2023 年第 2 期（总第 146 期）
3 月 15 日出版

东北财经大学 学报

〔DOI〕10. 19653/j. cnki. dbcjdxxb. 2023. 02. 003
〔引用格式〕王鹏超，韩立彬. 多时点双重差分法的潜在问题与解决措施[J]. 东北财经大学学报，2023(2)：27-39.

多时点双重差分法的潜在问题与解决措施

王鹏超，韩立彬

（东北财经大学 经济与社会发展研究院，辽宁 大连 116025）

〔摘 要〕多时点双重差分法具有准自然试验特征，可以相对干净地识别因果效应，广泛应用于与政策评估相关的研究中，但必须重视其可能存在的估计偏差问题。本文总结了多时点双重差分法存在的问题和相应的解决措施。通过梳理最新文献发现，多时点双重差分法回归系数识别的是组别—时间处理效应的加权平均，而非受处理个体的平均处理效应。在异质性处理效应下，多时点双重差分法估计系数有偏，严重时估计系数符号会与真实系数符号相反。目前文献上提出的解决措施可以归结为一个诊断方法和三类解决方法。其中，诊断方法为Goodman-Bacon的系数分解定理，三类解决方法分别是加总方法、两步回归法和堆叠型双重差分法。

〔关键词〕双重差分法（DID）；多时点双重差分法；异质性处理效应；组别—时间处理效应

中图分类号：F064. 1　　**文献标识码**：A　　**文章编号**：1008-4096(2023)02-0027-13

一、问题的提出

计量经济学可信性革命推动了实证经济学进展，因果推断成为实证经济学研究的显学。2021年，诺贝尔经济学奖授予Card、Angrist和Imbens三位学者，表彰Card对劳动经济学领域的实证贡献，以及Angrist和Imbens对因果推断方法的贡献。这充分肯定了因果推断方法在经济学中的应用与发展。双重差分法（Difference-In-Difference，DID）作为应用最广的因果推断方法之一，可以相对干净地识别因果效应，在政策评估中受到国内外学者青睐。本文统计了2005—2020年使用DID方法的中文期刊文章数量。DID已成为国内实证研究者进行学术研究的重要工具。

根据政策实施时点的不同，DID一般可分为单时点DID（Staggered DID）和多时点DID（Multiple DID）。然而学界对多时点DID识别的系数含义与正确性却较少讨论。单时点DID识别的

① 多时点DID有时又被称为渐进DID或交错型DID。

收稿日期：2022-11-12
基金项目：国家自然科学基金青年项目“土地资源配置对人力资本空间分布的影响研究：理论、机制与对策”（72003020）
作者简介：王鹏超(1996—)，男，山西晋城人，博士研究生，主要从事区域和城市经济研究。E-mail：pengchaowang1996@163.com
韩立彬(1988—)，男，山东临沂人，副教授，博士，主要从事区域和城市经济研究。E-mail：hanlibin@dufe. edu. cn

是受处理个体的平均处理效应（Average Treatment Effect on the Treated，ATT），多时点DID识别的是否也是受处理个体的平均处理效应？多时点DID估计系数是否有偏？现有文献并未过多讨论。《American Economic Review》2020年第9期，Chaisemartin和D′Haultfoeuille[1]探讨了多时点DID存在的问题，《Journal of Econometrics》在2021年第2期发布了“处理效应”专题，其中3篇文章与多时点DID识别直接相关。表明学术界对这一方法存在问题的高度关注。

最新研究发现，多时点DID估计系数识别的并不是受处理个体的平均处理效应，而是组别—时间处理效应的加权平均。当存在异质性处理效应时，估计系数有偏[1-5]。Goodman-Bacon[2]认为，多时点DID估计系数可分解为多个单时点DID系数的加权平均，权重与每个单时点DID的样本份额和解释变量方差相关，且都为正值。然而，部分单时点DID把早接受处理组作为晚接受处理组的对照组，在异质性处理效应下，这部分系数可能为负，从而总体估计系数会存在较大偏差。Chaisemartin和D′Haultfoeuille[1]、Borusyak和Jaravel[3]以及Borusyak等[4]认为，组别—时间处理效应为正，但部分权重为负，导致最终估计结果有偏。虽然事件研究法可以将不同处理时点转化为处理时点一致的相对时点，但Sun和Abraham[6]证明，事件研究法设定中的每一相对时点系数不仅与该相对时点系数相关，还与回归方程中其他相对时点系数及被剔除在方程之外的相对时点系数相关。在异质性处理效应下，利用相对时点系数大小检验平行趋势假定是否满足也会存在问题。

针对多时点DID存在的问题，学者们提出了不同的解决方法，本文将其归结为一个诊断方法和三类解决方法。其中，诊断方法为Goodman-Bacon[2]的系数分解定理，该方法用于诊断估计系数的偏差程度。第一类解决方法为Sun和Abraham[6]、Callaway和Sant′Anna[7]提出的加总方法，即分别估计每个时期每个组别平均处理效应，再将其加总得到所有受处理个体的平均处理效应；第二类解决方法为Gardner[8]和Borusyak和Jaravel等[3]提出的两步回归法；第三类解决方法为堆叠型DID（stacked DID）[9]，将每一政策时点前后一段时期内的处理组和干净的对照组形成一个数据集，之后把所有的数据集堆叠并进行回归。

二、DID方法的基本原理

为阐明DID的基本原理，考虑包含2个组别和2个时期的2 × 2 DID情形。组别包括一个处理组（$treat_i = 1$）和一个对照组（$treat_i = 0$），时期包括政策前（$post_t = 0$）和政策后（$post_t = 1$）。政策效果通过式（1）进行估计：

$$Y_{it} = \beta_0 treat_i \times post_t + \beta_1 treat_i + \beta_2 post_t + \varepsilon_{it} \tag{1}$$

其中，β_0为政策评估所关注系数，识别的是受处理个体的平均处理效应。这一系数可以用式（2）和式（3）加以分解说明：

$$\hat{\beta}_0 = [\mathbb{E}(Y_{it}(1)|treat_i=1,\ post_t=1) - \mathbb{E}(Y_{it}(0)|treat_i=1,\ post_t=0)] - [\mathbb{E}(Y_{it}(0)|treat_i=0,\ post_t=1) - \mathbb{E}(Y_{it}(0)|treat_i=0,\ post_t=0)] \tag{2}$$

$$\begin{aligned} &= [\mathbb{E}(Y_{it}(1)|treat_i=1,\ post_t=1) - \mathbb{E}(Y_{it}(0)|treat_i=1,\ post_t=1)] + [\mathbb{E}(Y_{it}(0)|treat_i=1,\ post_t=1) \\ &\quad - \mathbb{E}(Y_{it}(0)|treat_i=1,\ post_t=0)] - [\mathbb{E}(Y_{it}(0)|treat_i=0,\ post_t=1) - \mathbb{E}(Y_{it}(0)|treat_i=0,\ post_t=0)] \end{aligned} \tag{3}$$

其中，$Y_{it}(1)$和$Y_{it}(0)$是潜在结果。式（2）是处理组事后与事前均值的差异减去对照组事后与事前均值的差异，式（3）在式（2）上各加减一项$\mathbb{E}(Y_{it}(0)|treat_i = 1,\ post_t = 1)$。式（3）中第一项表示事后所有受处理个体处理效应的均值，第二项是处理组事后假若未经处理的结果均值减去处理组事前的结果均值，第三项同理第二项。若满足平行趋势假定（第二项与第三项相减为0），得到$\hat{\beta}_0 = \{\mathbb{E}(Y_{it}(1) - Y_{it}(0)|treat_i = 1,\ post_t = 1)\}$，$\hat{\beta}_0$识别的是受处理个体的平均处理效应。

