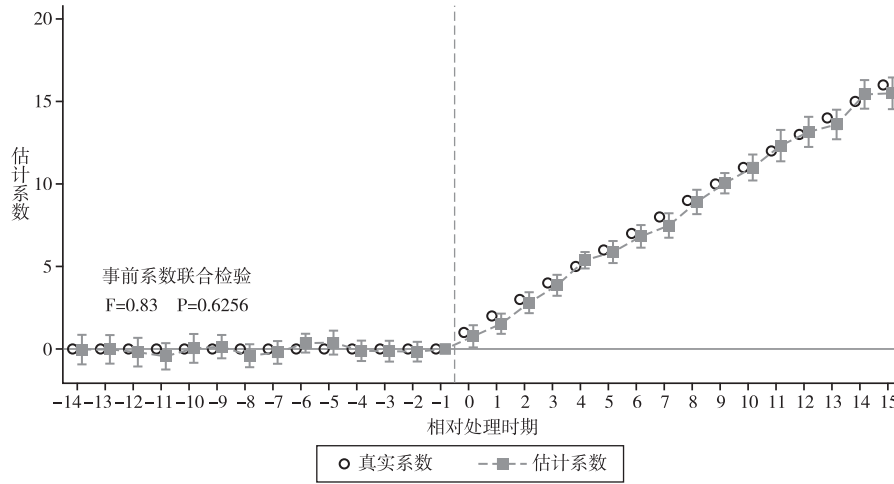


图3 事件研究法的估计结果

(二) 事件研究法的识别假设

Sun 和 Abraham (2021) 证明估计系数 μ_l 和 β_l 可以分解为如下形式:

$$\begin{aligned} \mu_l(\beta_l) = & \sum_{g=1}^G \omega_g^l \left(\underbrace{E[Y_{i,T_g^*+l} - Y_{i,0} | T^* = T_g^*]}_{\text{组群}g\text{的前后变化}} - \underbrace{E[Y_{i,T_g^*+l}^C - Y_{i,0}^C]}_{\text{控制组的前后变化}} \right) \\ & + \sum_{l' \neq l} \sum_{g=1}^G \omega_{g,l'}^l \left(E[Y_{i,T_g^*+l} - Y_{i,0} | T^* = T_g^*] - E[Y_{i,T_g^*+l}^C - Y_{i,0}^C] \right) \\ & + \sum_{l' \in l^{excl}} \sum_{g=1}^G \omega_{g,l'}^l \left(E[Y_{i,T_g^*+l} - Y_{i,0} | T^* = T_g^*] - E[Y_{i,T_g^*+l}^C - Y_{i,0}^C] \right) \end{aligned} \quad (3)$$

第一行括号内是组群 g 与控制组从0期到 $T_g^* + l$ 期的时间趋势差异, T_g^* 是组群 g 的处理时点。第二行括号内是对于除 l 期之外的其他相对时期 l' 期,组群 g 与控制组从0期到 $T_g^* + l'$ 期的时间趋势差异。第三行的括号内是回归中被排除掉的第 l^{excl} 期(一般为基期)组群 g 与控制组从0期到 $T_g^* + l^{excl}$ 期的时间趋势差异。式(3)中的 ω_g^l 指组群 g 在第 l 期的权重,Sun 和 Abraham (2021)发现式(3)中的各项权重满足以下条件:第一,相对时期 l 的各组群权重和为1,即 $\sum_{g=1}^G \omega_g^l = 1$ 。第二,除 l 期外的其他相对时期 l' 的权重和为0,即 $\sum_{g=1}^G \omega_{g,l'}^l = 0$ 。第三,回归方程中被排除的基期 $l' \in l^{excl}$ 权重和为-1,即

$\sum_{l' \in l^{excl}} \sum_{g=1}^G \omega_{g,l'}^l = -1$ 。从式(3)中可以很清晰地看到事件研究法与双重差分法之间的内在联系:每一个

括号内其实都是一个由第 $T_g^* + l$ 期的处理组和同期控制组构成的 2×2 双重差分估计,因此事件研究法的估计系数是一系列基本 2×2 双重差分估计量的加权平均。接下来借助式(3)的分解结果逐步分析事件研究法所需要的识别假设。

识别假设1:平行趋势假设(Parallel Trends Assumption)。对于任意两个时期 $s < t$,处理组和控制组未受处理的潜在结果(记为 $Y_{i,t}^0$)具有相同时间趋势,即对任意两个组群 g_p 和 g_q 均有

$$E[Y_{i,t}^0 - Y_{i,s}^0 | i \in g_p] = E[Y_{i,t}^0 - Y_{i,s}^0 | i \in g_q] \quad (4)$$

若平行趋势假设成立,式(3)的括号内处理组和控制组前后差异的差异(双重差分)可以正确识别组群 g 在相对时期 l 的平均处理效应:

$$E[Y_{i,T_g^*+l} - Y_{i,0} | T^* = T_g^*] - E[Y_{i,T_g^*+l}^C - Y_{i,0}^C] = ATT_{g,l} \quad (5)$$

此时估计系数可以分解为三部分:

$$\mu_l(\beta_l) = \sum_{g=1}^G \omega_g^l ATT_{g,l} + \sum_{l' \neq l} \sum_{g=1}^G \omega_{g,l'}^l ATT_{g,l'} + \sum_{l' \in l^{excl}} \sum_{g=1}^G \omega_{g,l'}^l ATT_{g,l'} \quad (6)$$

式(6)中, $ATT_{g,l}$ 指组群 g 在相对时期 l 的平均处理效应。第一项是相对时期 l 的各组群平均处理效应 $ATT_{g,l}$ 的加权平均,第二项是除 l 期之外的其他相对时期 l' 的各组群平均处理效应 $ATT_{g,l'}$ 的加权平均,第三项是回归中排除掉的 l^{excl} 期(一般为基期)的各组群平均处理效应 $ATT_{g,l'}$ 的加权平均。

识别假设2:无预期效应假设(No Anticipatory Assumption)。任意未发生处理时期 l 的平均处理效应均等于0。即对任意组群 g 和处理前时期 $t < T_g^*$ 均有

$$E[Y_{i,t} - Y_{i,t}^0 | i \in g, t < T_g^*] = 0 \quad (7)$$

若无预期效应假设成立,选取处理前时期(如-1期)作为基期时式(6)可进一步化简为

$$\mu_l(\beta_l) = \sum_{g=1}^G \omega_g^l ATT_{g,l} + \sum_{l' \neq l} \sum_{g=1}^G \omega_{g,l'}^l ATT_{g,l'} \quad (8)$$

式(8)表明,即使满足平行趋势假设和无预期效应假设,事前估计系数 μ_l 或事后估计系数 β_l 会受到其他相对时期 l' 的平均处理效应的影响,仍然不能正确估计当期的平均处理效应 ATT_l 。^①基于双向固定效应模型的事件研究法估计系数能否正确识别相对时期 l 的平均因果效应取决于真实平均处理效应的具体形式。根据处理效应在组群和时间两个维度是否存在异质性,可以将处理交错发生条件下的真实处理效应分为三类:

①同质性处理效应(Homogeneous Treatment Effects)。如图4(a)所示,对于任意组群 $g \in G$ 和任意处理后时期 $l \geq 0$,平均处理效应均相同,即

$$ATT_{g,l} = ATT \quad g \in G, \quad l \in \{0, 1, \dots, L\} \quad (9)$$

②同质性处理效应路径(Homogeneous Treatment Effects Path)。如图4(b)所示,平均处理效应会随着时间推移产生动态变化,但任意组群的平均处理效应变化路径均相同,即

$$ATT_{g,l} = ATT_l \neq ATT_{l'} \quad g \in G, \quad l, l' \in \{0, 1, \dots, L\} \text{ 且 } l \neq l' \quad (10)$$

③异质性处理效应(Heterogeneous Treatment Effects)。如图4(c)所示,平均处理效应会随着时间推移产生动态变化,并且不同组群的平均处理效应变化路径存在差异,即

$$ATT_{g,l} \neq ATT_{g',l} \text{ 且 } ATT_{g,l} \neq ATT_{g',l'} \quad g, g' \in G, \quad l, l' \in \{0, 1, \dots, L\} \quad (11)$$

① 按照 Goldsmith-Pinkham 等(2022)的观点,若一个线性回归中存在多个相互不独立的政策变量,特定政策变量的估计系数不仅会捕捉到自身的效应,还会部分地捕捉到其他政策的效应。如果政策效果存在异质性,可能导致估计系数出现“污染偏误”(Contamination Bias)。式(3)的分解结果可以看作一种特定形式的“污染偏误”。

