

# 当代 DID 方法：新 DID 与异质性趋势

## 从 Zhang and Xu(2025,JEBO) 开始

许文立, wxu@cityu.edu.mo

澳门城市大学金融学院

2025 年 7 月 7 日



澳門城市大學  
Universidade da Cidade de Macau  
City University of Macau

- ① Zhang and Xu(2025,JEBO): TV 暴露对孩子成绩的影响
- ② Wooldridge(2025) 的灵活、统一框架
- ③ 新 DID 估计量
- ④ 扩展
- ⑤ 参考文献

- ① Zhang and Xu(2025,JEBO): TV 暴露对孩子成绩的影响
- ② Wooldridge(2025) 的灵活、统一框架
- ③ 新 DID 估计量
- ④ 扩展
- ⑤ 参考文献

## 方法论的应用

- 首次以 Wooldridge(2021) 的 ETWFE 为基准方程声明:

$$\begin{aligned}
Y_{i,t} = & \alpha + \sum_{g \in G} \sum_{t=t_0}^{g-1} \theta_{g,t}^{pre} D_{i,g,t} + \sum_{g \in G} \sum_{t=g}^T \theta_{g,t}^{post} D_{i,g,t} \\
& + \sum_{g \in G} \sum_{t=t_0}^{g-1} D_{i,g,t} \tilde{x}'_i \beta_{g,t}^{pre} + \sum_{g \in G} \sum_{t=g}^T D_{i,g,t} \tilde{x}'_i \beta_{g,t}^{post} \\
& + x'_i \beta + \sum_{t=t_0}^T D_{i,t} x'_i \beta_t + \sum_{g \in G} D_{i,g} x'_i \beta_g \xi_i + \xi_t + \varepsilon_{i,t}
\end{aligned}$$

# 方法论的应用

- Wooldridge(2021) 的 ETWFE 是课本中的 ols 估计，透明，易于理解；
- 非常灵活：可以应对非常多的混淆因子带来的内生性问题；
- 应对交叠处理可能存在的异质性处理效应偏误
- 容易扩展：混合截面，队列，连续处理，非线性 DID 等等
- 容易与 honestdid 结合检验平行趋势的敏感性

- ◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ▶ ↺ 🔍 ↻

- A set of navigation icons typically found in Beamer presentations, including symbols for back, forward, search, and other slide controls.

- ◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ▶ ↺ 🔍 ↻



- ◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ▶ ↺ 🔍 ↻

- T 期，第一期没有处理发生
- 第一个处理的个体发生在  $t = q \leq T$
- 处理发生，就一直留在处理组
- 有从未处理的纯控制组（可以放松）
- 对于每个处理类别（cohort） $g \in \{q, \dots, T\}$ ,  $Y_t(g)$  是 cohort  $g$  的结果， $Y_t(\infty)$  表示从未处理个体的结果
- 主要关注的处理效应：

$$TE_{gt} = Y_t(g) - Y_t(\infty)$$

- 互斥的、所有的处理类别虚拟变量:

$$D_g = 1, \text{ if } i \in g$$

$$D_\infty = 1 - (D_q + D_{q+1} + \dots + D_T)$$

其中,  $D_{\infty} = 1$  意味着个体  $i$  从未处理



- ◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡

① Zhang and Xu(2025,JEBO): TV 暴露对孩子成绩的影响

② Wooldridge(2025) 的灵活、统一框架

③ 新 DID 估计量

插补法 imputation 估计量

事件研究估计量

处理类别的异质性线性趋势

④ 扩展

⑤ 参考文献

imputation

- BJS(2024,RES) 和 Gardner(2021) 用两步估计法:
  - 1、 用还未处理/从未处理的观测值来识别潜在结果:

$$Y_{jt}(0) = \alpha + \beta_g + \gamma_t + E(x|D_g = 1)(\kappa + \xi_g + \pi_t)$$