单个政策时点情形下，为灵活地估计回归系数，常用个体固定效应λ_i和时间固定效应η_t代替

式（1）中的$treat_i$和$post_t$。在多个政策时点情形下，若将D_{it}表示为处理状态，处理组在政策后受到影响为1，否则为0，由于$post_t$与个体i和时间t同时相关，多时点DID的回归方程如式（4）所示：

$$Y_{it} = \beta_0 D_{it} + \lambda_i + \eta_t + \varepsilon_{it} \tag{4}$$

在多时点DID实际应用中，由于存在多个处理组，无法直接通过对比处理组和对照组结果均值的时间变化，以此检验是否满足平行趋势假定，因而事件研究法通常被当作替代方法。这一方法不仅可以检验事前是否满足平行趋势假定，还可以观察事后政策效果的动态变化。事件研究法的模型设定为式（5）：

$$Y_{i,\ t} = \sum_{l=-k}^{-2} \beta_l^0 D_{i,\ t}^l + \sum_{l=0}^{L} \beta_l^1 D_{i,\ t}^l + \lambda_i + \eta_t + \varepsilon_{i,\ t} \tag{5}$$

其中，$(-k,\ L)$是相对时点l的范围，$D_{i,\ t}^l$是每一相对时点l是否接受处理，接受处理为1，未接受处理为0。实证中常剔除-1期这一相对时点作为基准，每一相对时点系数都表示为相对-1期这一时点系数的大小。假若政策前每一相对时点系数β_l^0都无法拒绝系数为零的假设，则满足平行趋势假定，政策后每一相对时点系数β_l^1反映的是政策效果随时间的变化。式（4）和式（5）的识别策略被广泛用于渐进性试点和多期试点政策研究中。

三、多时点DID存在的问题

最新研究指出在异质性处理效应下多时点DID估计结果有偏。Baker等[10]利用蒙特卡罗方法，分析了不同处理效应下多时点DID估计系数存在偏差的六种情况，并分别进行了模拟。①图1和图2为模拟1—模拟3的结果。由图1可知，模拟1和模拟2得到的估计系数围绕真实系数呈正态分布，表明在单时点DID情况下，无论处理效应是否随时间变化，估计系数都是无偏的。在存在多个处理时点，处理效应不随时间和处理组组别变化时，模拟3得到的估计系数依然无偏。

图3和图4为模拟4—模拟6的结果。由图3可知，在模拟4—模拟6中，多时点DID估计系数与真实系数存在偏差，偏差不断增大，并且在模拟6中估计系数符号与真实系数符号相反。

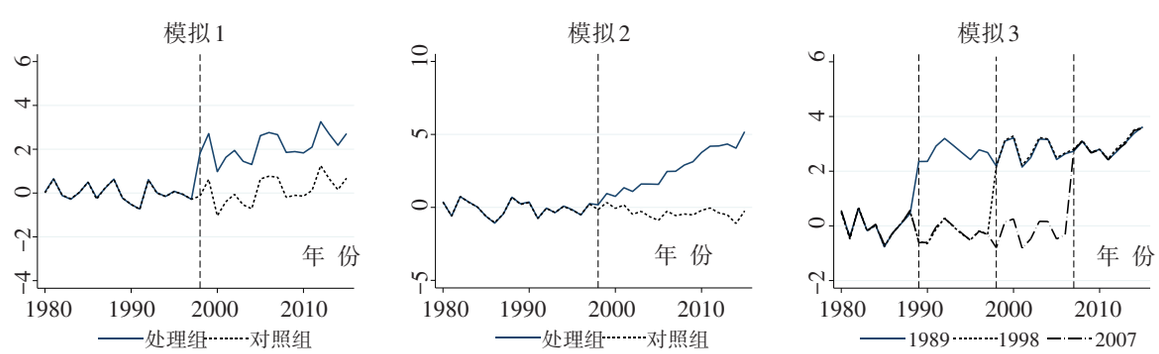

图1　模拟1—模拟3处理组和控制组结果均值的时间路径

① 研究中每个模拟都生成1980—2015年1 000家企业平衡面板数据，企业在50个城市均匀分布。在单时点情形下，假定1/2城市在1997年接受处理，1/2的城市未接受处理；在多时点情形下，假定1/3的城市在1989年接受处理，1/3的城市在1998年接受处理，剩余城市在2007年接受处理。研究的六种模拟情况：模拟1，单时点DID，假定处理组的处理效应恒定；模拟2，单时点DID，假定处理组处理效应随时间变化；模拟3，多时点DID，假定每个处理组组内的处理效应恒定，处理组之间的处理效应相同；模拟4，多时点DID，假定每个处理组组内处理效应恒定，处理组之间的处理效应不同；模拟5，多时点DID，假定每个处理组组内处理效应随时间变化，处理组之间处理效应的时间趋势完全相同；模拟6，多时点DID，假定每个处理组组内处理效应随时间变化，处理组之间的处理效应时间趋势不同。