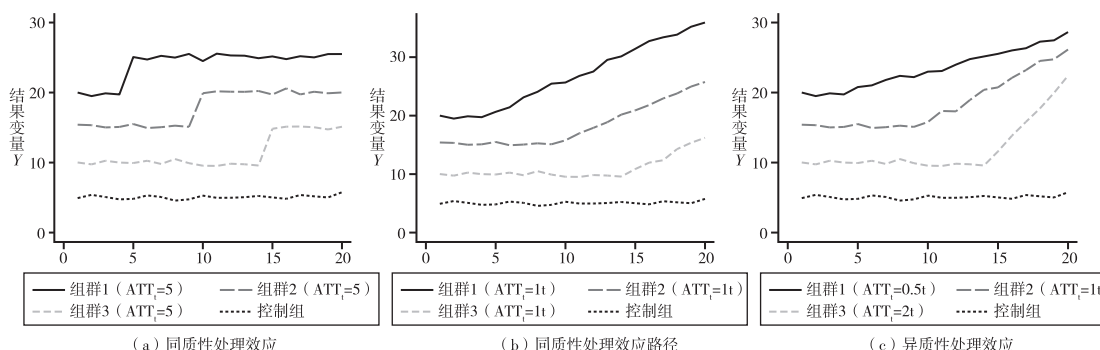


图4 三类处理效应假设示意

识别假设3:同质性处理效应路径假设(Homogeneous Treatment Effects Path Assumption)。处理效应满足同质性处理效应或同质性处理效应路径假设。

若同质性处理效应路径假设成立,式(8)可以进一步简化为如下形式:

$$\mu_t(\beta_t) = \sum_{g=1}^G \omega_g^l ATT_{g,l} + \sum_{l' \neq l} \sum_{g=1}^G \omega_{g,l'}^l ATT_{g,l'} = ATT_t \underbrace{\sum_{g=1}^G \omega_g^l}_{=1} + \sum_{l' \neq l} ATT_{l'} \underbrace{\sum_{g=1}^G \omega_{g,l'}^l}_{=0} = ATT_t \quad (12)$$

式(12)表明,当同时满足平行趋势假设、无预期效应假设和同质性处理效应路径假设三个识别假设时,基于双向固定效应模型的事件研究法能够正确估计各个相对时期的平均处理效应。如果识别假设不满足,特别是同质性处理效应路径假设不满足时,事件研究法的估计系数无法正确识别各个时期的平均处理效应。

(三)事件研究法与双重差分法的比较

在拥有多期面板数据的情形下,实证研究中使用更为普遍的识别策略是双重差分法:

$$\begin{aligned} \text{单一政策时点: } y_{it} &= \alpha_i + \lambda_t + \beta treat_i \times post_t + \epsilon_{it} \\ \text{多个政策时点: } y_{it} &= \alpha_i + \lambda_t + \beta D_{it} + \epsilon_{it} \end{aligned} \quad (13)$$

当政策在单一时点施行时, $treat_i$ 表示个体是否属于受政策影响的处理组, $post_t$ 表示政策是否发生,交互项系数 β 是双重差分估计量,衡量政策平均处理效应。当政策在多个时点逐步施行时,无法直接以交互项形式表示双重差分估计,此时可以用代表个体 i 在第 t 期是否受到政策影响的虚拟变量 D_{it} 代替交互项,系数 β 是交错双重差分形式下的政策平均处理效应估计。事实上,对比式(13)和式(1)的模型设定可以很清楚地发现,在满足平行趋势假设和无预期效应假设的前提下,如果对事件研究法施加事后各期平均处理效应相等的约束条件($\beta_0 = \dots = \beta_t$),式(1)形式的事件研究法就会退化为式(13)形式的双重差分法(Miller, 2023),因此事件研究法可以看作双重差分法的一般化。

事件研究法和双重差分法最重要的区别是二者关于处理效应的识别假设不同。双重差分法只有在满足处理效应在组群和时间维度上均不存在异质性(同质性处理效应)时才能够正确估计平均处理效应(Goodman-Bacon, 2021),而事件研究法允许处理效应在时间维度上存在动态变化,只要这种动态变化在组群间保持一致(同质性处理效应路径)即可正确估计各时期的平均处理效应。考虑到现实中绝大部分政策都需要一段时间才能逐渐发挥全部作用,处理效应在时间维度上往往呈现出动态变化特征,此时双重差分法的估计结果有偏,事件研究法则能一致估计政策效应。本文建议,在政策交错发生的实证研究中应该首先使用事件研究法分析并展示相应的图形结果,观察政策

效应是否存在动态特征。如果政策效应不存在明显的动态特征,可以使用双重差分法估计政策在整个样本期内的平均处理效应;如果政策效应呈现出显著的动态性特征,由于双重差分法估计有偏,研究者需要手动计算样本期内的政策平均处理效应。^①

四、事件研究法实践中的若干问题

(一)基期选择

基于双向固定效应模型的事件研究法在估计时需要丢弃一个相对时期变量以避免共线性问题,这个被丢弃的相对时期一般被称为基期(Base Period)。在实践中,研究者们通常选择处理前一期(-1期)或事件窗口的第一期(-K期)作为基期,根据式(6),在满足平行趋势假设、无预期效应假设和同质性处理效应路径假设三个识别假设的前提下,事件研究法第 l 期的估计系数等于当期平均处理效应 ATT_l 与被排除在回归模型外的第 l' 期(基期)平均处理效应 $ATT_{l'}$ 之差:

$$\mu_l(\beta_l) = ATT_l - ATT_{l'} \quad (14)$$

无预期效应假设意味着处理发生前时期($l < 0$)的平均处理效应为0,只要选择任意处理前时期 $l' < 0$ 作为基期,系数 μ_l 和 β_l 就能够正确估计当期平均处理效应 ATT_l 。反之,如果选择处理后时期中的任意一期作为基期,此时系数 μ_l 和 β_l 估计的是第 l 期和基期的平均处理效应之差,而非当期的平均处理效应本身。因此,必须选择处理前时期作为基期。那么,在处理前时期中应该选取哪一期作为基期呢?若无预期效应假设满足,选取任一处理前时期作为基期均可,此时建议选择处理前一期(-1期)作为基期,可以更为清晰直观地观测到处理前时间趋势是否稳定以及处理发生后是否发生了明显的变化。但是,若无预期效应假设不满足,距离处理时点较近的处理前时期(如-1期)的平均处理效应可能不为0,此时若使用这些时期作为基期就会导致估计结果出现偏误。本文建议,研究者可首先选择不太可能发生预期效应的较为靠前的相对时点(如-3期)作为基期,如果估计结果显示-1期没有出现显著变化,说明预期效应可能并不明显,此时可以将基期更换为-1期来更加清晰地展示处理发生前后的结果变量变化情况。

虽然采用提前基期的方法可以检验和规避预期效应可能导致的估计偏误,但如果在实践中确实发现了个体对处理提前反应的经验证据,更为重要的问题是探究为何个体会提前做出反应。如果这种提前反应行为还涉及是否接受处理和接受处理具体时点的选择性,这可能预示着处理组和控制组之间不具有可比性,使得平行趋势假设的可信性受到威胁。因此,在实践中仅仅更换靠前的基期来解决预期效应问题是不充分的,更重要的是对产生预期效应的根本原因和可能后果做出充分的解释和稳健性检验。

除了选取某一期作为统一基期(Universal Base Period)的做法外,还有另一种被称为可变基期(Varying Base Period)的基期选取方法。^②具体而言,处理前第 k 期的估计系数是基于 k 和 $k-1$ 两期数据、以 $k-1$ 作为基期的双重差分估计结果;对于处理后各时期,统一使用-1期作为基期。当选择可变基期时,处理前第 k 期估计系数代表着一种“伪平均处理效应”(Pseudo-ATTs),若其不等于0则意味着事前平行趋势不满足或存在预期效应。可变基期的优点在于检验预期效应时不依赖于具体的基期选择,但缺点在于其无法直观呈现事前组间趋势差异;传统事件研究法的优缺点恰好与之

^① 一种简单的办法是对处理后各期系数求简单平均,可以方便地使用stata中的lincom命令实现。另一种方法考虑样本权重,以各相对时期出现的样本数量为权重进行加权平均。

^② 详细内容见 <https://bcallaway11.github.io/posts/event-study-universal-v-varying-base-period>。

相反,其优点在于可以直观呈现组间趋势差异,但缺点是事前各期系数的单一显著性依赖于基期选择(见下节)。研究者可根据具体的研究情景和需要选择更为适合的做法。

(二)对事前平行趋势假设的检验

实践中研究者们一般会通过关注事前估计系数 μ_i 是否接近0来判断处理前控制组和处理组的组间均值变化是否满足事前平行趋势。然而,如何判断事前平行趋势是否满足,至今仍然没有较为公认的标准和做法。一种做法是观察处理前每一期系数的显著性,如果全部事前估计系数 μ_i 都不显著,则认为满足事前平行趋势。但是,这种做法面临着两个问题:第一,事前平行趋势要求处理前每一期的平均处理效应为0,其对应的原假设是 $\mu_{-2} = \mu_{-3} = \dots = \mu_{-K} = 0$,需要进行系数联合检验而不是关注单一系数的显著性。全部系数都不显著并不意味着事前平行趋势检验一定通过,有个别系数显著也并不代表事前平行趋势检验一定不通过。第二,单一估计系数 μ_i 的显著性和基期的选择有关,结论不稳健。更理想的做法是对事前系数显著性进行联合检验。

