微观计量经济学
Heterogeneous difference-in-differences estimation

阮 睿

中央财经大学
中国财政发展协同创新中心

May 13, 2024

Contents

1 为什么 heterogeneous treatment 会产生影响？

1.1 交错 DID

- Static TWFE specification:

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

- $\hat{\beta}$ 识别的是什么？
- -定义处理组 k 在 W 期间的平均处理效应，即是一种 ATET

$$ATT_k(W) = \frac{1}{|W|}\sum_{t\in W}\mathbb{E}\left(Y_{it}^k - Y_{it}^0 \mid T_i^* = k\right)$$

- 定义未经处理的潜在结果 Y^0 的变化趋势

$$\Delta Y_k^0\left(W_1, W_0\right) = \frac{1}{|W_1|}\sum_{t\in W_1}\mathbb{E}\left(Y_{it}^0 \mid T_i^* = k\right) - \frac{1}{|W_0|}\sum_{t\in W_0}\mathbb{E}\left(\hat{Y}_{it}^0 \mid T_i^* = k\right)$$

对于多期面板，可以采用灵活估计

$$Y_{it} = \sum_{l=2}^{T} \beta_l \cdot \left(D_i \times T_t^l\right) + u_i + \eta_t + \varepsilon_{it}$$

$$T_t^l = \begin{cases} 1 & t = l \\ 0 & t \neq l \end{cases}$$

易见

$$T_t = \sum_{l \geq T^*} T_t^l$$

我们应该期望得到

$$\beta_2 \approx \beta_3 \approx \cdots \beta_{T^*-1} \approx 0; \beta_{T^*}, \beta_{T_1^*+1}, \cdots, \beta_T \neq 0$$

- 如果处理对不同的个体发生在不同的时间点，就叫做“staggered did”，交错型 DID
- 交错型 DID(staggered DID):T^* 因个体而异，即处理组个体接受处理时间不一致 (variation in treatment timing)。
- Bacon decomposition
  Goodman-Bacon, Difference-in-differences with variation in treatment timing, Journal of Econometrics, 2021

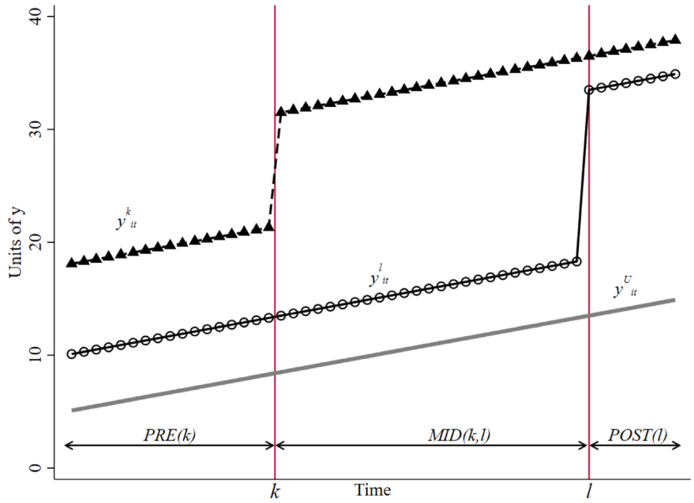


Fig. 1. Difference-in-Differences with variation in treatment Timing: Three groups. Notes: The figure plots outcomes in three timing groups: an untreated group, U; an early treatment group, k, which receives a binary treatment at $k = \frac{34}{100}T$; and a late treatment group, ℓ, which receives the binary treatment at $\ell = \frac{85}{100}T$. The x-axis notes the three sub-periods: the pre-period for timing group k, $[1, \mathrm{k} - 1]$, denoted by $PRE(k)$; the middle period when timing group k is treated and timing group ℓ is not, $[k, \ell - 1]$, denoted by $MID(k, \ell)$; and the post-period for timing group ℓ, $[\ell, T]$, denoted by $POST(\ell)$. The treatment effect is 10 in timing group k and 15 in timing group ℓ.

此时直接构造表示接受处理的虚拟变量。

$$Y_{it} = \beta \cdot D_{it} + u_{it} + \eta_t + \varepsilon_{it}$$

$$D_{it} = \begin{cases} 1, & \text{个体}i\text{ 在第}t\text{ 期接受处理} \\ 0, & \text{其它情形} \end{cases}$$

根据 Frisch-Waugh-Lovell 定理，定义

$$\tilde{\tilde{D}}_{it} = D_{it} - \bar{D}_i - \bar{D}_t + \bar{\bar{D}}$$

$$\tilde{\tilde{Y}}_{it} = Y_{it} - \bar{Y}_i - \bar{Y}_t + \bar{\bar{Y}}$$

$$\hat{\beta} = \frac{\widehat{\mathbf{Cov}}\left(\tilde{\tilde{D}}_{it}, \tilde{\tilde{Y}}_{it}\right)}{\widehat{\mathbf{Var}}\left(\tilde{\tilde{D}}_{it}\right)} = \frac{\frac{1}{NT}\sum_i \sum_t \tilde{\tilde{D}}_{it} \cdot \tilde{\tilde{Y}}_{it}}{\frac{1}{NT}\sum_i \sum_t \tilde{\tilde{D}}_{it}^2}$$