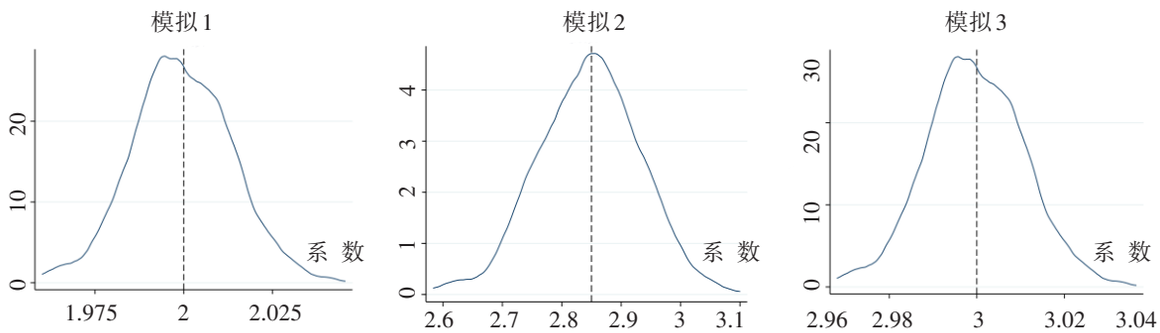

图2　模拟1—模拟3经过500次模拟所得系数的分布情况

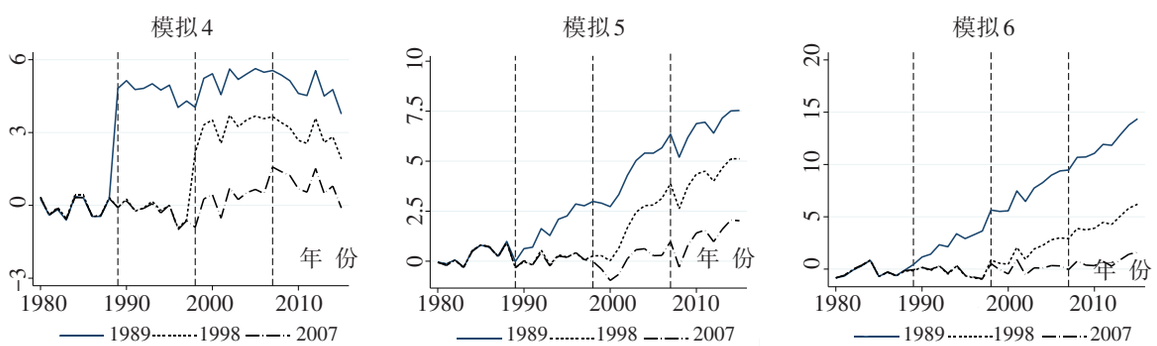

图3　模拟4—模拟6处理组和控制组结果均值的时间路径

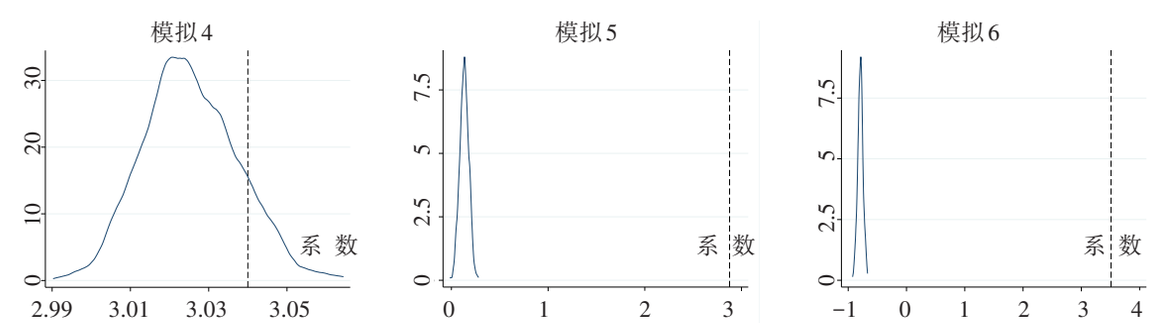

图4　模拟4—模拟6经过500次模拟所得系数的分布情况

异质性处理效应既包括处理效应随不同处理组组别发生变化，也包括同一个处理组在时间维度发生变化。无论处理效应是否随时间变化，单时点DID都不存在估计系数有偏问题。然而，由于早接受处理组在回归中被作为晚接受处理组的对照组，导致多时点DID估计系数有偏，尤其在异质性处理效应下，偏差会更大。不仅如此，估计系数识别的也不是受处理个体的平均处理效应，而是组别—时间处理效应的加权平均[1-5]。关于加权处理，Goodman-Bacon[2]对多时点DID系数的加权给出了直观解释。①考虑个体为N时期为T的平衡面板数据，假设存在2个政策时点：k和l，假设存在3个组别：早接受处理组k、晚接受处理组l和未受处理组U。

$(0，k)$时期为$PRE(k)$，$(k，l)$时期为$MID(k，l)$，$(l，T)$时期为$POST(l)$。在不考虑控制变量

① 参见Goodman-Bacon[2]附录中图1。

时，DID 回归方程为式（6）：

$$y_{it} = \beta^{DD} D_{it} + \lambda_i + \eta_t + \varepsilon_{it} \tag{6}$$

通过分解$\hat{\beta}^{DD}$，证明该系数可表示为 4 个 2×2 DID 系数的加权平均[2]，如式（7）所示：

$$\hat{\beta}^{DD} = s_{ku}\hat{\beta}^{2\times2}_{ku} + s_{lu}\hat{\beta}^{2\times2}_{lu} + s^{k}_{kl}\hat{\beta}^{2\times2,\ k}_{kl} + s^{l}_{kl}\hat{\beta}^{2\times2,\ l}_{kl} \tag{7}$$

其中，$s_{ku} = \frac{(n_k + n_U)^2 \hat{V}^D_{kU}}{\hat{V}^D}$，$s_{lu} = \frac{(n_l + n_U)^2 \hat{V}^D_{lU}}{\hat{V}^D}$，$s^k_{kl} = \frac{\left((n_k + n_l)(1-\bar{D}_l)\right)^2 \hat{V}^{D,\ k}_{kl}}{\hat{V}^D}$，$s^l_{kl} = \frac{\left((n_k + n_l)\bar{D}_k\right)^2 \hat{V}^{D,\ l}_{kl}}{\hat{V}^D}$表示每个系数的权重，$n_{j,\ j\in\{k,\ l,\ U\}}$是组别$j$样本量占总体样本量比重，$\bar{D}_{i,\ i\in\{k,\ l\}}$是$(i,\ T)$该时期占总时期$T$的比重，$\hat{V}^D$是总体方差，$\hat{V}^D_{kU}$，$\hat{V}^D_{lU}$，$\hat{V}^{D,\ k}_{kl}$，$\hat{V}^{D,\ l}_{kl}$是每个 2×2 DID 中$D_{it}$的方差。每一权重都由每组样本份额的平方和每组方差与总体方差之比两部分构成，4 个权重之和为 1。$\hat{\beta}^{2\times2}_{ku}$，$\hat{\beta}^{2\times2}_{lu}$，$\hat{\beta}^{2\times2,\ k}_{kl}$，$\hat{\beta}^{2\times2,\ l}_{kl}$分别是早接受处理组$k$和未受处理组$U$在$(0,\ T)$的系数、晚接受处理组$l$和未受处理组$U$在$(0,\ T)$的系数、早接受处理组$k$和晚接受处理组$l$在$(0,\ l)$的系数、晚接受处理组$l$和早接受处理组$k$在$(k,\ T)$的系数。①其中，$\hat{\beta}^{2\times2}_{ku}$，$\hat{\beta}^{2\times2}_{lu}$，$\hat{\beta}^{2\times2,\ k}_{kl}$含义是处理组事后与事前$y$均值的差异减去对照组事后与事前$y$均值的差异。$\hat{\beta}^{2\times2,\ l}_{kl}$是将早接受处理组$k$作为对照组，含义为晚接受处理组$l$事后和事前$y$均值的差异减去早接受处理组$k$作为对照组事后与事前$y$均值的差异。

由式（2）至（3）可知，$\hat{\beta}^{2\times2}_{ku}$，$\hat{\beta}^{2\times2}_{lu}$，$\hat{\beta}^{2\times2,\ k}_{kl}$分别是受处理个体平均处理效应+（处理组的时间趋势−对照组的时间趋势），在满足平行趋势假定条件下，系数识别的是受处理个体的平均处理效应。但$\hat{\beta}^{2\times2,\ l}_{kl}$不同，$\hat{\beta}^{2\times2,\ l}_{kl}$是受处理个体平均处理效应+（晚接受处理组$l$的时间趋势−早接受处理组$k$作为对照组的时间趋势）−早接受处理组$k$处理效应的时间趋势。若早接受处理组的处理效应随时间变化较大，$\hat{\beta}^{2\times2,\ l}_{kl}$可能为负值并最终影响总体系数$\hat{\beta}^{DD}$的估计。②因此，当$N$趋于无穷时，式（7）可以进一步表示为式（8）：

$$plim_{N\to\infty}\hat{\beta}^{DD} = \beta^{DD} = VWATT + VWCT - \Delta ATT \tag{8}$$

其中，*VWATT*（Variance-Weighted Average Treatment Effect on the Treated）是方差加权 ATT，*VWCT*（Variance-Weighted Common Trends）是方差加权的平行趋势，ΔATT是早接受处理组处理效应的时间趋势。在满足平行趋势假定和处理效应恒定（即ΔATT=0）条件下，多时点 DID 的估计系数为*VWATT*。当存在异质性处理效应时，多时点 DID 估计系数就会产生偏差。