- 2、用实际观测结果减去第一步估计的潜在结果:

$$ATT_{it} = Y_{it} - Y_{it}(0)$$

- 为了避免对函数形式敏感，条件于处理类别，协变量分布要有足够的 overlap

## 15 / 50

## imputation 与 pols 的等价性

- 如果我们感兴趣的是暴露时长  $W_{it}$  的处理效应:

$$\begin{aligned}
E(Y_{it}|\mathbf{W}_i, \mathbf{D}_i, \mathbf{X}_i) = & \alpha + \sum_{g=q}^T \beta_g D_{ig} + \mathbf{X}_i \kappa + \sum_{g=q}^T (D_{ig} \cdot \mathbf{X}_{ig}) \xi_g \\
& + \sum_{s=2}^T \gamma_s f_{st} + \sum_{s=2}^T (f_{st} \bullet \mathbf{X}_i) \pi_s \\
& + \sum_{g=q}^T \sum_{s=g}^T \tau_{gs} (W_{it} \bullet D_{ig} \bullet f_{st}) \\
& + \sum_{g=q}^T \sum_{s=g}^T (W_{it} \cdot D_{ig} \cdot f_{st} \cdot \dot{\mathbf{X}}_{ig}) \rho_{gs}
\end{aligned}$$

其中,  $\dot{\mathbf{X}}_g \equiv \mathbf{X} - E(\mathbf{X}|D_g = 1)$  表示去除了处理类别均值的协变量



# imputation 与 pols 的等价性

- 在没有协变量的情况下，上面两类 pols 估计等价于 Gardner(2021) 的 cohort 虚拟变量 imputation
- 在时间不变协变量和平衡面板，pols 与 BJS(2024,RES) 等价
  - BJS 用个体固定效应替换 cohort dummy
- 用 cohort dummy 可以极大提高计算速度
- BJS 估计量的推断更加保守
- cohort imputation = pols = RE=FE= BJS(FE imputation)
- FE 舍弃  $D_{ig}, x_i, D_{ig} * x_i$
- $x_i$  可以包含比 cohort 层级低的虚拟变量：如果  $i$  是县级层面，处理在省级变化，可以包含  $prov \times year$  固定效应；也允许处理前趋势在省级变动

## pols 的透明性和灵活性：允许排除许多异质性因素干扰

- 明确控制多种内生选择进入处理的机制： $D_{ig}$  和  $D_{ig} * \dot{X}_i$
- 允许协变量的效应随着 cohort 和时间发生变化：  
 $D_{ig} * \dot{X}_i, fs_t * X_i$
- 允许处理效应随可观测的协变量变化： $D_g * fs_t * \dot{X}_i$
- ATT 不受限制：允许异质性处理效应

① Zhang and Xu(2025,JEBO): TV 暴露对孩子成绩的影响

② Wooldridge(2025) 的灵活、统一框架

③ 新 DID 估计量

插补法 imputation 估计量

事件研究估计量

处理类别的异质性线性趋势

④ 扩展

⑤ 参考文献

## SA(2021,JoE) 估计量

- 前面的新 DID 估计量都没有处理前系数，只有处理后 (only lags) 系数
- SA(2021) 提出一种饱和处理变量的 TWFE 方式：

$$D_{ig} * fs_t, \text{ for all } s \neq g - 1$$

其中， $g - 1$  是基期

- Wooldridge(2024) 指出，SA(2021) 估计量等价于包括 cohort dummies  $D_{ig}$  和时间虚拟变量
- 完全灵活的 ES 回归：也包括处理前交互项  
 $D_{ig} * X_i, fs_t * X_i, D_{ig} * fs_t * X_i, s = 1, \dots, g - 2$
- 相比于 SA(2021)，pols 事件研究更加灵活的控制协变量

# 完全灵活的事件研究

- 上述完全灵活的事件研究 pols:

$Y_{it}$  on  $D_{iq} \bullet f1_t, \dots, D_{iq} \bullet f(q-2)_t, D_{iq} \bullet fq_t, \dots, D_{iq} \bullet fT_t,$

...

$D_{iT} \bullet f1_t, \dots, D_{iT} \bullet f(T-2)_t, D_{iT} \bullet fT_t,$

$D_{iq} \bullet f1_t \bullet \dot{X}_{iq}, \dots, D_{iq} \bullet f(q-2)_t \bullet \dot{X}_{iq}, D_{iq} \bullet fq_t \bullet \dot{X}_{iq}, \dots, D_{iq}$

$D_{iT} \bullet f1_t \bullet \dot{X}_{iT}, \dots, D_{iT} \bullet f(T-2)_t \bullet \dot{X}_{iT}, D_{iT} \bullet fT_t \bullet \dot{X}_{iT},$