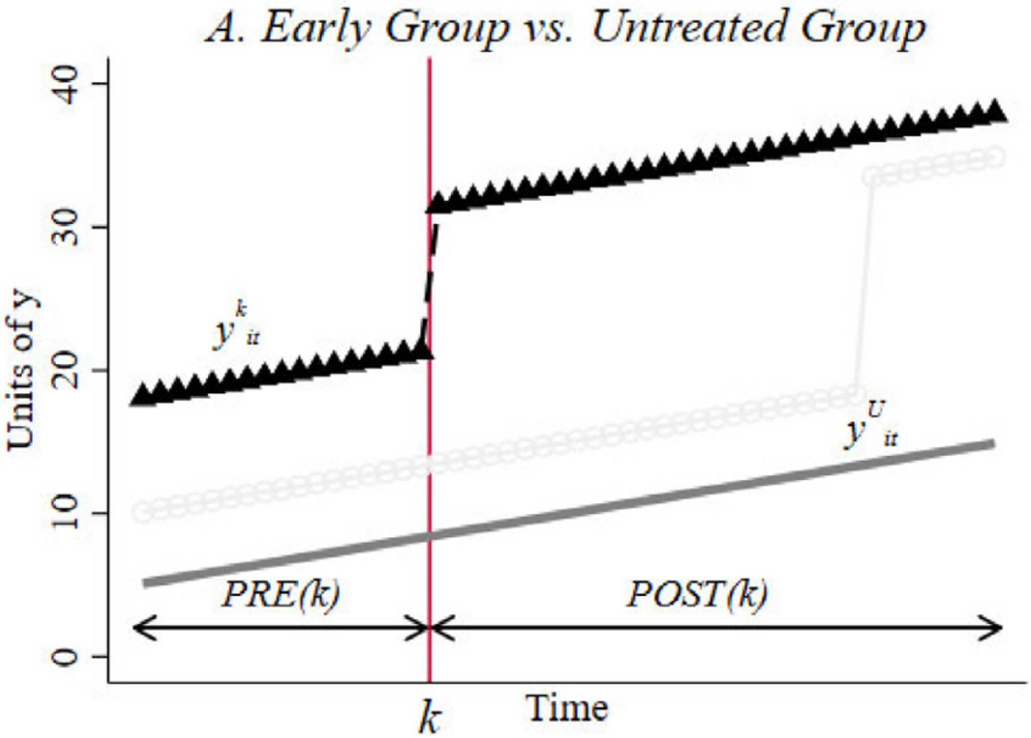
- $\hat{\beta}$ 识别的是什么？
- 定义处理组 k 在 W 期间的平均处理效应

$$ATT_k(W) = \frac{1}{|W|} \sum_{t \in W} \mathbb{E}\left(Y_{it}^k - Y_{it}^0 \mid T_i^* = k\right)$$

- 定义未经处理的潜在结果 Y^0 的变化趋势

$$\Delta Y_k^0\left(W_1, W_0\right) = \frac{1}{|W_1|} \sum_{t \in W_1} \mathbb{E}\left(Y_{it}^0 \mid T_i^* = k\right) - \frac{1}{|W_0|} \sum_{t \in W_0} \mathbb{E}\left(\hat{Y}_{it}^0 \mid T_i^* = k\right)$$

$$\begin{aligned}\beta_{ku} =& ATT_k(\mathbf{POST}(k)) \\ &+ \left[\Delta Y_k^0(\mathbf{POST}(k), \mathbf{PRE}(k)) - \Delta Y_u^0(\mathbf{POST}(k), \mathbf{PRE}(k))\right]\end{aligned}$$



$$
\begin{aligned}
\beta_{lk} = & ATT_l(\mathbf{POST}(l)) \\
& + \left[\Delta Y_l^0(\mathbf{POST}(l), \mathbf{MID}(k,l)) - \Delta Y_k^0(\mathbf{POST}(l), \mathbf{MID}(k,l))\right] \\
& - \left[ATT_k(\mathbf{POST}(l)) - ATT_k(\mathbf{MID}(k,l))\right]
\end{aligned}
$$