图5的模拟结果展示了关注单一系数显著性和联合检验显著性的区别和优劣。图5(a)分别估计了-1期和-8期作为基期时的事件研究法结果。可以看到,基期变化仅仅改变了“参照点”,所有的估计系数随着“参照点”的变化而整体向上或向下平行移动,单一估计系数的显著性也随之改变。选择-1期为基期时,处理前的所有相对时期系数都不显著;但若选择-8期为基期, μ_{-5} 和 μ_{-6} 在5%的水平上显著。从逻辑上看,改变“参照点”不会改变处理组和控制组的组间时间趋势,不应该对平行趋势检验的结果产生影响,显然根据单一估计系数的显著性与否来判断事前平行趋势是否成立的做法不能得到稳健、一致的结论。从实践角度看,如果以单一系数显著性作为事前平行趋势是否成立的判断标准,会存在人为操纵事前平行趋势检验的空间,研究者们可以选择更为接近处理前估计系数均值的基期[如图5(a)的-1期]而避开较为极端的基期[如图5(a)的-8期],人为地降低事前各期估计系数的显著性,使得事前平行趋势检验更容易被通过。总的来看,单一系数显著性检验既不符合事前平行趋势检验的本质,又存在人为操纵的可能性,因此不能成为检验事前平行趋势是否成立的良好标准(Roth, 2022)。就目前而言,更为合理的做法是对事前估计系数的显著性进行联合检验。联合检验不仅更加符合事前平行趋势的本质,其另一个不可忽视的优势在于使联合检验的结果不受基期选择的影响。图5(b)使用相同的数据和模型进行事前估计系数联合检验,模拟结果清晰地表明无论选择处理前哪一期作为基期,都不会改变事前估计系数的联合检验结果。因此,联合检验作为事前平行趋势检验的评判标准具有更好的一致性和稳健性。

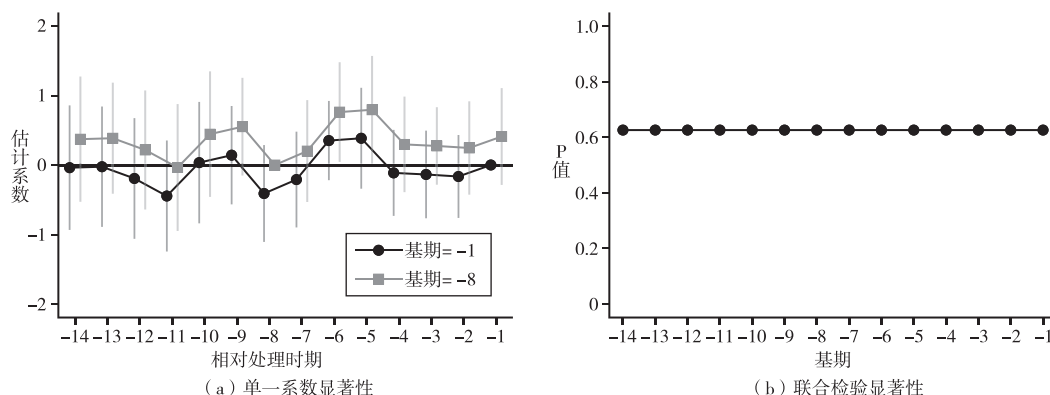


图5 改变基期对事前平行趋势检验结果的影响

在实践中,如果事前估计系数的联合检验通过,但估计系数本身体现出某种明显的时间趋势,此时应该对平行趋势检验的结果保持高度的小心谨慎。Roth(2022)分析了事前平行趋势检验的统计效力等问题,发现联合检验一个潜在问题是统计效力较低。更为重要的是,他们的分析进一步说明如果组间差异确实存在某种时间趋势,在给定事前估计系数的联合检验没有拒绝原假设(统计意义上的第二类错误)的情况下,平均处理效应的估计偏误要更为严重。^①

综上所述,本文对于实践中的事前平行趋势检验提出三点建议:第一,必须更加关注事前估计系数是否呈现出明显的时间趋势。个别事前估计系数显著可能仅仅是一种数据噪声导致的偶然现象,然而如果存在明显的事前趋势,可能预示着平行趋势假设不成立,这会直接威胁到事件研究法的识别假设可靠性,可能会导致估计结果产生严重偏误。因此,事前时间趋势(即使不显著)是比个别事前系数显著严重得多的大问题。若估计结果提示可能存在这一类问题,必须对该问题的来源进行详细的分析和讨论,并尽可能地做针对性处理以减轻对估计结果的干扰。第二,由于事前时间趋势是关注重点,除了某些特殊情况外,必须对估计结果进行图形化以直观判别是否存在事前时间趋势。第三,检验事前平行趋势应该对事前估计系数做联合检验而不是关注单一系数的显著性。

(三)归并和截断数据

对于平衡面板数据,当个体接受处理的时点不同时,相对处理时期的样本分布是不平衡的,在比较靠前或靠后的相对时期可能只有非常少的观测值,因此研究者们通常会对数据做一些预处理。实践中主要有数据归并和数据截断两种处理方式。数据归并(Binning)是指设定一个事件窗口 $[-K, L]$,将早于事件窗口的观测值($l \leq -K$)的相对时期重新设定为 $-K$ 期,将晚于事件窗口的观测值($l \geq L$)的相对时期重新设定为 L 期。数据截断(Trimming)同样是设定一个事件窗口,区别在于将事件窗口外的样本做删除而非归并。

归并数据的优势有二:一是在样本中没有从未接受处理的控制组时可以避免共线性问题以识别平均处理效应(Schmidheiny 和 Siegloch, 2022),二是保留了全部样本从而提高了估计效率。但其缺点在于需要施加一些额外的假设条件,否则可能导致比较明显的估计偏误。Sun 和 Abraham (2021)证明,只有当被归并在一起的相对时期具有相同的平均处理效应时,归并后的事件研究法估计结果才会具有一致性。然而在现实中绝大多数的处理效应都具有动态性,各个相对时期的处理效应并不相等,此时做归并处理可能会存在导致较为明显的估计偏误。图6的模拟数据说明了这一问题:图6(a)中对-5期之前的样本做归并处理,由于处理前各期的平均处理效应相等(均为0),归并处理并未对估计结果造成显著影响,各个估计系数仍然非常接近真实系数;图6(b)中对第5期之后进行归并,由于真实处理效应具有动态性,被归并的各个相对时期平均处理效应不相等,此时事前估计系数和事后估计系数均产生了严重偏误。因此,在实践中应该避免对处理发生后时期进行归并;若事前平行趋势满足,可以对处理前时期做适当归并。

截断数据的优势在于不需要像归并数据那样施加额外的同质性处理效应假设,从图6(b)可以看到,无论是处理前还是处理后时期截断样本,产生的估计偏误都要比归并数据小得多。但是,截断数据面临着其他的风险:第一,截断数据意味着非随机地删除掉了一部分样本,这可能会引起一定的样本选择问题,保留下的数据不再具有样本代表性(Borusyak 等, 2023)。第二,截断数据会减少样本量,降低估计效率。对比图6中归并和截断数据的估计结果,可以看出截断数据时估计系数的置信区间明显要更宽一些,这是因为截断数据会减少样本数从而放大标准误、降低估计精度。在