更具体地，若将每个 2×2 DID 表述为是每个组，当每个组组内处理效应恒定，组间各自的处理效应也相同时，如图 1 模拟 3 所示，由于权重之和为 1，*VWATT*就是*ATT*；当每个组组内处理效应恒定，组间各自的处理效应不同时，如图 3 的模拟 4 所示，估计系数不再是 ATT，而是表现为*VWATT*，权重与每个组的样本份额和处理变量方差相关；当处理效应随时间和组别变化时，如图 3 的模拟 5 和模拟 6 所示，由于式（8）中还需减去ΔATT，因而估计系数会存在较大偏差。更为严重时，估计系数符号会与真实系数符号相反，无法准确识别政策效果。

学者们对多时点 DID 估计系数有偏的解释存在差别。Goodman-Bacon[2]认为，所有 2×2 DID 系数的权重为正，但部分系数符号为负，导致多时点 DID 估计系数存在偏差。Chaisemartin 和 D′Haultfoeuille[1]、Borusyak 和 Jaravel[3]以及 Borusyak 等[4]却认为，多时点 DID 系数是组别—时间处理效应的加权平均，组别—时间处理效应为正，但部分权重为负，导致最终估计结果有偏。本文以 Chaisemartin 和 D′Haultfoeuille[1]的研究来解释这类观点。考虑城市g-年份t-企业i层面的数据

① 参见 Goodman-Bacon[2]附录中图 2。

① 早接受处理组由于较早受到处理，随着时间推移政策效果可能会增强。晚接受处理组由于较晚受到处理，受政策影响可能较小，当早接受处理组作为对照组时，该 2×2 DID 的估计系数为负。

结构。假定每个城市每年至少有一家企业，处理发生在城市层面，城市受处理后所有企业也同样受到处理。令$Y_{i,\,c,\,t}(1)$和$Y_{i,\,c,\,t}(0)$是企业的潜在结果，$D_{c,\,t}$是城市的受处理状态，$D_{c,\,t}$为1表示城市c在年份t受到处理，否则为0，$D_{i,\,c,\,t}$为企业受处理状态。$N_1=\sum_{i,\,c,\,t}D_{i,\,c,\,t}$是所有受处理企业的观测值数量，所有受处理企业的平均处理效应为：$\Delta^{TR}=\frac{1}{N_1}\sum_{(i,\,c,\,t):\ D_{c,\,t}=1}\left[Y_{i,\,c,\,t}(1)-Y_{i,\,c,\,t}(0)\right]$。令$\delta^{TR}=\mathbb{E}(\Delta^{TR})$，$\delta^{TR}$是回归估计所要识别的ATT。将$N_{c,\,t}$是城市—时间组$cell(c,\,t)$的观测值数量，每个城市—时间组的平均处理效应定义为：$\Delta_{c,\,t}=\frac{1}{N_{c,\,t}}\sum_{i=1}^{N_{c,\,t}}\left[Y_{i,\,c,\,t}(1)-Y_{i,\,c,\,t}(0)\right]$。

δ^{TR}等价于所有受处理城市—时间组平均处理效应$\Delta_{c,\,t}$的加权平均，权重为每个城市—时间组样本与所有受处理企业样本之比，即式（9）：

$$\Delta^{TR}=\mathbb{E}\left[\sum_{(c,\,t):\ D_{c,\,t}=1}\frac{N_{c,\,t}}{N_1}\Delta_{c,\,t}\right] \tag{9}$$

但是在满足平行趋势假定的双向固定效应模型下，实际得到的估计系数为式（10）：

$$\beta_{fe}=\mathbb{E}\left[\sum_{(c,\,t):\ D_{c,\,t}=1}\frac{N_{c,\,t}}{N_1}w_{c,\,t}\Delta_{c,\,t}\right],$$

$$w_{c,\,t}=\frac{\varepsilon_{c,\,t}}{\sum_{(c,\,t):\ D_{c,\,t}=1}\frac{N_{c,\,t}}{N_1}\varepsilon_{c,\,t}} \tag{10}$$

其中，$w_{c,\,t}$等式中的$\varepsilon_{c,\,t}$由$D_{c,\,t}=\alpha+\gamma_c+\lambda_t+\varepsilon_{c,\,t}$得到。[①]$\frac{N_{c,\,t}}{N_1}w_{c,\,t}$是每个受处理城市—时间组平均处理效应的权重，由城市—时间组的样本份额，以及$\varepsilon_{c,\,t}$与$\varepsilon_{c,\,t}$均值之比两部分构成。式（9）和式（10）表明，β_{fe}与Δ^{TR}并不相等，估计系数识别的不是ATT，而是组别—时间处理效应的加权平均。部分组别—时间处理效应权重为负时，多时点DID的估计系数有偏。

考虑2个组别和3个时点的DID，第1组在第3时点受到处理，第2组在第2时点受到处理。$\varepsilon_{c,\,t}$可由式$\varepsilon_{c,\,t}=D_{c,\,t}-\bar{D}_c-\bar{D}_t+\bar{\bar{D}}$得到，其中$\bar{D}_g$是$g$城市企业受处理状态变量的均值，$\bar{D}_t$是$t$年企业受处理状态变量的均值，$\bar{\bar{D}}$是所有企业在所有年份受处理状态变量的均值。不同组别和不同时间的$\varepsilon_{c,\,t}$分别是$\varepsilon_{1,\,3}=\frac{1}{6}$，$\varepsilon_{2,\,2}=\frac{1}{3}$，$\varepsilon_{2,\,3}=-\frac{1}{6}$。$\varepsilon_{c,\,t}$均值为$(\varepsilon_{1,\,3}+\varepsilon_{2,\,2}+\varepsilon_{2,\,3})\times\frac{1}{3}=\frac{1}{9}$，$w_{c,\,t}$分别是$w_{1,\,3}=\frac{3}{2}$，$w_{2,\,2}=3$，$w_{2,\,3}=-\frac{3}{2}$，$\beta_{fe}$是$\beta_{fe}=\frac{1}{2}\mathbb{E}(\Delta_{1,\,3})+\mathbb{E}(\Delta_{2,\,2})-\frac{1}{2}\mathbb{E}(\Delta_{2,\,3})$，如图5所示。

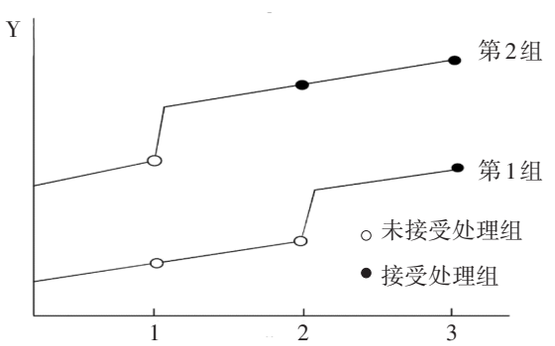

图5　2个组别3个时点的DID

① Chaisemartin和D′Haultfoeuille[1]的证明可以参考Borusyak等[4]附录中的命题2。

由图5可知，第2组在第3时点处理效应的权重为负，导致多时点DID估计系数有偏。负权重的来源可以通过Goodman-Bacon[2]的分解定理进一步解释。β_{fe}可分解为2个2×2 DID系数的加权平均，即$\beta_{fe}=(DID_1+DID_2)/2$。在满足平行趋势假定前提下，$DID_1$和$DID_2$分别为$\begin{cases}DID_1=\left[\mathbb{E}(Y_{2,2})-\mathbb{E}(Y_{2,1})\right]-\left[\mathbb{E}(Y_{1,2})-\mathbb{E}(Y_{1,1})\right]\\DID_2=\left[\mathbb{E}(Y_{1,3})-\mathbb{E}(Y_{1,2})\right]-\left[\mathbb{E}(Y_{2,3})-\mathbb{E}(Y_{2,2})\right]\end{cases}$，$DID_1$识别的是ATT，因而$DID_1=\mathbb{E}(\Delta_{2,2})$，$DID_2$表示为$DID_2=\mathbb{E}(\Delta_{1,3})-\left[\mathbb{E}(\Delta_{2,3})-\mathbb{E}(\Delta_{2,2})\right]$，即ATT减去早接受处理组处理效应的时间趋势，最终可得系数为$\beta_{fe}=\frac{1}{2}\mathbb{E}(\Delta_{1,3})+\mathbb{E}(\Delta_{2,2})-\frac{1}{2}\mathbb{E}(\Delta_{2,3})$，表明负的权重也是源于早接受处理组被作为晚接受处理组的对照组。在相同时间点上，受处理时间越长的城市，其处理效应的权重也越有可能为负[1]。当负权重非常大时，会导致估计结果产生大的偏差。