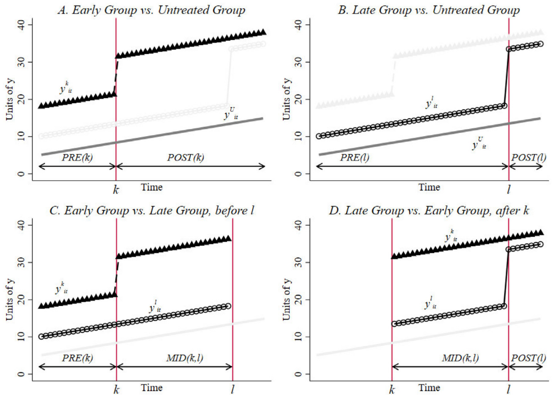


Fig. 2. The four simple (2x2) difference-in-differences estimates in the three group case. Notes: The figure plots outcomes for the subsamples that generate the four simple 2x2 difference-in-difference estimates in the three timing group case from Fig. 1. Each panel plots the data structure for one 2x2 DD. Panel A compares early treated units to untreated units ($\hat{\beta}_{kU}^{DD}$); panel B compares late treated units to untreated units ($\hat{\beta}_{\ell U}^{DD}$); panel C compares early treated units to late treated units during the late timing group's pre-period ($\hat{\beta}_{k\ell}^{DD,k}$); panel D compares late treated units to early treated units during the early timing group's post-period ($\hat{\beta}_{k\ell}^{DD,\ell}$). The treatment times mean that $\overline{D}_k = 0.67$ and $\overline{D}_\ell = 0.16$, so with equal group sizes, the decomposition weights on the 2x2 estimate from each panel are 0.365 for panel A, 0.222 for panel B, 0.278 for panel C, and 0.135 for panel D.

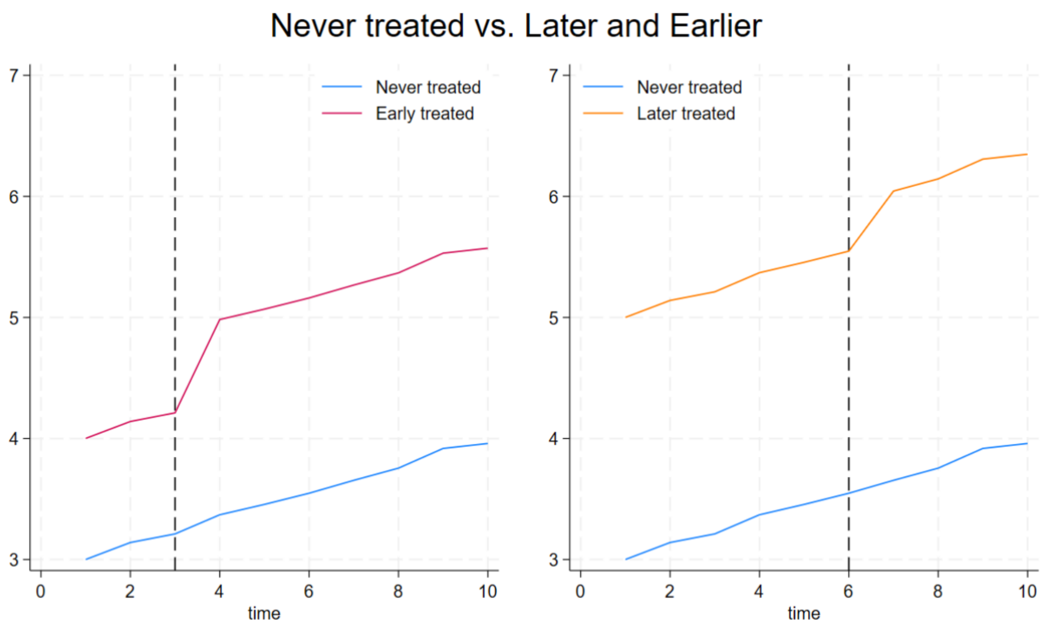
$$
\begin{aligned}
&\operatorname{plim}_{n\to\infty}\hat{\beta}\\
&= \text{Variance-weighted } ATT\\
&+ \text{Variance-weighted (un)common trends}\\
&+ \Delta ATT
\end{aligned}
$$