^① Rambachan 和 Roth(2023)提出了对平行假设不成立时可能造成的偏误程度进行量化评估的方法。

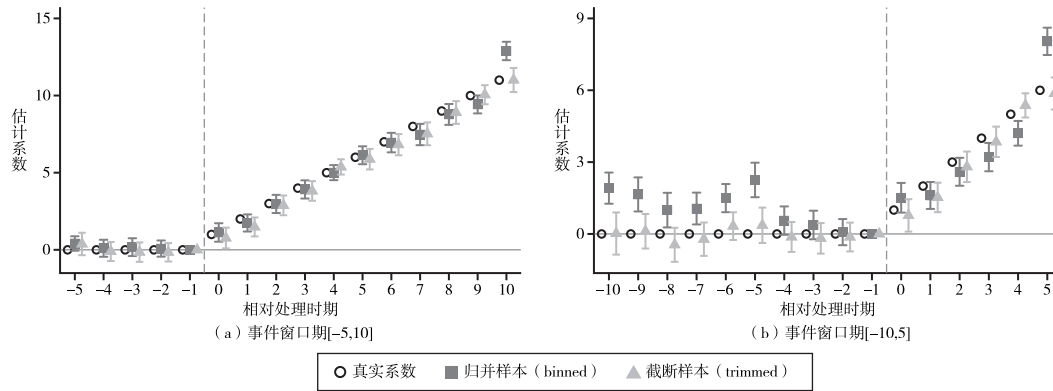


图6 归并和截断数据对事件研究法估计结果的影响

实践中需要对上述优缺点进行充分的考虑和权衡。

(四)正确控制组群异质性时间趋势

事件研究法的核心识别假设是平行趋势假设。然而,由于涉及处理发生后不可观测的潜在结果,平行趋势假设本身是不可能被直接检验所证实或证伪的。因此,实践中研究者们一方面会检验事前平行趋势是否成立从而间接地为平行趋势假设提供支持,另一方面也考虑尽可能放松平行趋势假设,即允许处理组和控制组存在某种确定形式的趋势差异。最常见的是假设组间存在异质性的线性时间趋势,如图7所示,3个处理组分别存在斜率为1、0.5、0.2的线性时间趋势,控制组不存在时间趋势。该数据生成过程显然不满足平行趋势假设,直接使用事件研究法无法得到平均处理效应的一致估计结果。针对这种情况,可以在回归方程的基础上进一步控制组群特定时间趋势:

$$y_{it} = \alpha_i + \lambda_t + \gamma_g f(t) + \sum_{l=-K}^{-2} \mu_l D_{it}^l + \sum_{l=0}^L \beta_l D_{it}^l + \epsilon_{it} \quad (15)$$

式(15)中的 $f(t)$ 代表时间趋势项的函数形式,例如 $f(t) = t$ 代表线性时间趋势, $f(t) = t + t^2$ 代表二次时间趋势; γ_g 代表时间趋势项的参数, g 意味着不同组群的时间趋势项存在差异。以图7(a)所示的数据为线性时间趋势 $[f(t) = t]$,控制组的时间趋势斜率 $\gamma_0 = 0$,处理组中组群1的时间趋势斜率 $\gamma_1 = 1$,组群2的时间趋势斜率 $\gamma_2 = 0.5$,组群3的时间趋势斜率 $\gamma_3 = 0.2$ 。实践中研究者们通常直接回归式(15),认为 $\gamma_g f(t)$ 可以良好地控制潜在的组间时间趋势差异,从而确保估计系数 μ_l 和 β_l 能够一致估计各期平均处理效应。然而事实并非如此。图8(a)描述了基本的事件研究法和

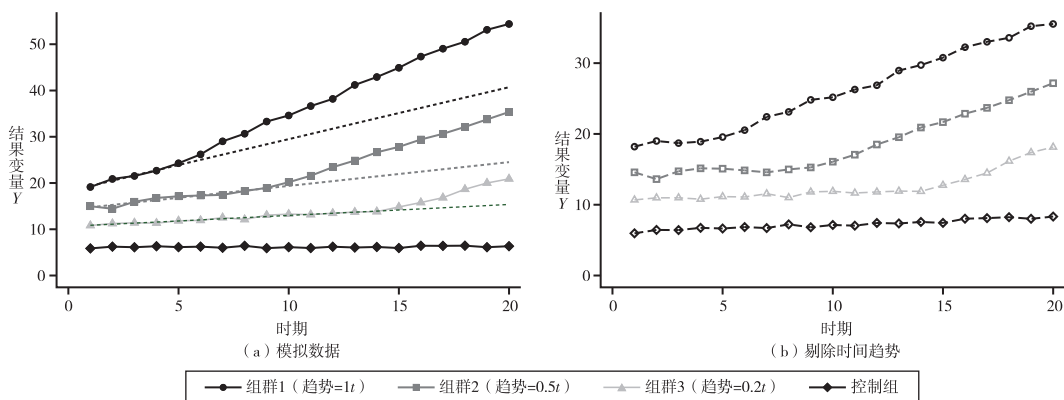


图7 存在组间时间趋势差异的原始数据与去趋势处理结果

控制组群线性时间趋势项 $\gamma_g \times t$ 的事件研究法估计结果。首先,可以发现由于平行趋势假设不满足,事件研究法无法正确识别处理前和处理后的真实平均处理效应。其次,令人惊讶的是,当控制住组群时间趋势项 $\gamma_g \times t$ 后,事件研究法不仅无法一致估计各期的平均因果效应,甚至会导致更大的估计偏误!并且,如果没有对数据做归并或截断处理,在包含全部相对时间变量的“完全动态”回归模型中需要额外去除一个相对时期变量以避免共线性问题,此时事件研究法的估计结果对基期选择非常敏感(Borusyak等,2023)。图8(b)描述了控制组群时间趋势时更换基期对估计结果的影响,可以清晰地看到使用-10期和-5期作为基期与使用-1期和5期作为基期得到的估计结果大相径庭。这些证据都表明在事件研究法中直接控制组间趋势可能存在导致严重估计偏误的潜在风险。

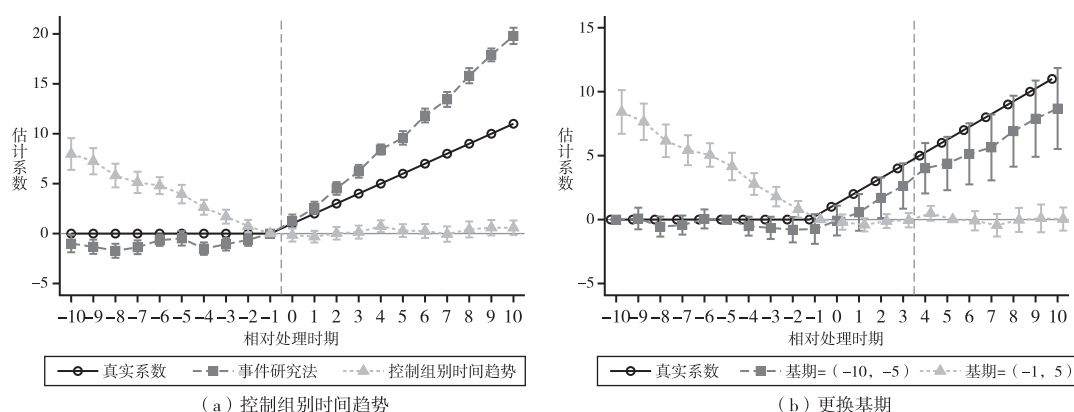


图8 控制组群时间趋势的事件研究法估计结果

为何在回归中直接控制组群时间趋势无法确保事件研究法的一致性呢?本文提供一种直观理解:根据Frisch-Waugh-Lovell定理,估计式(15)中的系数 μ_i 和 β_i 等价于先“剔除”双向固定效应和组群时间趋势对结果变量 y_{it} 和相对时期变量 D_{it}^* 的影响后再进行回归(Borusyak等,2023;Liu等,2022)。然而,如果使用全样本进行回归,组群时间趋势项会错误地“吸收”一部分处理后的动态平均处理效应作为组间异质性时间趋势,此时事件研究法不能再一致地估计各个相对时期的平均处理效应。表1的估计结果说明了这一问题。第(1)列为真实组群时间趋势系数。第(2)列使用未接受处理的子样本估计组群时间趋势,可以看到估计系数与第(1)列中的真实系数非常接近,表明估计结果具有一致性。但是,第(3)列使用全样本估计组群时间趋势,得到的估计系数与第(1)列中的真实系数具有显著差异,表明使用全样本并不能一致估计组群时间趋势。

如何才能正确控制可能存在的组群间异质性时间趋势呢?目前文献中出现的主要有两种方法。第一种方法是手动剔除事前趋势的两阶段估计法。第一阶段使用未接受处理的子样本(控制组和尚未接受处理的处理组)估计组群异质性时间趋势并外推至处理后时期,从结果变量中剔除时间趋势(Detrend)。第二阶段使用剔除事前趋势后的结果变量进行事件研究法估计(Goodman-Bacon,2021;2021;Dustmann等,2022)。这种方法背后的逻辑很直观:如果组群时间趋势在处理前和处理后时期保持不变,使用未接受处理的子样本可以正确估计组群时间趋势,经过去趋势处理的结果变量重新满足平行趋势假设[见图7(b)],事件研究法能够正确地估计出各期平均处理效应。不过,在使用这种方法时需要注意以下几点:第一,该方法成立的前提条件是组群时间趋势在处理前后不发生系统

性变化,但该假设和平行趋势假设一样无法进行验证;第二,需要知道组群时间趋势的形式(线性或更高次项),否则会由于模型错误设定(Misspecification)无法一致估计时间趋势项,如果处理后时期较长的话可能导致较为严重的外推偏误(Exploitation Bias);第三,处理前期数要尽可能地长一些以提高事前趋势的估计准确度。两阶段估计法的一个潜在问题是第二阶段估计得到的标准误并未考虑到第一阶段拟合组间时间趋势时可能存在的不确定性,可能会导致低估标准误。^①