$1, D_{iq}, \dots, D_{iT}, \mathbf{X}_i, D_{iq} \bullet \mathbf{X}_i, \dots, D_{iT} \bullet \mathbf{X}_i,$

$f2_t, \dots, fT_t, f2_t \bullet \mathbf{X}_i, \dots, fT_t \bullet \mathbf{X}_i$

- 检验处理前  $D_{ig} \bullet fs_t$  的系数, 来进行处理前趋势检验,  
 $s = 1, \dots, g-2$

- $g = q - 1$  是基期
- 等价于 RE 估计
- 等价于 TWFE, 舍弃了时间不变的变量
- 对于处理前系数和处理后系数, 用  $2 \times 2$  DID( $g-1, s$ ):  
$$Y_{it} = \alpha + \beta D_{ig} + \mathbf{X}_i \gamma + (D_{ig} \cdot \dot{\mathbf{X}}_{ig}) \delta + \gamma_s f_{st} + (f_{st} \cdot \mathbf{X}_i) \pi_s$$
$$+ \tau_{gs} (D_{ig} \bullet f_{st}) + (D_{ig} \bullet f_{st} \bullet \dot{\mathbf{X}}_{ig}) \mathbf{p}_{gs} + U_{it}, t \in (g-1, s)$$
- 上式等同于 CS(2021) 的回归调整法, 即对 “长差分” 分别跑 RA:

$Y_{is} - Y_{i,g-1}$  on  $1, D_{ig}, \dot{\mathbf{X}}_{ig}, D_{ig} \cdot \dot{\mathbf{X}}_{ig}$  using  $D_{ig} = 1$  or  $D_{i\infty} = 1$

- 在强的正序列相关下, ES 估计量比 only lags 估计量更有效率

# imputation 版的 ES 估计

- 与 imputation 版本的 ES 估计也是等价的：
- 第一步：包含所有处理前 dummies，及其与协变量  $x$  的交乘项
- 回归声明：

$$\begin{aligned}
 E[Y_t(\infty)|\mathbf{D}, \mathbf{X}] = & \alpha + \sum_{g=q}^T \beta_g D_g + \mathbf{X}\kappa + \sum_{g=q}^T (D_g \cdot \mathbf{X}) \xi_g \\
 & + \sum_{s=2}^T \gamma_s f_{st} + \sum_{s=2}^T (f_{st} \cdot \mathbf{X}) \pi_s \\
 & + \sum_{g=q}^T \sum_{s=1}^{g-2} \theta_{gs} (D_g \bullet f_{st}) + \sum_{g=q}^T \sum_{s=1}^{g-2} (D_g \bullet f_{st} \bullet \dot{\mathbf{X}}_g) \lambda_{gs}
 \end{aligned}$$

- 注意：技术上看，平行趋势的 ES 版本是更弱的假设，但是当 PT 在处理后不成立时，估计量的偏误更大
- 因此，如果研究者想用基于灵活回归的估计量，实际上就是在“only lags”和“ES”之间进行选择
- only lags 用所有 CPT 的信息来保证处理效应
- ES 可以进行处理前趋势检验
- 所有的新 DID 都是在偏误和效率之间进行 trade off，没有孰优孰劣



① Zhang and Xu(2025,JEBO): TV 暴露对孩子成绩的影响

② Wooldridge(2025) 的灵活、统一框架

③ 新 DID 估计量

插补法 imputation 估计量

事件研究估计量

处理类别的异质性线性趋势

④ 扩展

⑤ 参考文献

- 允许处理类别的异质性线性趋势的灵活回归:

$$E[Y_t(\infty) | \mathbf{D}, \mathbf{X}] = \alpha + \sum_{g=q}^T \beta_g D_g + \mathbf{X}\boldsymbol{\kappa} + \sum_{g=q}^T (D_g \cdot \mathbf{X}) \boldsymbol{\xi}_g \\ + \sum_{s=2}^T \gamma_s f_s + \sum_{s=2}^T (f_{s_t} \cdot \mathbf{X}) \boldsymbol{\pi}_s + \sum_{g=q}^T \psi_g (D_g \cdot t)$$