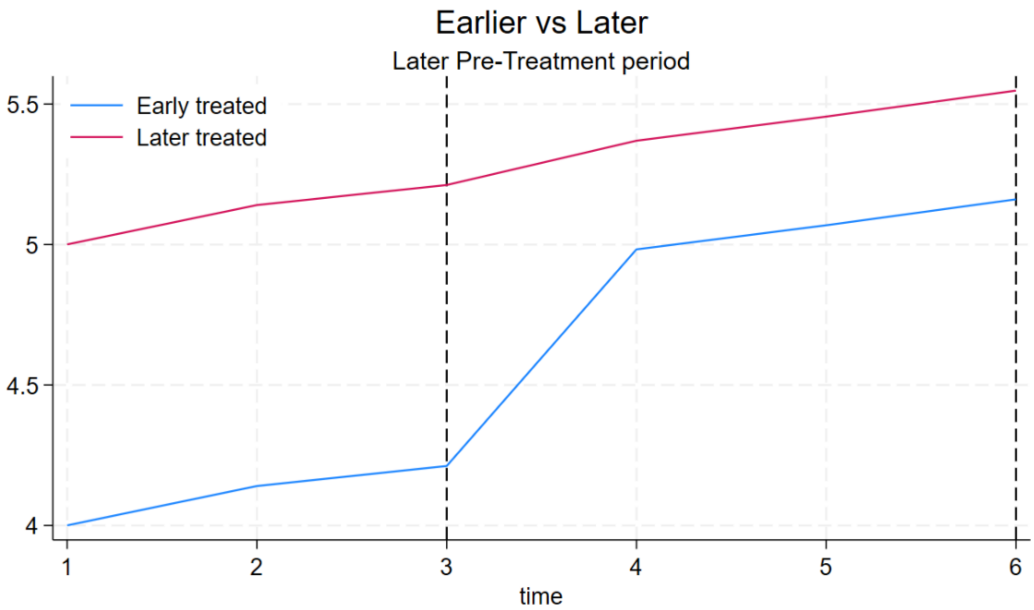
1.2 两个模拟结果

1.2.1 多期同质处理

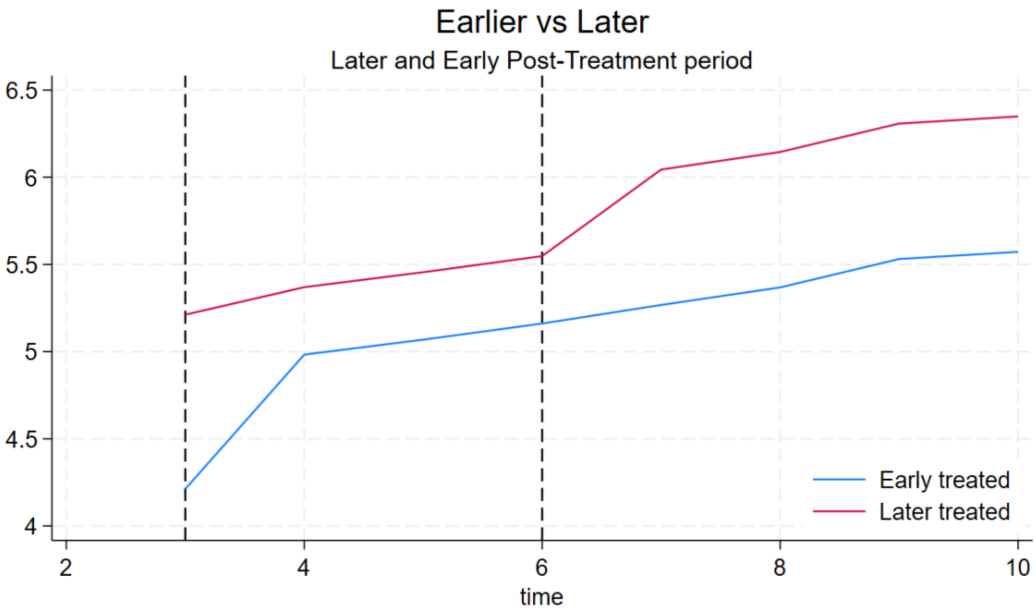
模拟设定：

- 10 期
- 部分个体在第 3 期接受处理
- 部分个体在地 6 期接受处理
- 其余个体一直不被处理


Never treated vs. Later and Earlier
Never treated
Early treated
Never treated
Later treated
time
time
7
6
5
4
3
0
2
4
6
8
10



Earlier vs Later
Later Pre-Treatment period
Early treated
Later treated
5.5
5
4.5
4
1
2
3
4
5
6
time

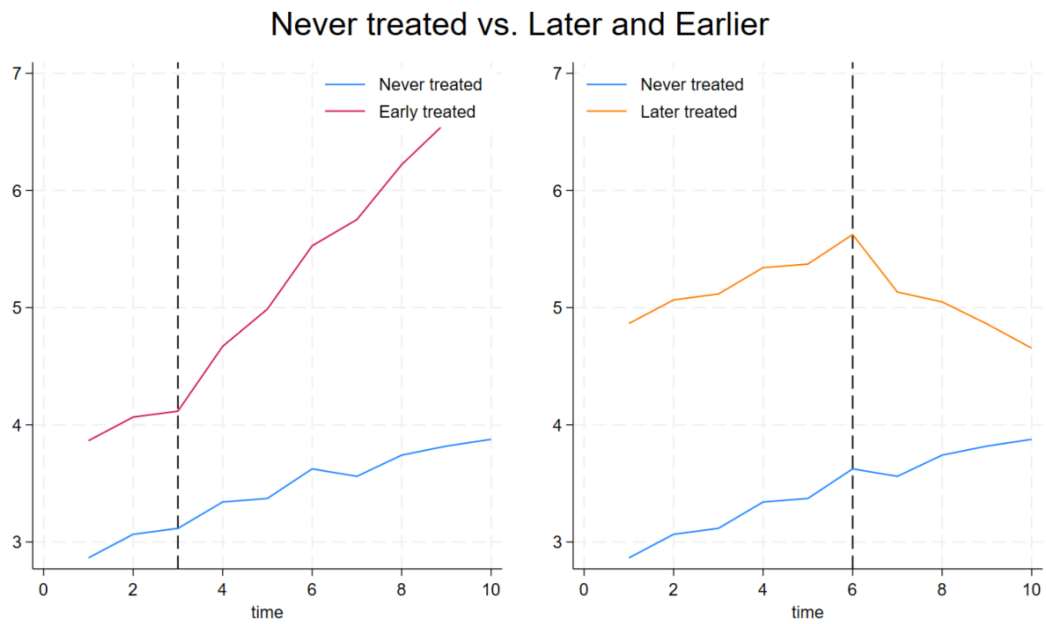


Earlier vs Later
Later and Early Post-Treatment period
6.5
6
5.5
5
4.5
4
2
4
6
8
10
time
Early treated
Later treated