表 1 使用全样本和未处理样本估计组群时间趋势

解释变量	被解释变量:结果变量 y		
	(1) 真实系数	(2) 未处理样本	(3) 全样本
组群 1×时间趋势	1	1.042*** (0.156)	1.933*** (0.017)
组群 2×时间趋势	0.5	0.509*** (0.047)	1.102*** (0.017)
组群 3×时间趋势	0.2	0.238*** (0.028)	0.447*** (0.017)
个体固定效应		是	是
时间固定效应		是	是
观测值		4700	8000
R ² 值		0.785	0.953

注:(1)由于与时间固定效应存在共线性问题,需要去掉 1 组时间趋势作为基准组,本文选取不存在时间趋势的控制组作为基准组;(2)*、**、***分别表示在 10%、5%、1% 的水平上显著,括号内为聚类在个体层面的稳健标准误。

第二种方法是 Borusyak 等(2023)、Liu 等(2022)和 Gardner(2022)的“插补法”思路^②,其基本思想与第一种做法类似:既然问题出在使用全样本无法一致估计组群时间趋势项,那么一个自然的解决思路是使用未接受处理的子样本来正确地识别出组群时间趋势项 $\gamma_g f(t)$ 以及双向固定效应 α_i 和 λ_t ,将其从结果变量 y_{it} 中剔除后再进行回归估计。图 9 使用 Liu 等(2022)提出的广义合成控制法以及 Gardner(2022)提出的两阶段双重差分法(Two-Stage DID)重新估计了式(15),可以看到该方法确实能够正确地控制组群时间趋势,从而一致地估计各个相对时期的平均处理效应。第二种方法的优势是可以同时考虑一阶段拟合和二阶段估计的不确定性,获得正确的标准误。从图 9 可以看到,去除事前趋势的两阶段估计法得到的标准误要显著低于插补法估计量的标准误,可能导致“过度拒绝”(Over-rejection)问题。与两阶段估计法相比,Borusyak 等(2023)和 Liu 等(2022)采用自助法(Bootstrap)估计标准误、Gardner(2022)使用联合广义矩估计方法都可以获得正确的渐进标准误估计从而进行稳健的统计推断。

五、异质性处理效应条件下的事件研究法

(一)异质性处理效应情形下事件研究法的偏误来源

在许多研究情景中,同样的政策可能对不同个体产生显著的异质性影响,例如工业机器人和人工智能的兴起对低技能工人产生明显的替代效应,但对高技能工人的替代效应可能更小甚至表现

① 这一问题与手动进行工具变量两阶段回归时面临的标准误低估问题类似。

② 第五部分会详细介绍“插补法”等异质性处理效应稳健估计量,感兴趣的读者也可进一步阅读刘冲等(2022)。

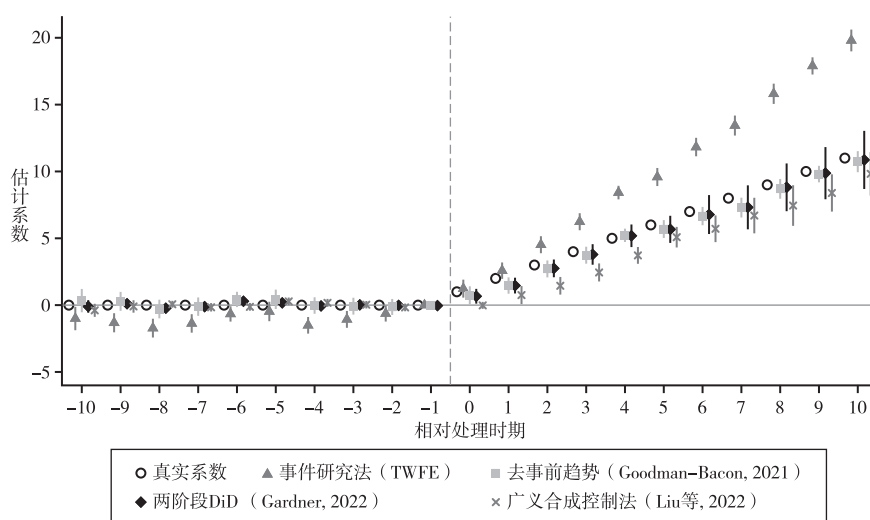


图9 使用去事前趋势和插补法正确控制组群时间趋势

注:为便于更清晰地展示估计结果,上图中只展示了-10期到10期的估计系数与置信区间。

为互补效应(Acemoglu 和 Restrepo, 2018)。因此,基于双向固定效应模型的事件研究法所需要的三个识别假设中的同质性处理效应路径假设在实践中很可能不满足。近年来涌现出的一批理论计量文献深入讨论了异质性处理效应条件下双重差分法和事件研究法可能存在的估计偏误问题(De Chaisemartin 和 D'Haultfoeuille, 2020; Goodman-Bacon, 2021; Sun 和 Abraham, 2021; Callaway 和 Sant'Anna, 2021; Gardner, 2022; Borusyak 等, 2023; Dube 等, 2023)。上述研究发现,如果组群具有异质性处理效应路径,即使满足平行趋势假设和无预期效应假设,基于双向固定效应模型的事件研究法仍然存在估计偏误,无法一致地估计各个相对时期的平均处理效应。

借助 Sun 和 Abraham(2021)的理论分析可以更清晰地理解为何异质性处理效应会导致基于双向固定效应模型的事件研究法产生估计偏误。根据式(6)可知,事件研究法的第 l 期估计系数可以分解为三部分:全部组群 g 在第 l 期、其他相对时期和基期的平均处理效应 $ATT_{g,l}$ 的某种加权平均,权重可以根据以下辅助回归(Auxiliary Regression)得到:

$$D_{it}^l \cdot \mathbf{1}\{G_i = g\} = \alpha_i + \lambda_t + \sum_{j=-K}^L \omega_{g,j}^l D_{it}^j + \epsilon_{it} \quad (16)$$

借助式(16)可以求得任一估计系数所包含的组群一时间平均处理效应 $ATT_{g,l}$ 所对应的权重。本文使用模拟数据,估计了式(2)中的估计系数 β_3 所包含的各组群一时间平均处理效应 $ATT_{g,l}$ 所占权重,结果展示在图 10。可以发现, β_3 中不仅包含了第 3 期各组群的平均处理效应(权重和为 1),还包含了基期(权重和为 -1)与其他相对处理时期(每一期的权重和为 0)的平均处理效应。若平均处理效应在组群一时间维度存在异质性,任意其他相对处理时期的加权平均处理效应均不等于 0,此时 β_3 受到其他相对处理时期加权平均处理效应的“污染”,导致估计结果偏离真实值。

(二)异质性处理效应情形下的稳健估计量

为了克服在异质性处理效应情形下的估计偏误,研究者们提出了一系列新的稳健估计方法,根据思路不同主要可分为组群一时间平均处理效应估计法、插补法和堆叠法三类。Baker 等(2022)和刘冲等(2022)对上述方法进行了精彩详细的综述,本文主要对三类估计量做一些简要说明。

如前文所述,事件研究法估计本质上是一系列简单的 2×2 双重差分法估计量的加权平均。双

向固定效应估计量的问题在于其中一部分 2×2 双重差分法在估计较晚接受处理组群的平均处理效应时使用了较早接受处理的组群充当控制组,这使得控制组的前后变化既包含了潜在时间趋势,又包含了动态处理效应。因此,处理组前后变化减去控制组前后变化时会错误地减去一部分动态处理效应,产生如图 10 所示的负权重问题。Borusyak 等(2023)将这类使用较早接受处理的组群充当控制组的双重差分比较称为“禁止比较”(Forbidden Comparisons)。