Sun和Abraham[6]研究表明，事件研究法可以解决图3中模拟4和模拟5存在的问题，但无法解决模拟6存在的问题，原因在于每一相对时点上的处理效应不仅与自身时点处理效应相关，还与回归中其他相对时点及被剔除在等式之外的其他相对时点处理效应相关。考虑个体为N，时期为$T+1$的样本。$Y_{i,t}$为结果变量，$Y_{i,e+l}$和$Y^{\infty}_{i,e+l}$分别为个体i在时点t的可观测结果和未接受处理的反事实结果。$D_{i,t}$为二值变量，表示个体受处理状态。将受处理个体i首次受到处理的时点表示为$E_i=\min\{t\,|D_{i,t}=1\}$，个体未接受处理的时点表示为$E_i=\infty$，队列（cohort）$e$为首次接受处理时点$E_i$相同的所有个体集合。相对时点为$l$，$l=t-e$。Sun和Abraham[6]将受处理队列$e$在相对时点$l$的平均处理效应（Cohort-Specific Average Treatment Effect on the Treated，CATT）定义为式（11）：

$$CATT_{e,l}=\mathbb{E}\left[Y_{i,e+l}-Y^{\infty}_{i,e+l}|E_i=e\right] \tag{11}$$

Sun和Abraham[6]基于这一定义得出了主要结论。假设$T=3$，存在3个队列。①相对时点l，$l\in\{-3,-2,-1,0,1,2\}$。回归中常剔除第-1期作为基期，在该例中，由于没有未受处理队列，还需额外再剔除1期，最终剔除第-3期和第-1期。②以μ_{-2}表示第-2期系数，在满足平行趋势假定条件下，经证明μ_{-2}可分解为式（12）至式（14）：

$$\mu_{-2}=\sum_{e\in\{2,3\}}\omega_{e,-2}CATT_{e,-2} \tag{12}$$

$$+\sum_{e\in\{1,2,3\}}\omega_{e,0}CATT_{e,0}+\sum_{e\in\{1,2\}}\omega_{e,1}CATT_{e,1}+\sum_{e\in\{1\}}\omega_{e,2}CATT_{e,2} \tag{13}$$

$$+\sum_{e\in\{1,2,3\}}\omega_{e,-1}CATT_{e,-1}+\sum_{e\in\{3\}}\omega_{e,-3}CATT_{e,-3} \tag{14}$$

其中，ω是$CATT$权重，式（12）是只和第-2期相关$CATT$的加权平均，所有权重之和为1；式（13）是回归等式中除第-2期之外其他时期$CATT$的加权平均，所有权重之和为0；式（14）是被剔除在回归等式之外其他时期$CATT$的加权平均，所有权重之和为-1。

若处理前不存在预期性行为，即个体不会在处理前受到影响，当$l<0$时，对于任意队列e，$CATT_{e,l}=0$，式（12）和式（14）都为0。在满足无预期性行为假设下，存在两类情形：（1）假设不同队列处理效应的时间趋势相同，如图3的模拟5所示，对于相对时点l的任意队列e，$CATT_{e,l}=CATT_l$，由于权重之和为0，所以式（13）为0，最终第-2期系数为0；（2）假设不同队列处理效应的时间趋势不同，如图3的模拟6所示，第-2期系数非0，表明这一相对时点的系数有偏。同理，分解事前其他相对时点和事后相对时点系数也会存在偏误。在现实案例中，不同队列更多表现为

① 由于篇幅所限，没有列示该表，留存备索。

② 当所有队列都接受处理时，因式(6)中$D^l_{it}=1\{t-e_i=l\}$，$l=t-e_i$本身会存在共线，所以除了剔除第-1期作为基期外，为了避免共线性还需再额外剔除一期。

异质性处理效应。这种情况下利用相对时点系数大小判别是否满足平行趋势假定，以及检验处理效应的动态变化是存在问题的。

四、多时点DID估计有偏的解决方法

针对在异质性处理效应下多时点DID存在的问题，本文将解决方法总结为：诊断方法，也就是系数分解定理，三类解决方法分别为加总方法、两步回归法及堆叠型DID。

（一）系数分解定理

由Goodman-Bacon[2]的DID系数分解可知，当存在2个处理时点加一个对照组时，多时点DID的估计系数表示为4个2×2 DID系数的加权平均。当存在K个不同的政策时点加一个对照组时，多时点DID的估计系数可表示为K×K个系数的加权平均。通过观察分解的不同系数大小和系数权重，即可诊断多时点DID存在的偏误多大程度会影响最终估计结果。

关于系数分解定理。Stevenson和Wolfers[11]研究了1969—1985年美国部分州推行的无过错离婚法案对妇女自杀率的影响，结果表明无过错离婚法案实行后女性的自杀率有所降低。在不考虑任何控制变量情形下，多时点DID估计系数可以分解为156个2×2 DID系数的加权平均。

图中所有三角和叉号表示的系数符号为负，且系数无偏，与整体估计系数的符号一致。菱形和圆圈表示的系数会令整体估计有偏，因为这部分样本把早接受处理组作为晚接受处理组的对照组，并且所有菱形表示的系数符号为正，与整体估计系数的符号相反。菱形和圆圈两部分所占比重较高，最终会严重低估法案实施后对女性自杀率的影响。因此，通过分解系数大小与系数所占权重可以诊断多时点DID的估计偏差。无过错离婚法案对女性自杀率回归系数的分解如图6所示。

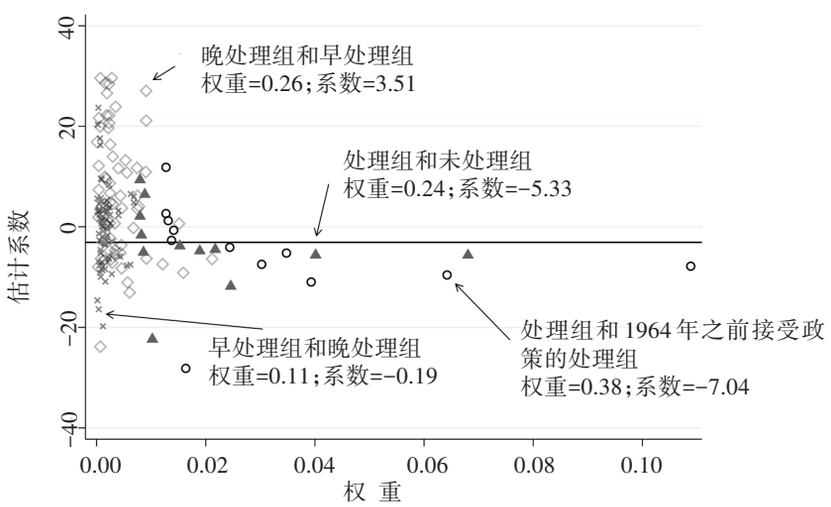

图6　无过错离婚法案对女性自杀率回归系数的分解

Goodman-Bacon[2]的系数分解定理并非解决方法，回归中加入控制变量后得到的图形也无法区分早接受处理组和晚接受处理组和晚处理组和早处理组的分解，但由于实证文章中通常汇报不加控制变量的回归结果，Baker等[10]认为此方法具有一般适用性，可以用于诊断只包括双向固定效应的回归结果偏差。