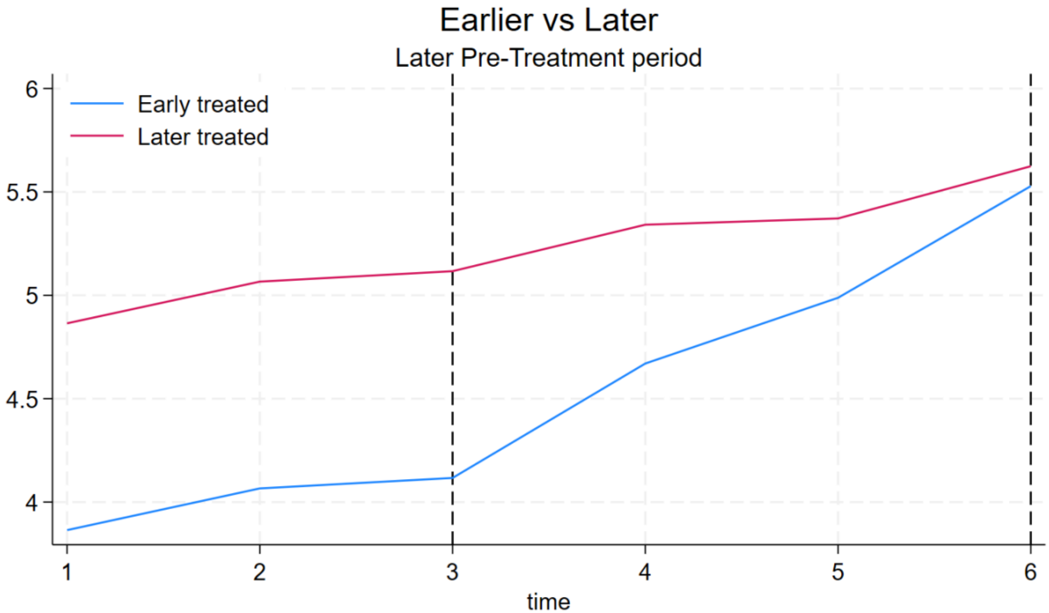
estat bdecomp

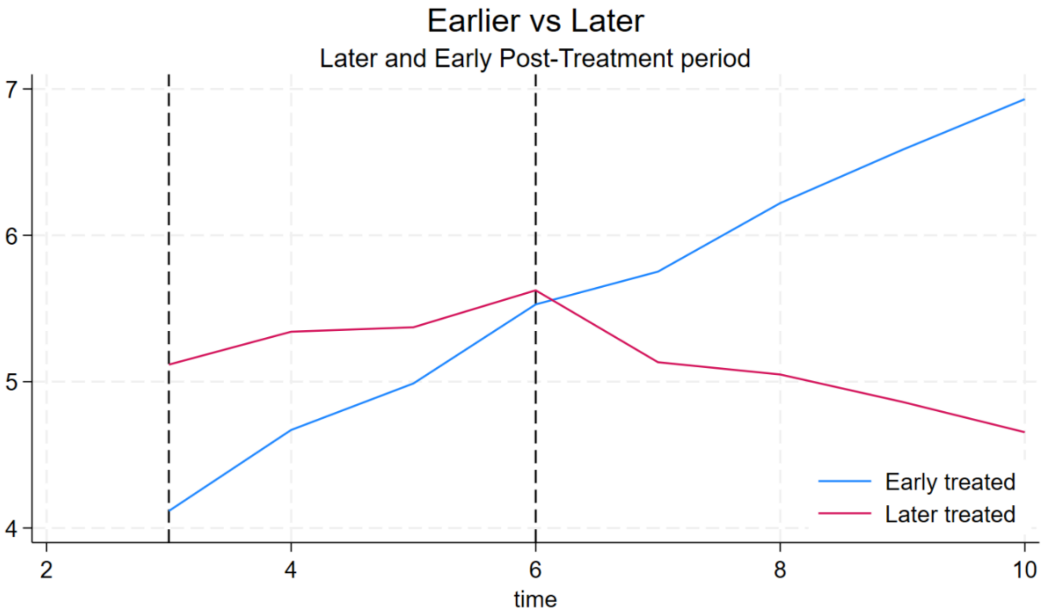
```
. estat bdecomp, summaryonly
DID treatment-effect decomposition
ATET = 4.041694                                    Number of obs     = 100,000
                                                   Number of groups  =  10,000
                                                   Number of cohorts =       3
ATET decomposition summary                ATET component                Weight
-------------------------------------------------------------------------------
Treated vs never treated                       4.0435917              0.865605
Treated earlier vs later                       4.0245287              0.057598
Treated later vs earlier                       4.0331799              0.076797
-------------------------------------------------------------------------------
Note: Number of cohorts includes never treated.
Note: The ATET reported by xtdidregress is a weighted average of the ATET
      components. If any component is substantially different from the ATET
      reported by xtdidregress and the weight is large, consider accounting
      for treatment-effect heterogeneity by using xthdidregress.
```

1.2.2 异质处理效应



Never treated vs. Later and Earlier
Never treated
Early treated
Never treated
Later treated
time
time