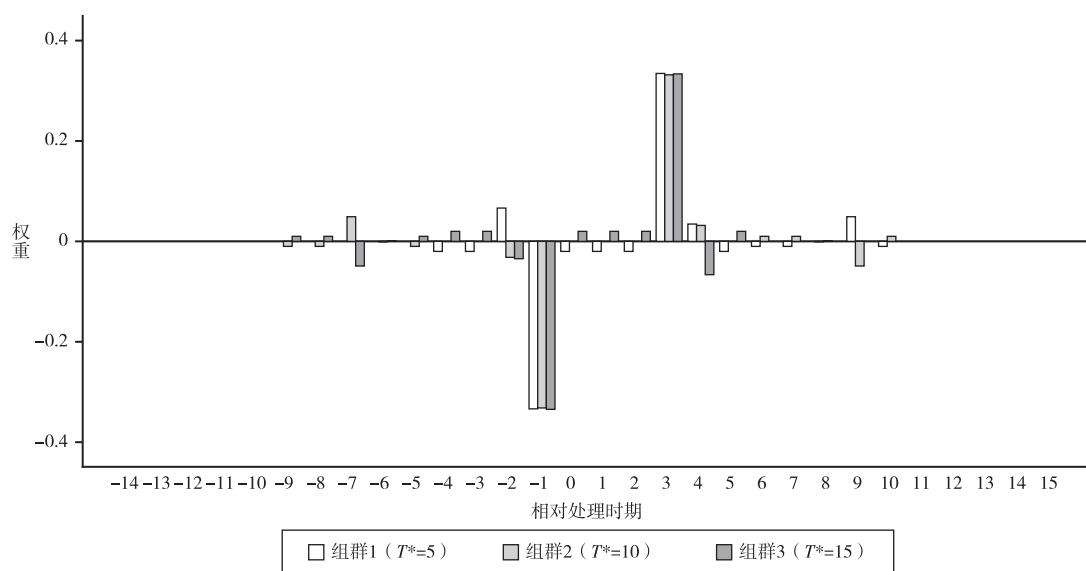


图 10 事件研究法估计系数的权重分解

三类稳健估计量采取不同的思路来处理“禁止比较”带来的负权重问题。第一类是组群一时期平均处理效应估计法,包括 De Chaisemartin 和 D’Haultfoeuille (2020)、Sun 和 Abraham (2021)、Callaway 和 Sant’Anna (2021) 提出的异质性处理效应稳健估计量以及 Dube 等(2023)提出的局部投影双重差分法(Local Projection DiD)。这类方法的基本思路是首先估计出组群一时期平均处理效应 $ATT_{g,l}$, 然后根据某种权重(等权重或组群规模权重)手动加权求出各相对时期的加权平均处理效应 ATT_l 。由于在手动加重的过程中人为避免了负权重问题,这类估计量可以修正双向固定效应模型的偏误,正确地估计各时期的平均处理效应。第二类是插补法,包括 Liu 等(2022)、Borusyak 等(2023)和 Gardner (2022) 提出的估计量,基本思路是使用未受处理的子样本(包括控制组和处理前时期的处理组)估计拟合出处理组个体的反事实结果,而后计算出个体处理效应,最后加总得到各个相对时期的平均处理效应。由于在估计拟合反事实结果时仅使用了未接受处理的子样本,因此不存在“禁止比较”的情形,不会导致负权重问题和估计偏误。最后一类是堆叠法(Cengiz 等, 2019),其基本思想是通过为每个处理组组群匹配一个由从未接受处理的控制组或尚未接受处理的处理组样本构成的特定控制组,再将匹配好的样本堆叠起来进行回归。堆叠法与插补法类似,都是通过避免使用已经接受处理的样本作为控制组来规避负权重问题。三类稳健估计量在识别假设、组群一时间平均处理效应的估计方法、加权方式、标准误估计等方面存在一些差异,但都能够在异质性处理效应情形下正确估计各时期的平均处理效应(见图 11)。

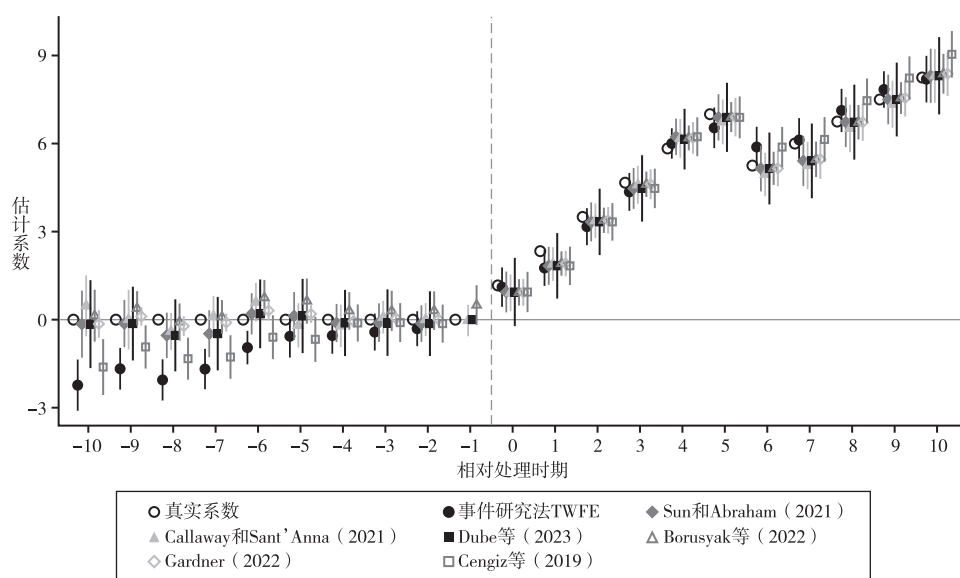


图 11 异质性处理效应稳健估计量

(三) 异质性处理效应情形下的事前平行趋势检验

在当前的研究实践中,事件研究法最重要的作用之一是检验事前平行趋势是否成立,在异质性处理效应情形下如何正确评估和检验事前平行趋势是一个重要的实践问题。由于基于双向固定效应模型的传统事件研究法估计系数存在偏误,即使事前各期系数联合检验不显著也无法作为事前平行趋势成立的支撑证据。研究者们可以使用异质性处理效应稳健估计量一致地估计各个相对时期的平均处理效应,根据事前各期系数的联合显著性来检验事前平行趋势是否成立。

事实上,在异质性处理效应情形下检验事前平行趋势同样可以使用双向固定效应模型。研究者只需使用未接受处理的子样本(包括全部控制组样本和处理前时期的处理组样本)进行事件研究法估计就能够一致估计处理前各时期的平均处理效应(Borusyak等,2023)。该方法的原理和插值法类似,都是通过选择未受处理的子样本进行回归来避免出现“禁止比较”,因此不会存在估计偏误。如图12所示,在异质性处理效应情形下,使用全样本时事件研究法的事前各期估计系数都是有偏的,联合检验的结果错误地拒绝了事前各期平行的原假设,无法正确地检验事前平行趋势。但是,使用未受处理的子样本时事件研究法的事前各期估计系数是一致的,联合检验结果正确地表明事前平行趋势成立。与使用异质性处理效应稳健估计量进行事前平行检验相比,这种基于线性回归模型的事前平行趋势检验方法在使用简便性和运算速度上都具有明显优势。

(四) 异质性处理效应稳健估计量和传统事件研究法的互补性

既然新的异质性处理效应稳健估计量能够克服异质性处理效应导致的估计偏误,其稳健性要优于传统的事件研究法,那么在实践中是否应该使用这类新方法来完全替代传统的事件研究法呢?本文认为新方法和传统的事件研究法两者之间并不是替代关系而是互补关系,二者搭配使用可以更好地服务于实证研究需求。新方法的重点在于给定数据满足平行趋势等识别假设时,在异质性处理效应情形下如何准确估计平均处理效应。然而,如果组群间存在异质性处理效应,根据一般逻辑推理,最大的可能性是某些协变量(不妨记为 X_g)在组群间存在显著差异(Sun和Abraham,2021)。由于组群处理时点不同,组间不平衡的协变量 X_g 很可能与处理时点存在某种相关性。那么此时必须要慎重对待的一个问题是,协变量 X_g 是否可能是一个决定组群处理时点的选择性变量?这种选

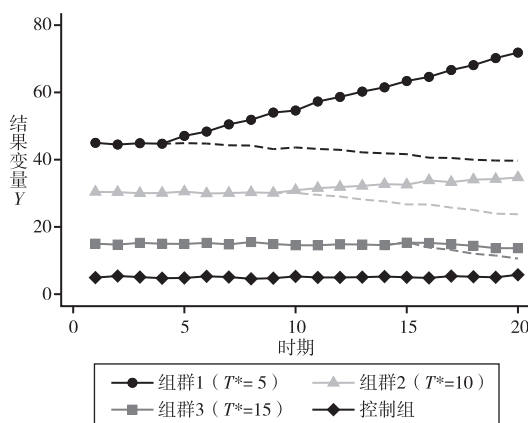
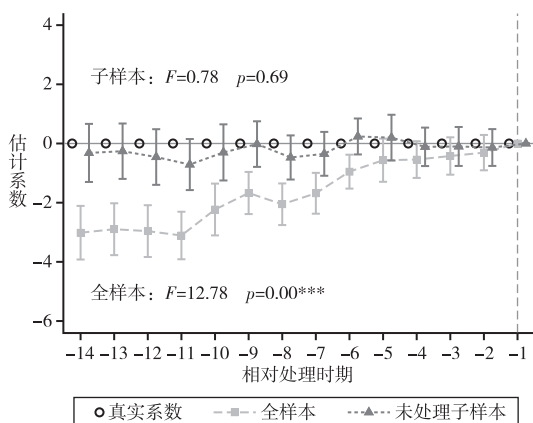