（二）加总方法

Sun和Abraham[6]对事件研究法处理多时点DID时存在的问题提出了相应解决方法，即求得每一相对时点下所有队列的$CATT_{e,l}$，用样本份额对这一时点下所有$CATT_{e,l}$加权平均。Sun和

Abraham[6] 将其提出的估计量称之为交互加权估计量（Interaction-Weighted Estimator，IW Estimator），分三步求得。第一步，利用双向固定效应回归估计$CATT_{e,\ l}$，回归等式为式（15）：

$$Y_{i,\ t} = \lambda_i + \eta_t + \sum_{e \notin C} \sum_{l \neq -1} \delta_{e,\ l} \left(\mathbf{1}\{E_i = e\} \cdot D_{i,\ t}^{l} \right) + \varepsilon_{i,\ t} \tag{15}$$

其中，C表示对照组，$D_{i,\ t}^{l}$表示是否为相对时点l，$\mathbf{1}\{E_i = e\}$表示是否为队列e。与每一相对时点l作对比的是相对时点-1，与每一组队列e作对比的是对照组C. 因此$\delta_{e,\ l}$识别的是每一相对时点l下每一队列e的平均处理效应$CATT_{e,\ l}$。[①]第二步，估计每个队列e在相对时点l的权重$Pr\{E_i = e | E_i \in [-l,\ T-l]\}$，该权重含义为：每个队列$e$在相对时点$l$下的样本占相对时点$l$里所有队列样本的比重。[②]第三步，每一相对时点的交互加权估计量表示为$\hat{v}_l = \sum_e \hat{\delta}_{e,\ l} \widehat{Pr}\{E_i = e | E_i \in [-l,\ T-l]\}$。这一方法可以得到每一相对时点$l$准确的估计系数。

类似地，Callaway 和 Sant'Anna[7] 提出的解决方法依赖于定义的组别—时点平均处理效应（Group-time Average Treatment Effect），将组别—时点平均处理效应加总，即可得到总的处理效应。组别—时点平均处理效应定义为式（16）：

$$ATT(g,\ t) = \mathbb{E}\left[Y_t(g) - Y_t(0) | G_g = 1 \right] \tag{16}$$

其中，g是个体首次受到处理的时点，G是首次接受处理时点为g的个体集合。G_g表示是否是首次接受处理时点为g的组别G。$Y_t(g)$是首次接受处理时点为g的组别在时点t的结果，$Y_t(0)$是这一组别在t这一时点假若未经处理的潜在结果。Callaway 和 Sant'Anna[7] 考虑了两类对照组，一类是将从未接受处理个体作为对照组，另一类是将从未接受处理个体以及尚未接受处理个体作为对照组。在满足无预期性行为假设和平行趋势假设条件下，若其他变量都不影响个体受处理状态，[③]则可以利用两个不同的对照组分别求得组别—时点平均处理效应，如式（17）：

$$\begin{aligned} ATT(g,\ t)_{unc}^{nev} &= \mathbb{E}\left[Y_t - Y_{g-1} | G_g = 1 \right] - \mathbb{E}\left[Y_t - Y_{g-1} | C = 1 \right] \\ ATT(g,\ t)_{unc}^{ny} &= \mathbb{E}\left[Y_t - Y_{g-1} | G_g = 1 \right] - \mathbb{E}\left[Y_t - Y_{g-1} | D_t = 0 \right] \end{aligned} \tag{17}$$

其中，nev表示将未接受处理个体作为对照组（$C = 1$），ny表示将从未接受处理个体以及未接受处理个体作为对照组（$D_t = 0$）。unc表示不考虑控制变量，$g-1$是首次接受处理时点g的前一时点。如表1所示，假设存在4个个体，6个时期，个体1和个体2未接受处理，个体3在t=3接受处理，个体4在t=4接受处理，表中数字表示个体的结果Y。根据上式可得：$\left\{ \begin{array}{l} ATT(3,\ 4)_{unc}^{nev} = (8-4) - ((5-3) + (4-2))/2 = 2 \\ ATT(3,\ 4)_{unc}^{ny} = (8-4) - ((5-3) + (4-2) + (7-5))/3 = 2 \end{array} \right\}$。

然而，个体受处理状态通常是非随机的，在考虑控制变量情形下，三种方法可以获得$ATT(g,\ t)$：结果回归方法（Outcome Regression，OR）、逆概率加权方法（Inverse Probability Weighting，IPW）、双重稳健方法（Doubly Robust，DR）。三种方法得到的$ATT(g,\ t)$在识别上相同，若需要进一步做统计推断，双重稳健方法会更加稳健[7]。

依靠求得的$ATT(g,\ t)$，通过不同形式加总，可以得到总体的平均处理效应、处理效应的动态变化、不同处理组组间处理效应、累积的平均处理效应。[④]相应表达式如式（18）至式（21）：

① Sun 和 Abraham[6] 表明若总体样本中存在从未接受处理的个体，则其为对照组；若总体样本中没有从未接受处理的个体，则将最后接受处理的个体作为对照组；若总体样本有一直接受处理的个体，则将这一组剔除。

② 其中，队列2在相对时点0的样本份额为$Pr\{E_i = 2 | E_i \in [0,3]\} = Pr\{E_i = 2 | E_i \in [1,3]\} = n_1/(n_1 + n_2 + n_3)$，其中，$n_{i,i \in \{1,2,3\}}$为每个队列$i$的样本份额。

③ 此处考虑的是有限预期行为。本文未考虑有限预期行为。

④ 在对$ATT(g,t)$加总时，作者是将从未接受过处理的组作为对照组。

$$\theta_{es}(e)=\sum_{g\in\mathbb{G}}\mathbf{1}\{g+e\leqslant T\}P(G=g|G+e\leqslant T)ATT(g,\ g+e) \tag{18}$$

式（18）为每一相对时点e的系数。其中，$\mathbf{1}\{g+e\leqslant T\}$是指示函数，$G$是所有个体首次接受处理的时点集合。以相对时点$e=-2$为例，$\theta_{es}(-2)$表示为$\theta_{es}(-2)=\mathbf{1}\{1\leqslant 6\}P\{G=3|G-2\leqslant 6\}ATT(3,\ 3-2)+\mathbf{1}\{3\leqslant 6\}P\{G=5|G-2\leqslant 6\}ATT(5,\ 3-2)=P\{G=3|G\leqslant 8\}ATT(3,\ 1)+P\{G=5|G\leqslant 8\}ATT(5,\ 3)$。因此，每一相对时点$e$的系数为所有不同组别$G=g$在相对时点$ATT(g,\ t)$的加权平均，权重为每个组别的样本份额。

在$[e,\ e']$数据结构平衡的相对时期里，每个相对时点e的系数为式（19）：

$$\theta_{es}^{bal}(e;\ e')=\sum_{g\in\mathbb{G}}\mathbf{1}\{g+e'\leqslant T\}P(G=g|G+e'\leqslant T)ATT(g,\ g+e) \tag{19}$$