Earlier vs Later
Later Pre-Treatment period
Early treated
Later treated
6
5.5
5
4.5
4
1
2
3
4
5
6
time


Earlier vs Later
Later and Early Post-Treatment period
7
6
5
4
2
4
6
8
10
time
Early treated
Later treated

```
. estat bdecomp, summaryonly
DID treatment-effect decomposition
ATET = 1.389688                                   Number of obs     = 50,000
                                                  Number of groups  =  5,000
                                                  Number of cohorts =      3
ATET decomposition summary              ATET component                Weight
-----------------------------------------------------------------------------
Treated vs never treated                     1.9728088              0.860939
Treated earlier vs later                     4.3441312              0.059597
Treated later vs earlier                    -7.1439268              0.079463
-----------------------------------------------------------------------------
Note: Number of cohorts includes never treated.
Note: The ATET reported by xtdidregress is a weighted average of the ATET
      components. If any component is substantially different from the ATET
      reported by xtdidregress and the weight is large, consider accounting
      for treatment-effect heterogeneity by using xthdidregress.
```

```
. estat aggregation
Overall ATET                                          Number of obs = 50,000
                                   (Std. err. adjusted for 50 clusters in state)
```

y	ATET	Robust std. err.	z	P>|z|	[95% conf.	interval]
tr						
(1 vs 0)	3.654175	.962046	3.80	0.000	1.768599	5.53975

总结：forbidden comparisons 导致了不可直接加总的分组 ATT 被错误地加总到一起。

Goodman-Bacon (2021) shows that when the treatment is binary and the design is staggered, meaning that groups can switch in but not out of treatment, we have

$$\widehat{\beta}_{fe} = \sum_{g \neq g', t < t'} v_{g,g',t,t'} \mathrm{DID}_{g,g',t,t'},$$

where $DID_{g,g',t,t'}$ is a DID comparing the outcome evolution of two groups g and g' from a pre period t to a post period t', and where $v_{g,g',t,t'}$ are non-negative weights summing to one, with $v_{g,g',t,t'} > 0$ if and only if g switches treatment between t and t' while g' does not. [9] Some of the $DID_{g,g',t,t'}$ s in Equation (5) compare a group switching treatment from t to t' to a group untreated at both dates, while other $DID_{g,g',t,t'}$s compare a switching group to a group treated at both dates. The negative weights in (3) originate from this second type of DIDs.

- 对于多期处理的情况，传统的 TWFE 估计方法实际上是假设了同质性处理。
- 而事实上，很有可能每次的处理并不是同质的。或者说，更加保守的做法——对样本做异质性处理的假设——更应被提倡。
- 一般而言，这些估计是通过双向固定效应（TWFE）DiD 回归获得的，它们是许多不同的“2×2”DiD 的方差加权平均值，每个 2×2 都涉及在受处理组接受处理之前和之后的一个窗口内，比较受处理组和有效对照组之间的差异。在一些 2×2 中，已经接受处理的个体可以作为有效比较单元，其结果变化可能反映出从后续接受处理的单元中减去的处理效应。换句话说，这些回归引入了一个“糟糕的比较”问题，这与违反平行趋势假设类似但同样有问题。当处理效应随时间变化（“动态处理效应”）时，错位 DiD 处理效应估计值实际上可能获得与真正的 ATT 相反的符号，即使研究人员能够随机分配处理（从而满足平行趋势假设）。这些理论结果对应用研究人员具有深远的影响。

2 解决异质性处理效应的方法

- 根据组别、时期进行调整的平均处理效应
  - DeChaisemartin 和 d'Haultfeuille (2020) 提出的估计量 (did_multiplegt)
  - Sun 和 Abraham (2021) 提出的估计量 (eventstudyinteract)
  - Callaway 和 Sant'Anna (2021) 提出的估计量 (drdid, csdid)
- 插补估计量
  - Borusyak et al.(2021) 提出的估计量 (did_imputation)
  - Gardner(2021) 提出的估计量 (did2s)
- Solution 1: Callaway and Sant'Anna (2021)
  - Use valid comparison groups
  - Split the problem into 2 by 2 comparisons
  - Compute $ATET(g, t)$, where g is the cohort and t is the time
- Solution 2: Wooldridge (2001)
  - (Do not blame the messenger) Include heterogeneity in your regression
  - Add the adequate interactions with cohort and time

- Compute $ATET(g,t)$, where g is the cohort and t is the time
- Computing $ATET(g)$ or $ATET(T)$ is also possible

- 一些 survey: Roth et al. (2022), de Chaisemartin and D’Haultfoeuille

2.1 分析框架

$$Y_{it} = Y_{it}(0) + \sum_{g=2}^{T} \left[Y_{it}(g) - Y_{it}(0)\right] G_{ig}$$

- Y_{it} observed outcome
- $Y_{it}(0)$ potential outcome of not being treated
- G_{ig} is an indicator for treatment group
- g is the time at which a group of individuals is treated (cohort)。g 代表在第 g 期受到处理的组。
- $Y_{it}(g)$ potential outcome for cohort g。$Y_{it}(g) = ATET(g, t)$，第 g 组个体在第 t 期的处理效应。