图 12 存在异质性处理效应的事前平行趋势检验 图 13 处理时点内生的数据生成过程

择性是否与潜在结果的时间趋势有关?若 X_g 确实是一个选择性变量且会影响反事实结果的时间趋势,这就意味着处理时点内生性可能导致平行趋势假设不再成立。

处理时点内生选择的一个典型例子是阿森费尔特沉降(Ashenfelter's Dip)。Ashenfelter(1978)在评估培训项目对工资的影响时发现职工平均收入在参加培训前有一个明显降低的趋势,后续研究发现这一现象的原因在于失业的工人为了更好地寻找到新工作而更有动机参加职业培训。由于参加职业培训的工人(处理组)与同时期未参加职业培训的工人(控制组)不具有良好的可比性,平行趋势假设很可能并不成立,导致事件研究法无法正确识别因果效应(Heckman和Smith,1999)。另一个常见的例子是离职对员工收入的影响。员工离职有两种可能情形,一种是主动辞职去新的企业工作以获得更高的收入,另一种是被企业强制解雇导致被动失业。显然,不同情形的离职对员工收入的影响存在巨大差异:主动辞职者通常可以获得更高的收入,而被迫失业者收入会大幅度降低。更为严重的问题在于,主动辞职者和被迫失业者(处理组)与持续工作者(控制组)的反事实收入很可能有不同的时间趋势,此时平行趋势假设不再成立。

本文使用模拟数据进一步说明该问题。数据生成过程如图13所示,结果变量是个人收入,事件为更换工作,虚线为未更换工作的个人收入反事实潜在结果。从图13中可以看到,当个人预期到企业未来经营困难、收入可能会下降时会考虑更换工作;能力越强的员工(事前收入越高)越容易找到新职位,因此会更早更换工作(处理时点靠前),并且新职位的收入也会越高(处理效应更强);反之,能力较低的员工更难找到新职位,所以更换工作较晚(处理时点靠后)并且新职位的收入更低(处理效应更弱)。由于处理组未受处理的反事实潜在结果与控制组不平行,不满足平行趋势假设。在这种情形下,无论是基于双向固定效应模型的传统事件研究法还是异质性处理效应稳健估计量都存在估计偏误,无法正确估计各期平均处理效应(见图14)。^①

综上所述,本文认为虽然传统事件研究法可能在异质性处理效应假设下存在估计偏误,但这并不意味着其没有使用价值。研究者可以通过对比传统事件研究法和异质性处理效应稳健估计方法的结果来获得更多关于组间可比性的信息。如果传统事件研究法和稳健估计量的结果不存在明显差异,表明异质性处理效应可能并不严重,此时可以继续使用传统事件研究法作为基准识别策略。如果二者结果存在很大差异则表明存在明显的异质性处理效应,此时研究者至少应该注意两个问

^① 图14的估计结果带来的另一个启示是,即使事前趋势平行也并不意味着事后平行趋势假设一定成立。因此,不能过度依赖事前趋势是否平行来验证平行趋势假设。

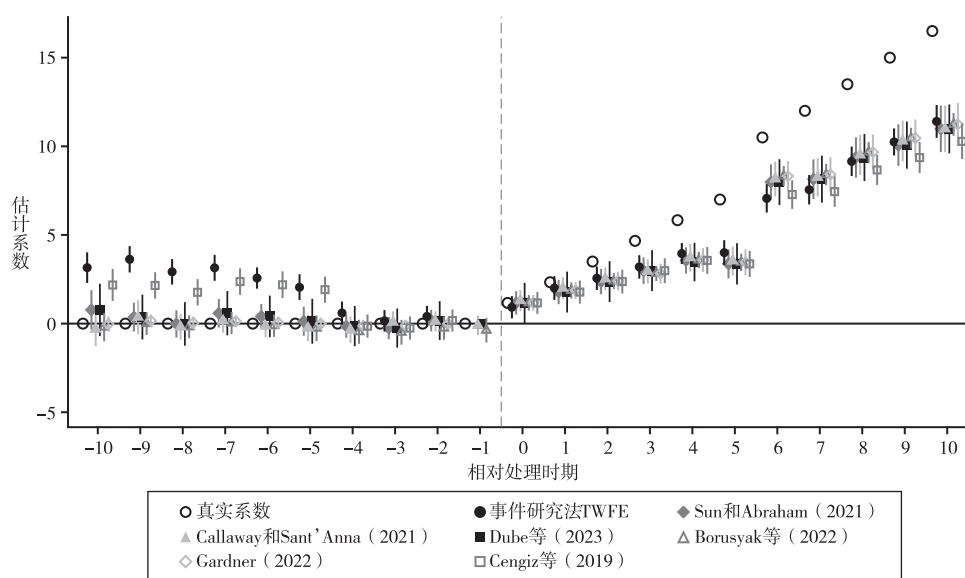


图 14 不满足平行趋势假设时各类事件研究法估计结果

题:首先,最重要的问题是寻找异质性处理效应的来源,特别是寻找可能存在的组间不平衡特征变量并判断该变量是否可能导致潜在的内生性问题,而后进一步对是否可能存在内生处理时点选择、是否可能威胁到平行趋势假设等一系列问题做出充分严谨的分析和处理。^①其次,若明确存在异质性处理效应,应以新发展的稳健估计量作为基准识别策略,以避免可能存在的估计偏误,获得更为可信的因果效应估计结果。

六、总结性评论

本文系统归纳总结了事件研究法的基本思想、实现方式和实践中的注意事项。近年来,大量的国内经济学和管理学领域的实证研究使用双重差分法来进行因果推断和政策评估。最新的前沿文献表明,事件研究法是传统的“两组一期”双重差分法的动态形式拓展,能够比传统双重差分法提供更多关于政策动态效应的信息。在政策交错发生和存在异质性处理效应的情形下,事件研究法能够修正时间维度上的处理效应异质性导致的估计偏误,相较于传统双重差分法的估计结果更稳健。在处理效应在“组群一时间”双重维度上存在异质性的情形下,虽然事件研究法本身存在估计偏误,但其估计结果仍然有助于判定异质性处理效应的存在性和来源。目前国内经济学和管理学等社会科学领域虽然已经逐渐开始运用事件研究法,但在许多文献中仍然只是将其作为双重差分法的辅助性工具,可能在无意中忽略了事件研究法所传达出来的丰富信息。本文试图详细地介绍事件研究法的用途和实现方法,针对研究者们应用过程中的一些常见问题进行分析讨论并提出相应的解决方案,期望能为中国社会科学领域的应用研究者们更好地使用事件研究法这一现代因

^① 从近两年经济权威期刊(如“五大刊”)发表的论文来看,绝大部分论文仍然采用传统的事件研究法作为主要识别策略,辅以异质性处理效应稳健估计量作为稳健性检验,二者一般不存在明显差异。Chiu等(2023)复现了38篇发表在政治学领域顶级期刊的经验研究论文,发现使用异质性稳健估计量后绝大多数论文结论没有改变。对于这一现象,笔者的一个猜想是,如果两种方法的估计结果存在显著差异,可能暗示存在文中所述的选择性问题。这类问题会直接威胁到实证结论的可信性,但常常又难以较为理想地解决,导致论文无法通过同行评议。因此,最终绝大部分发表论文会呈现出两种估计方法结论基本一致的现象。

果推断和政策评估工具提供一些有意义的参考。

事件研究法作为一种很早就诞生的实证研究方法,在其发展过程中不断吸收、借鉴新的理论计量研究成果,直到今天仍然保持着旺盛的生命力。在异质性处理效应、连续处理变量、包含时变协变量等情形下事件研究法的识别假设和估计方法仍是目前理论计量领域的最前沿问题,相关文献不断涌现。本文的分析和建议是基于笔者对当前的前沿理论计量和实证研究文献的理解和自身研究经验归纳总结而来,作为一家之言必定存在谬误和不足。期待本文的“抛砖引玉”能引发更广泛的讨论,以促成经济学、管理学、政治学、社会学等中国社会科学各领域的共同进步。