其中，$\theta_{es}^{bal}(e;\ e')$表示为在$e$到$e'$范围内相对时点$e$的系数。相对时点$e$在受处理时点附近范围内，会包含所有受处理队列。当相对时点e'距离受处理时点较远时，可能只包含部分受处理队列。由于所含队列数量不同，其系数大小难以解释，因而可以保留包含所有受处理队列的相对时期。

每个组别$G=\tilde{g}$的处理效应为组别内所有受处理时点$ATT(\tilde{g},\ t)$的均值，如式（20）所示：

$$\theta_{sel}(\tilde{g})=\frac{1}{T-\tilde{g}+1}\sum_{t=\tilde{g}}^{T}ATT(\tilde{g},\ t) \tag{20}$$

总体系数为每个组别$G=\tilde{g}$处理效应$\theta_{sel}(\tilde{g})$的加权平均，权重为每个组别的样本份额，如式（21）所示：

$$\theta_{sel}^{O}=\sum_{g\in\mathbb{G}}\theta_{sel}(\tilde{g})P(G=g|G\leqslant T) \tag{21}$$

上述两种方法均是先获得组别—时点平均处理效应，然后加总得到总的处理效应。两者的不同之处在于：第一，Sun和Abraham[6]定义的是相对时点上不同队列的平均处理效应，可以加总得到每一相对时点系数，但Callaway和Sant′Anna[7]定义的是正常时点上不同队列的平均处理效应，将其加总可得相对时点的平均处理效应、不同组别的平均处理效应、总体的平均处理效应，适用更多情形；第二，Sun和Abraham[6]的解决方法没有考虑控制变量对结果的影响，反观Callaway和Sant′Anna[7]提出的获得$ATT(g,\ t)$的三种方法都依赖于控制变量，因而更具有一般性。

表1　个体为4时期为6的数据结构

	t=1	t=2	t=3	t=4	t=5	t=6
个体1	1	2	3	4	5	6
个体2	2	3	4	5	6	7
个体3	3	4	6	8	10	12
个体4	4	5	6	7	10	13

注：表中下划线数字表示个体接受处理结果，其他数字表示个体未受处理结果。

（三）两步回归法

Gardner[8]开发了两阶段回归法（Two-Stage Regression Approach），与Borusyak等[4]开发的方法类似，这是因为Borusyak等[4]认为Gardner[8]构造的估计量采取了“插补”形式（Imputation Form），且通过两步可以获得无偏的估计系数。本文将这两种方法统称为两步回归法。

Gardner[8]首先对多时点DID存在的问题给出直观解释，然后在此基础上提出通过两步法解决多时点DID存在的问题。考虑个体i和时间t层面的数据结构，将受处理时点相同的个体归为g组，$g\in\{0,\ 1,\ \ldots,\ G\}$，$g=0$是对照组，若$g=2$，是处理时点发生在$t=2$时的个体集合。$g$组受处理后的时期为$p$，$p\in\{0,\ 1,\ \ldots,\ P\}$，$p=0$是受处理之前，若$p=2$，是组别$g=2$接受处理后的时期为2。$Y_{gpit}$、$Y_{1gpit}$和$Y_{0gpit}$分别是个体可观测结果、处理组潜在结果和对照组潜在结果。$D_{g,\ p}$表示组别g在时期p是否接受处理。

当存在两个时期和两个组别（包括一个处理组和一个对照组）时，2×2 DID回归等式为$Y_{gpit}=\lambda_g+\gamma_p+\beta_{gp}D_{gp}+\varepsilon_{gpit}$，系数$\beta_{gp}$识别的是受处理个体的平均处理效应，即$\beta_{gp}=\beta_{11}=\mathbb{E}\left(Y_{1gpit}-\right.$

$Y_{0gpit}|D_{gp}=1\big)$，回归等式用均值形式表示为式（22）：

$$\mathbb{E}(Y_{gpit}|g,\ p,\ D_{gp})=\lambda_g+\gamma_p+\beta_{gp}D_{gp} \tag{22}$$

当存在多个处理组时，受处理个体的平均处理效应为$\mathbb{E}(\beta_{gp}|D_{gp}=1)=\mathbb{E}\big(Y_{1gpit}-Y_{0gpit}|D_{gp}=1\big)$，即多个处理组平均处理效应的均值。多时点DID回归等式用均值形式表示，如式（23）所示：

$$\mathbb{E}(Y_{gpit}|g,\ p,\ D_{gp})=\lambda_g+\gamma_p+\mathbb{E}(\beta_{gp}|D_{gp}=1)D_{gp}+\big[\beta_{gp}-\mathbb{E}(\beta_{gp}|D_{gp}=1)\big]D_{gp} \tag{23}$$

其中，$\mathbb{E}(\beta_{gp}|D_{gp}=1)$是需要识别的受处理个体的平均处理效应，$\beta_{gp}$是回归得到的估计系数。由前文可知，$\beta_{gp}$估计的是多个处理组处理效应的加权平均，$\beta_{gp}$与$\mathbb{E}(\beta_{gp}|D_{gp}=1)$并不相等。当仅有一个处理组时，两项相减为0。当处理效应为同质时，β_{gp}等于受处理个体平均处理效应，两项相减也为0。因此，最后一项可看作扰动项，其会影响多时点DID的最终估计结果。

Gardner[8]提出通过两阶段回归法解决多时点DID存在的问题。第一阶段，保留$D_{gp}=0$的样本，用Y_{gpit}对λ_g和γ_p进行回归；第二阶段，在全样本里将$Y_{gpit}-\hat{\lambda}_g-\hat{\gamma}_t$对$D_{gp}$回归，最终估计系数识别的是受处理个体的平均处理效应。该方法的优点在于易于操作，并且第一阶段回归中也可加入控制变量。此外，该方法还可以拓展到事件研究法的应用中，在第二阶段将D_{gp}变为如式（5）中的相对时点虚拟变量即可。虽然这一方法通过回归得到无偏的估计系数，但在统计推断时还需对标准误进行调整，为此原文作者提供了Stata命令包did2s供实证研究者使用。

相比于Gardner[8]的两阶段回归法，Borusyak等[4]提出的方法更为直观。若处理组和对照组事前的差异可以完全由个体固定效应和时间固定效应捕捉，在控制双向固定效应后，对照组可以看作是处理组的反事实状态。政策后某一时点t上，处理组个体结果减去对照组个体结果，两者的差值便是受处理个体i在时间t的处理效应，通过对这一处理效应加权平均，即可得到所有受处理个体的平均处理效应。类似地，若处理组和对照组事前的差异还受其他变量影响，通过控制这些变量也可得到所有受处理个体的平均处理效应。

该方法可以通过两步得到多时点DID的无偏估计系数，本文以只考虑固定效应情形加以说明。第一步，保留没有受到处理的样本，回归估计处理组个体未接受处理的潜在结果$Y_{it}(0)$，$Y_{it}(0)=\lambda_i+\eta_t+\varepsilon_{it}$。第二步，对于处理组样本，令$\hat{Y}_{it}(0)=\hat{\lambda}_i+\hat{\eta}_t$，个体$i$在时间$t$的处理效应表示为$\hat{\tau}_{it}=Y_{it}-\hat{Y}_{it}(0)$。最终的总体估计系数为所有受处理个体处理效应的加权平均，权重为样本份额。