- Wooldridge(2024) 指出, 有足够的处理前时期, 直接将  $D_{ig} * t$  引入 pols/TWFE
- $\sum_{s=2}^T \gamma_s f_s$  运行未处理组有不受限制的趋势,  
 $\sum_{s=2}^T (f_{s_t} \cdot \mathbf{X}) \boldsymbol{\pi}_s$  允许差异化的协变量趋势,  
 $\sum_{g=q}^T \psi_g (D_g \cdot t)$  则允许处理类别差异化的线性趋势

- 为了允许处理类别差异化趋势，每个类别处理前至少要有两期
- 引入异质性线性趋势会与处理指标  $w_{it} * D_{ig} * f_{st} = D_{ig} * f_{st}$  共线
- 有两个及以上处理前时期， $D_{ig} * f_{st}$  和  $D_{ig} * t$  就不完美相关，仍正相关
- 共线性会提高标准误，但是会被降低的余值方差（如果趋势确实有助于解释结果）补偿回来
- 此外，包括这个趋势项也可以降低非序列相关异方差
- 通常来说，可以采用更加灵活的回归，即引入  $D_{ig} * t * \dot{X}_i$ ，并不影响 ATTs 的估计
- 在多个处理前时期下，研究者还可以引入高阶时间趋势与  $D_{ig}$  和  $D_{ig} * \dot{X}_i$  的交乘项，但是这可能会降低 ATTs 的精度

## ① Zhang and Xu(2025,JEBO): TV 暴露对孩子成绩的影响

## ② Wooldridge(2025) 的灵活、统一框架

## ③ 新 DID 估计量

## ④ 扩展

非线性模型

所有个体最终都被处理

交叠进入-退出

时变控制变量

连续/多值处理

非平衡面板

重复截面/队列

DDD

非随机抽样/聚类配置

## ① Zhang and Xu(2025,JEBO): TV 暴露对孩子成绩的影响

## ② Wooldridge(2025) 的灵活、统一框架

## ③ 新 DID 估计量

## ④ 扩展

非线性模型

所有个体最终都被处理

交叠进入-退出

时变控制变量

连续/多值处理

非平衡面板

重复截面/队列

DDD

非随机抽样/聚类配置

## 30 / 50



## 所有个体最终都被处理

- 如果所有的个体最终都接受处理，那么，
  - 不能估计最后处理组的 ATT
  - Cohort T 变成了对照组
  - 在 T 之前，ATTs 用从未处理作为对照组
  - 在 T 期，更早处理的 cohort 必然与 cohort T 比较



- ◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ↺ 🔍 ↻



## 非线性模型

所有个体最终都被处理

## 交叠进入-退出

## 时变控制变量

## 连续/多值处理

## 非平衡面板

重复截面/队列

DDD

## 非随机抽样/聚类配置



## 非线性模型

所有个体最终都被处理

## 交叠进入-退出

## 时变控制变量

## 连续/多值处理

## 非平衡面板

重复截面/队列

DDD

## 非随机抽样/聚类配置

许文立, [wlxu@cityu.edu.mo](mailto:wlxu@cityu.edu.mo)

澳門城市大學金融學院

## 连续处理

- 用连续处理变量替换  $W_{it}$
- 通过空间权重矩阵，允许空间溢出效应









## 重复截面/队列

- POLS 和 ETWFE 很容易扩展到重复截面或队列 DID
- 参考 Deb et al.(2025,NBER),Zhang and Xu(2025,JEBO)

## 非线性模型

所有个体最终都被处理

## 交叠进入-退出

## 时变控制变量

## 连续/多值处理

## 非平衡面板

重复截面/队列

DDD

## 非随机抽样/聚类配置

## DDD

- 时间维度上的 DDD，例如，Strezhnev(2024), Wooldridge (2024)
- 截面维度的 DDD，例如，Xu(2025, Work in Progress)



# 聚类

- 所有的 pooled 估计都允许调查权重和聚类抽样
- 所有的 pooled 估计方法都允许比个体更高层级的聚类

- ① Zhang and Xu(2025,JEBO): TV 暴露对孩子成绩的影响
- ② Wooldridge(2025) 的灵活、统一框架
- ③ 新 DID 估计量
- ④ 扩展
- ⑤ 参考文献



- 1 Lin Zhang, Wenli Xu, Television exposure in early childhood and subsequent cognitive outcomes: Evidence from rural China, Journal of Economic Behavior & Organization, Volume 230, 2025, 106878

*Thanks!*