参考文献

- [1]黄炜,任昶宇,周羿.退休制度、劳动供给与收入消费动态[J].经济研究,2023,(1):141~157.
- [2]黄炜,张子尧,刘安然.从双重差分法到事件研究法[J].产业经济评论,2022,(2):17~36.
- [3]刘冲,沙学康,张妍.交错双重差分:处理效应异质性与估计方法选择[J].数量经济技术经济研究,2022,39(9):177~204.
- [4]Acemoglu D., Restrepo P., 2018, *The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment* [J], *American Economic Review*, 108 (6), 1488~1542.
- [5]Angrist J. D., Pischke J. S., 2009, *Mostly Harmless Econometrics: An Empiricist's Companion* [M], Princeton: Princeton University Press.
- [6]Angrist J. D., Pischke J. S., 2010, *The Credibility Revolution in Empirical Economics, How Better Research Design Is Taking the Con out of Econometrics* [J], *Journal of Economic Perspectives*, 24 (2), 3~30.
- [7]Armitage S., 1995, *Event Study Methods and Evidence on Their Performance* [J], *Journal of Economic Surveys*, 9 (1), 25~52.
- [8]Ashenfelter O., 1978, *Estimating the Effect of Training Programs on Earnings* [J], *Review of Economics and Statistics*, 60 (1), 47~57.
- [9]Baker A.C., Larcker D.F., Wang C.C.Y., 2022, *How Much Should We Trust Staggered Difference-in-Differences Estimates?* [J], *Journal of Financial Economics*, 144 (2), 370~395.
- [10]Ball R., Brown P., 1968, *An Empirical Evaluation of Accounting Income Numbers* [J], *Journal of Accounting Research*, 6 (2), 159~178.
- [11]Borusyak K., Jaravel X., Spiess J., 2023, *Revisiting Event Study Designs, Robust and Efficient Estimation* [J/OL], DOI: 10.48550/arXiv.2108.12419.
- [12]Callaway B., Sant'Anna P. H. C., 2021, *Difference-in-Differences with Multiple Time Periods* [J], *Journal of Econometrics*, 225 (2), 200~230.
- [13]Cengiz D., Dube A., Lindner A., Zipperer B., 2019, *The Effect of Minimum Wages on Low-Wage Jobs* [J], *The Quarterly Journal of Economics*, 134(3), 1405~1454.
- [14]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* [J/OL], DOI:10.2139/ssrn.4490035.
- [15]Currie J., Kleven H., Zwieters E., 2020, *Technology and Big Data Are Changing Economics, Mining Text to Track Methods* [J], *AEA Papers and Proceedings*, 110, 42~48.
- [16]De Chaisemartin C., D'Haultfoeuille X., 2020, *Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects* [J], *American Economic Review*, 110 (9), 2964~2996.
- [17]Dobkin C., Finkelstein A., Kluender R., Notowidigdo M J., 2018, *The Economic Consequences of Hospital Admissions* [J], *American Economic Review*, 108 (2), 308~352.

- [18]Dolley J. C., 1933, *Characteristics and Procedure of Common Stock Split-ups* [J], Harvard Business Review, 11 (3), 316~326.
- [19]Dube A., Girardi D., Jordà Ò., Taylor A.M., 2023., *A Local Projections Approach to Difference-in-Differences Event Studies* [R], NBER Working Paper, No.31184.
- [20]Dustmann C., Lindner A., Schönberg U., Umkehehe M., Vom Berge P., 2022, *Reallocation Effects of the Minimum Wage* [J], Quarterly Journal of Economics, 137 (1), 267~328.
- [21]Fama E. F., Fisher L., Jensen M. C., Roll R., 1969, *The Adjustment of Stock Prices to New Information* [J], International Economic Review, 10 (1), 1~21.
- [22]Gardner J., 2022, *Two-stage Differences-in-Differences* [J/OL], DOI: 10.48550/arXiv.2207.05943.
- [23]Goldsmith-Pinkham P., Hull P., Kolesár M., 2022, *Contamination Bias in Linear Regressions* [R], NBER Working Paper, No.30108.
- [24]Goodman-Bacon A., 2021, *Difference-in-Differences with Variation in Treatment Timing* [J], Journal of Econometrics, 225 (2), 254~277.
- [25]Guriev S., Melnikov N., Zhuravskaya E., 2021, *3G Internet and Confidence in Government* [J], Quarterly Journal of Economics, 136 (4), 2533~2613.
- [26]Heckman J. J., Smith J. A., 1999, *The Pre-Program Earnings Dip and The Determinants of Participation in A Social Programme. Implications for Simple Programme Evaluation Strategies* [J], Economic Journal, 109 (457), 313~348.
- [27]Imbens G. W., Lemieux T., 2008, *Regression Discontinuity Designs, A Guide to Practice* [J], Journal of Econometrics, 142 (2), 615~635.
- [28]Jacobson L. S., LaLonde R. J., Sullivan D. G., 1993, *Earnings Losses of Displaced Workers* [J], American Economic Review, 83(4), 685~709.
- [29]LaLonde R. J., 1986, *Evaluating the Econometric Evaluations of Training Programs with Experimental Data* [J], American Economic Review, 76 (4), 604~620.
- [30]Lee D. S., Lemieux T., 2010, *Regression Discontinuity Designs in Economics* [J], Journal of Economic Literature, 48, 281~355.
- [31]Liu L., Wang Y., Xu Y., 2022, *A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data* [J], American Journal of Political Science, Forthcoming.
- [32]MacKinlay A.C., 1997, *Event Studies in Economics and Finance* [J], Journal of Economic Literature, 35 (1), 13~39.
- [33]Miller D.L., 2023, *An Introductory Guide to Event Study Models* [J], Journal of Economic Perspectives, 37 (2), 203~230.
- [34]Peterson P. P., 1989, *Event Studies: A Review of Issues and Methodology* [J], Quarterly Journal of Business and Economics, 28 (3), 36~66.
- [35]Rambachan A., Roth J., 2023, *A More Credible Approach to Parallel Trends* [J], Review of Economic Studies, Forthcoming.
- [36]Roth J., 2022, *Pretest with Caution, Event-Study Estimates after Testing for Parallel Trends* [J], American Economic Review: Insights, 4 (3), 305~322.
- [37]Roth J., Sant' Anna P.H.C., Bilinski A., Poe J., 2023, *What's Trending in Difference-in-Differences? A Synthesis of The Recent Econometrics Literature* [J], Journal of Econometrics, 235 (2), 2218~2244.
- [38]Schmidheiny K., Siegloch S., 2022, *On Event Study Designs and Distributed-Lag Models, Equivalence: Generalization and Practical Implications* [J], Quarterly Journal of Business and Economics.
- [39]Sun L., Abraham S., 2021, *Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects* [J], Journal of Econometrics, 225 (2), 175~199.

Questions, Applications and Extensions in Event Study Approach

ZHANG Ziyao^{1,2} HUANG Wei³

(1. School of Public Finance and Taxation, Zhongnan University of Economics and Law;
2. Innovation and Talent Base for Income Distribution and Public Finance, Zhongnan University
of Economics and Law; 3. National School of Development, Peking University)

Summary: Event study is an empirical research method used to study the impact of specific events or policy shocks on individual behavior. In recent years, event study has gained increasing importance in empirical research and has been widely applied by researchers in various social science fields. The popularity of event study in recent years is largely due to its close connection with the canonical difference-in-differences (DID) method. Using event study to test pre-parallel trends has become one of the standard steps in the current DID approach. The role of event study goes far beyond testing pre-parallel trends. Compared with the DID method, event study has many attractive features. First, it can visually demonstrate the dynamic changes in policy effects, which is the most appealing characteristic of the method. Second, recent advances in theoretical econometrics have demonstrated that the estimated results of the DID method may be biased if there are heterogeneous treatment effects at the group or time level. An event study is more robust than the DID method in this scenario. This study attempts to provide a comprehensive analysis and summary of the implementation, identification assumptions, and practice suggestions of the event study method to guide researchers in various social science fields.

Based on the cutting-edge econometric literature, we found that 1) the estimated coefficient of the TWFE model in the event study approach is essentially a weighted average of a bunch of canonical 2×2 DID estimates. 2) Heterogeneity in treatment effects at the group and time (or both) levels can lead to estimation bias in the DID approach based on the two-way fixed effects model. If the dynamic paths of treatment effects remain consistent across groups (homogeneous treatment effect paths assumption), the event study approach can consistently estimate the average treatment effects for each relative period of the treatment. 3) Researchers can assess the existence of heterogeneous treatment effects by comparing the differences between the estimation results of the TWFE event study and new HTE-robust estimators and then identify potential factors that may cause heterogeneous treatment effects. This evaluation can help assess whether endogeneity issues may arise in the timing of treatment.

This study provides several recommendations for using the event study method. 1) We suggest that a pre-event period should be used as the base period, e.g., -1 period. 2) When testing pre-parallel trends, we suggest that a joint test of the significance of pre-event coefficients should be conducted. 3) Binning data after the treatment period may lead to estimation bias, so it is essential to avoid binning post-treatment data in practice. If pre-parallel trends are met, binning can be performed for the pre-treatment period. Trimming data requires consideration of potential sample selection issues and loss of estimation efficiency. 4) Controlling group heterogeneity time trends directly in the TWFE model may cause estimation bias. Using the imputation method can effectively control potential differences in pre-trends between groups, thereby avoiding estimation bias.

Keywords: Event Study Approach; Dynamic Difference-in-Differences; Two-way Fixed Effect; Heterogeneous Treatment Effects

JEL Classification: C13; C21; C22

(责任编辑:唐跃桓)