（四）堆叠型DID

Cengiz等[9]在研究最低工资对就业的影响时，使用了堆叠型DID。堆叠型DID是指在受处理时点前后$(-j,\ k)$范围内，为受处理时点相同的队列寻找干净的对照组并形成一个数据集，数据集中包括这段时期内受处理个体的样本、从未接受处理个体的样本和尚未接受处理个体的样本，之后将所有数据集堆叠并进行回归。回归方程设定为式（24）：

$$Y_{mit}=\sum_{l=-j}^{k}\beta_l D_{mit}^{l}+\lambda_{mi}+\eta_{mt}+\varepsilon_{mit} \tag{24}$$

其中，m是数据集，i是个体固定效应，t是时间固定效应，D_{mit}^{l}表示是否为相对时点l，λ_{mi}为数据集—个体联合固定效应，η_{mt}是数据集—时间联合固定效应。回归系数β_l的大小可以观测处理效应的动态变化。虽然这一方法较为直观，但并没有相应的理论证明。

总结而言，Goodman-Bacon[2]的分解定理应用于诊断多时点DID系数的偏差。本文比较了不考虑控制变量时的不同解决方法。①在这一数据生成设定下，这几类方法分别得到的相对时点估计

① 由于篇幅所限，没有列示该图，留存备索。数据为模拟6中1980—2006年样本，其中包括1989年和1998年两个政策时点，回归中未加入控制变量。Sun和Abraham[6]动态趋势所用Stata命令为eventstudyinteract；Callaway和Sant'Anna[7]动态趋势所用Stata命令为csdid；Gardner[8]动态趋势所用Stata命令为did2s；堆叠型DID的动态趋势为手工合成数据后的回归结果。

系数相近，与真实系数也较为接近，然而不同解决方法仍存在差异。虽然堆叠型DID在实证研究中得到了应用，但缺乏理论证明，且该方法需要手工合成数据，较为繁琐。Sun和Abraham[6]提出的方法只在事件研究法下适用，也没有考虑加入控制变量的情形。Callaway和Sant′Anna[7]、Gardner[8]、Borusyak等[4]提出的方法既可以估计多时点DID系数，也可以应用于事件研究法，回归中还可以加入控制变量，因而更具有一般性。

五、结论与建议

多时点DID识别的不是受处理个体的平均处理效应，而是组别—时间处理效应的加权平均。特别是，如果存在异质性处理效应，多时点DID的估计系数与真实系数会存在偏差。依据文献给出的解决思路，本文将解决方法归纳为一个诊断方法和三类解决方法。

随着DID在经济学实证研究中的广泛应用，学者有必要了解多时点DID问题的应对方式，以保证实证中估计系数的一致性。基于此，本文给出以下四点建议：第一，重视对政策背景和研究设计的讨论，清晰地阐述不同组别受政策影响的可能方向；第二，在处理组较少时，可以通过观察不同处理组和对照组的时间趋势，以此检验异质性处理效应对结果的影响及平行趋势假定是否满足；第三，如果是面板数据结构，可利用系数分解定理评估系数偏差大小；第四，从实践角度出发，在利用多时点DID作为识别策略时，研究者应至少在以上三类解决方法中选择其一，加强实证结果的可靠性和稳健性。

参考文献：

[1] CHAISEMARTIN C D, D′HAULTFOEUILLE X. Two-way fixed effects estimators with heterogeneous treatment effects [J]. American economic review, 2020, 110(9): 2964-2996.
[2] GOODMAN-BACON A. Difference-in-differences with variation in treatment timing [J]. Journal of econometrics, 2021, 225(2): 254-277.
[3] BORUSYAK K, JARAVEL X. Revisiting event study designs [R]. SSRN working paper, 2017.
[4] BORUSYAK K, JARAVEL X, SPIESS J. Revisiting event study designs: robust and efficient estimation [R].CEPR discussion papers, 17247.
[5] ATHEY S, IMBENS G W. Design-based analysis in difference-in-differences settings with staggered adoption [J]. Journal of econometrics, 2022, 226(1): 62-79.
[6] SUN L, ABRAHAM S. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects [J]. Journal of econometrics, 2021, 225(2): 175-199.
[7] CALLAWAY B, SANT′ANNA P H C. Difference-in-differences with multiple time periods [J]. Journal of econometrics, 2021, 225(2): 200-230.
[8] GARDNER J. Two-stage differences in differences [R]. Working paper, 2021.
[9] CENGIZ D, DUBE A, LINDNER A, et al. The effect of minimum wages on low-wage jobs [J]. The quarterly journal of economics, 2021, 134(3): 1405-1454.
[10] BAKER A C, LARCKER D F, WANG C C Y. How much should we trust staggered difference-in-differences estimates? [J]. Journal of financial economics, 2022, 144(2): 370-395.
[11] STEVENSON B, WOLFERS J. Bargaining in the shadow of the law: divorce laws and family distress [J]. The quarterly journal of economics, 2006, 121(1): 267-288.
[12] SANT′ANNA P H C, ZHAO J. Doubly robust difference-in-differences estimators [J]. Journal of econometrics, 2020, 219(1): 101-122.

Potential Problems and Solutions of Staggered Difference-in-Difference Approach

WANG Peng-chao，HAN Li-bin

（Economic and Social Development Research Institute， Dongbei University of Finance and Economics，
Dalian 116025， China）

Summary： As one of the mainstream causal inference methods， difference-in-difference approach has the characteristics of quasi-natural experiment and can relatively exogenously identify the causal effect， so it is favored by scholars at home and abroad. The interpretation of treatment effect estimated by difference-in-difference approach with a single treatment period is well known， but there are only a few studies discussing the interpretation and accuracy of treatment effect estimated by staggered difference-in-difference approach. Recently， some latest studies discuss these problems in detail. This paper， by means of sorting such literature， summarizes the interpretation of treatment effect， potential problems， and corresponding solutions of staggered difference-in-difference approach.

The latest literature shows that the coefficient estimated by difference-in-difference approach with a single treatment period is unbiased regardless of the heterogeneity effect. Staggered difference-in-difference approach identifies the weighted average of different group-time treatment effects. The estimated coefficient is unbiased with homogeneity treatment effect but biased with heterogeneous treatment effect. Because some early-treated groups are taken as the control groups of the late-treatment groups， the estimated coefficients of this part are negative and finally result in the bias of the aggregated coefficient. In severe cases， the symbols of both estimated coefficient and real coefficient are opposite. According to the solutions given by the latest literature， the methods to solve the bias of estimated coefficient can be divided into 'a diagnosis method' and 'three kinds of solutions'. The diagnosis method is Goodman-Bacon decomposition theorem. It is used to diagnose the degree of bias by estimating the sizes and weights of different group-time treatment effect. The first is the aggregation method including two ways， all of which are to find comparable control groups for each treatment group and estimate each group-time treatment effect. Then the unbiased estimated coefficient can be obtained by averaging all the group-period effects weighted by the sample share. The second is two-step regression method involving two ways. The reason for uniformly terming the two ways as this name lies in that they are similar in the solutions and the unbiased coefficient can be gained by two steps. The third is the stacked difference-in-difference approach. It aims to find comparable control group for cohorts with the same treatment period and form a data set. This dataset includes the samples of treated group， never-treated group， and not-yet-treated group. Then it is to stack all the data sets and regress an augmented difference-in-difference specification.

With the wide application of staggered difference-in-difference approach in empirical research of economics， it is necessary for empirical researchers to know how to deal with the problems of this method.

Key words： difference-in-difference ； staggered difference-in-difference ； heterogeneous treatment effect ； group-time treatment effect

（责任编辑：李明齐）