其他假设：

- Treatment is staggered
- Parallel trends
- No anticipation
- Overlap

三种方法：

- Regression adjustment (RA)
- Inverse-probability weighting (IPW)
- Augmented inverse-probability weighting (AIPW)

2.2 Regression Adjustment

$$\text{ATET}(g,t) = E\left\{\frac{G_g}{E\left(G_g\right)}\left[Y_t - Y_{g-1} - m_{gt}(X)\right]\right\} \tag{1}$$

- $m_{gt}(X) = E\left(Y_t - Y_{g-1} \mid X, C = 1\right)$
  - 对于 never treated group ($C = 1$)，计算 t 期和 g 期之间的差
- $C = 1$ is the never treated group $\left(G_g = \infty\right)$
- $\left[Y_t - Y_{g-1} - m_{gt}(X)\right]$ 意思是对于每个 g 组的个体，计算他们的前后比较 (t 期和被处理前 $g-1$ 期的差异) 的之后，要减去 never treated 组在相同时期的变化。

$$\text{ATET}(g,t) = E\left\{\frac{G_g}{E\left(G_g\right)}\left[Y_t - Y_{g-1} - m_{gt}(X)\right]\right\}$$

- ATET(g,t) is calculated using two groups: g and $C = 1$, never treated
- Outcomes are computed for two points in time
- 2 by 2 idea
- This is done for all g and all t
- We could have other 2 by 2 comparisons, i.e, using the not yet treated Identification assumptions are the same but need to hold for each 2 by 2

2.3 IPW 逆概率加权

$$\mathrm{ATET}(g,t)=E\left\{\left(\frac{G_g}{E\left(G_g\right)}-\frac{\frac{p_g(X)}{1-p_g(X)}}{E\left[\frac{p_g(X)}{1-p_g(X)}\right]}\right)\left[Y_t-Y_{g-1}\right]\right\}$$

- $p_g(X)=P\left(G_g=1 \mid X, G_g+C=1\right)$, i.e., conditional on the sample we keep
- Steps are similar to before with the additional computation of $\widehat{p}_g(X)$ and the quotient $\frac{\frac{\hat{p}_g(X)}{1-\hat{p}_g(X)}}{\widehat{E}\left[\frac{\hat{\rho}_g(X)}{1-\hat{p}_g(X)}\right]}$

2.4 AIPW（double robust）

$$\text{ATET}(g,t)=E\left\{\left(\frac{G_g}{E\left(G_g\right)}-\frac{\frac{p_g(X_1)}{1-p_g(X_1)}}{E\left[\frac{p_g(X_1)}{1-p_g(X_1)}\right]}\right)\left[Y_t-Y_{g-1}-m_{gt}\left(X_2\right)\right]\right\}$$

- Notice $m_{gt}\left(X_2\right)$. Emphasizes we could have different covariates.
- AIPW is doubly robust. You may incorrectly specify at least one of $m_{gt}\left(X_2\right)$ or $p_g\left(X_1\right)$ and still recover $\text{ATET}(g,t)$
- 文献：
  Sant'Anna, Zhao, Doubly Robust Difference-in-Differences Estimators, JoE2020
  Callaway, Sant'Anna, Difference-in-Differences with multiple time periods, JoE202
  stata: csdid drdid
  drdid 可以得到任何可能的已处理和未处理组的 2×3DID 估计，相当于实现了 Sant'Anna and Zhao (2020)。
  csdid 实现了 Callaway and Sant'Anna (2021)，可以生产一个比较完整的表格。

3 命令

[CAUSAL] xthdidregress — Heterogeneous difference in differences for panel data
(View complete PDF manual entry)

Syntax

Two-way fixed effects

xthdidregress twfe (_ovar_ [_omvarlist_]**)** **(**_tvar_**)** [_if_] [_in_] [_weight_]**, group(**_groupvar_**)** [_options_]

Regression adjustment

xthdidregress ra (_ovar_ [_omvarlist_]**)** **(**_tvar_**)** [_if_] [_in_] [_weight_]**, group(**_groupvar_**)** [_options_]

Inverse-probability weighting

xthdidregress ipw (_ovar_**)** **(**_tvar_ [_tmvarlist_]**)** [_if_] [_in_] [_weight_]**, group(**_groupvar_**)** [_options_]

Augmented inverse-probability weighting

xthdidregress aipw (_ovar_ [_omvarlist_]**)** **(**_tvar_ [_tmvarlist_]**)** [_if_] [_in_] [_weight_]**, group(**_groupvar_**)** [_options_]

ovar is a continuous outcome of interest.
omvarlist specifies the covariates in the outcome model and may contain factor variables; see fvvarlist.
tvar must be a binary variable indicating observations subject to treatment.
tmvarlist specifies the covariates in the treatment model and may contain factor variables; see fvvarlist.
groupvar is a categorical variable that indicates the group level at which the treatment occurs.

3.1 Wooldridge

$$Y_{it} = \eta + \sum_{g=q}^{T} G_{ig}\alpha_g + \sum_{s=q}^{T} f_s\alpha_s + \sum_{g=q}^{T}\sum_{s=g}^{T} D_{it}G_{ig}f_s\tau_{gs}$$

- q is the first treatment time and $q \dots T$ the post period
- f_s is 1 if the time period is s and 0 otherwise
- We have group and time effects α_g and α_s
- Heterogeneity is captured by interacting group and time effects

$$\tau_{gs} \equiv \text{ATET}(g,t)$$

- We use all our data and do not partition them
- If I have covariates, they enter fully interacted
- Extended two-way fixed effects

3.2 De Chaisemartin 和 d'Haultfœuille (2020)

did_multiplegt

de Chaisemartin 和 D' Haultfœuille (2020) 提出了 DIDM 估计量，这是一个加权平均值，涵盖了两种类型的（DID）：

- a DID comparing the $t-1$ to t outcome evolution of groups going from untreated to treated from $t-1$ to t, the "switchers in", and of groups untreated at both dates.
- a DID comparing the $t-1$ to t outcome evolution of groups treated at both dates, and of groups going from treated to untreated from $t-1$ to t, the "switchers out".

这个方法思想是，将政策处理从无到有、从有到无两个正负方向的效果进行加权平均，而将两期的处理不变 (如：两期都是 1 或两期都是 0) 的样本作为控制组。对比多期 DID 模型，仅考虑了政策从无到有的情况。

$$W_{TC} = \sum_{t=1}^{\bar{t}} \left(\frac{N_{1,0,t}}{N_S} DID_{+,t} + \frac{N_{0,1,t}}{N_S} DID_{-,t} \right)$$

$DID_{+,t}$ 指的是政策从 0 到 1 的平均结果减去稳定组 (两期都是 0) 的平均结果，$DID_{-,t}$

指的是稳定组 (两期都是 1) 的平均结果减去政策从 0 到 1 的平均结果。

$$DID_{+,t} = \sum_{g:D_{g,t}=1,D_{g,t-1}=0} \frac{N_{g,t}}{N_{1,0,t}} \left(E\left(Y_{g,t}\right) - E\left(Y_{g,t-1}\right)\right) - \sum_{g:D_{g,t}=D_{g,t-1}=0} \frac{N_{g,t}}{N_{0,0,t}} \left(E\left(Y_{g,t}\right) - E\left(Y_{g,t}\right.\right.$$

$$DID_{-,t} = \sum_{g:D_{g,t}=D_{g,t-1}=1} \frac{N_{g,t}}{N_{1,1,t}} \left(E\left(Y_{g,t}\right) - E\left(Y_{g,t-1}\right)\right) - \sum_{g:D_{g,t}=0,D_{g,t-1}=1} \frac{N_{g,t}}{N_{0,1,t}} \left(E\left(Y_{g,t}\right) - E\left(Y_{g,t}\right.\right.$$

共同趋势假定: 稳定组组内共同趋势，即对于在 t 和 $t-1$ 都无政策的组有共同趋势，而在 t 和 $t-1$ 都有政策的组有共同趋势。这个设定比上面的设定要宽松，固定效应的模型要求不同组的共同趋势也相同。

3.3 Borusyak et al.(2024)

did_imputation

- 先试用未处理组的数据跑一个固定效应回归。
- 然后利用这个固定效应模型预测处理组假如未被处理时的反事实 outcome,
- 处理组的真是 outcome 减反事实 outcome，即为处理效应。
- 最后对这些处理效应进行加权求和。

Callaway 和 Sant’Anna (2020) 以及 Sun 和 Abraham (2020) 都用 $t-1$ 期的结果作为基准结果，而 Borusyak et al. (2024) 则使用 1 到 $t-1$ 期的平均结果作为基准.

- Goodman-Bacon（2021）表明，“静态”交错 DiD TWFE 处理效应估计（方程（2）的 δ^{DD}）是“所有可能的两组/两期 DiD 估计量在数据中的加权平均值”。例如，当样本期间有三组时（从最早的期间 $-t_0-$ 到数据的最后一个期间 $-T$），TWFE 估计量包括四个可能的 2×2，即一个从未接受过处理的组（表示为 U），一个较早接受处理的组（表示为 k），该组在时间 t_k^* 接受处理，以及一个较晚接受处理的组（表示为 l），该组在时间 t_l^* 接受处理。
- 可能的前两个 2×2 DiD 比较涉及一个处理组（早期或后期接受处理的企业之一）和整个样本窗口（从 t_0 到 T）的未接受处理的组（作为对照组）。另外两个可能的 2×2 涉及不同的处理组之间的比较。其中一个“仅时间”的 2×2 比较早期接受处理的企业与后期接受处理的企业（作为对照组）在从 t_0 到 t_l^* 的时间窗口内的情况（即在这种情况下，早期接受处理的企业接受处理，而后期接受处理的企业尚未接受处理）。另一个“仅时间”的 2×2 比较后期接受处理的企业与早期接受处理的企业在从 t_k^* 到 T 的时间窗口内的情况（即在这种情况下，后期接受处理的企业接受处理，而早期接受处理的企业已经接受了处理）。在后一个比较中，早期接受处理的企业被用作对照组，用于比较后期接受处理的结果。