

- 1 因果推断中的内生性
- 2 DID 中的协变量
- 3 DID 中的机器学习
- 4 内生的处理配置
- 5 DID-IV 与 IV 型 DID
- 6 参考文献

① 因果推断中的内生性

② DID 中的协变量

③ DID 中的机器学习

④ 内生的处理配置

⑤ DID-IV 与 IV 型 DID

⑥ 参考文献

因果推断回顾

- 回忆：线性回归方程的斜率系数 $\beta = \frac{\partial Y}{\partial T}$ 测度了处理变量 T 对结果变量 Y 的因果效应

因果推断回顾

- 回忆：线性回归方程的斜率系数 $\beta = \frac{\partial Y}{\partial T}$ 测度了处理变量 T 对结果变量 Y 的因果效应
- β 的无偏估计要求“外生性”：

$$\text{cov}(T, u) = 0$$

当随机扰动项 u 中包含遗落变量 W ，外生性就不满足，“内生性”问题就出现了

因果推断回顾

- 回忆：线性回归方程的斜率系数 $\beta = \frac{\partial Y}{\partial T}$ 测度了处理变量 T 对结果变量 Y 的因果效应
- β 的无偏估计要求“外生性”：

$$\text{cov}(T, u) = 0$$

当随机扰动项 u 中包含遗落变量 W ，外生性就不满足，“内生性”问题就出现了

- 计量经济学从以下几个方面来尝试解决内生性问题：

因果推断回顾

- 回忆：线性回归方程的斜率系数 $\beta = \frac{\partial Y}{\partial T}$ 测度了处理变量 T 对结果变量 Y 的因果效应
- β 的无偏估计要求“外生性”：

$$\text{cov}(T, u) = 0$$

当随机扰动项 u 中包含遗落变量 W ，外生性就不满足，“内生性”问题就出现了

- 计量经济学从以下几个方面来尝试解决内生性问题：
 - a 做实验，在实验中，我们可以控制所有其它因素为常数（即 $u = c$ ），从而使得 $\text{cov}(T, c) = 0$

因果推断回顾

- 回忆：线性回归方程的斜率系数 $\beta = \frac{\partial Y}{\partial T}$ 测度了处理变量 T 对结果变量 Y 的因果效应
- β 的无偏估计要求“外生性”：

$$\text{cov}(T, u) = 0$$

当随机扰动项 u 中包含遗落变量 W ，外生性就不满足，“内生性”问题就出现了

- 计量经济学从以下几个方面来尝试解决内生性问题：
 - 做实验，在实验中，我们可以控制所有其它因素为常数（即 $u = c$ ），从而使得 $\text{cov}(T, c) = 0$
 - 在 RCT 中，随机配置处理 T ($T = \text{rand}$)，从而使得 $\text{cov}(\text{rand}, c) = 0$

因果推断回顾

- 回忆：线性回归方程的斜率系数 $\beta = \frac{\partial Y}{\partial T}$ 测度了处理变量 T 对结果变量 Y 的因果效应
- β 的无偏估计要求“外生性”：

$$\text{cov}(T, u) = 0$$

当随机扰动项 u 中包含遗落变量 W ，外生性就不满足，“内生性”问题就出现了

- 计量经济学从以下几个方面来尝试解决内生性问题：
 - 做实验，在实验中，我们可以控制所有其它因素为常数（即 $u = c$ ），从而使得 $\text{cov}(T, c) = 0$
 - 在 RCT 中，随机配置处理 T ($T = \text{rand}$)，从而使得 $\text{cov}(\text{rand}, c) = 0$
 - 从 u 中把遗落遗漏变量 W “拿”出来，例如，回归调整或者匹配，控制住 W ，比较相同 W 下的处理组与控制组结果差异

因果推断回顾

- 回忆：线性回归方程的斜率系数 $\beta = \frac{\partial Y}{\partial T}$ 测度了处理变量 T 对结果变量 Y 的因果效应
- β 的无偏估计要求“外生性”：

$$\text{cov}(T, u) = 0$$

当随机扰动项 u 中包含遗落变量 W ，外生性就不满足，“内生性”问题就出现了

- 计量经济学从以下几个方面来尝试解决内生性问题：
 - 做实验，在实验中，我们可以控制所有其它因素为常数（即 $u = c$ ），从而使得 $\text{cov}(T, c) = 0$
 - 在 RCT 中，随机配置处理 T ($T = \text{rand}$)，从而使得 $\text{cov}(\text{rand}, c) = 0$
 - 从 u 中把遗落遗漏变量 W “拿”出来，例如，回归调整或者匹配，控制住 W ，比较相同 W 下的处理组与控制组结果差异
 - 准自然实验方法，IV、RD、panel data、SC 和 DID 等

1 因果推断中的内生性

2 DID 中的协变量

条件平行趋势 PTA

重叠 (overlap) 假设

带协变量的 TWFE

带协变量的 TWFE 为什么有偏误

解决方法

3 DID 中的机器学习

4 内生的处理配置

5 DID-IV 与 IV 型 DID

① 因果推断中的内生性

② DID 中的协变量

条件平行趋势 PTA

重叠 (overlap) 假设

带协变量的 TWFE

带协变量的 TWFE 为什么有偏误

解决方法

③ DID 中的机器学习

④ 内生的处理配置

⑤ DID-IV 与 IV 型 DID

- 在大多数应用中，研究者通常使用下列包含时变协变量 X_{it} 的 TWFE 方程：

$$Y_{it} = \theta_t + \eta_i + \alpha D_{it} + X'_{it}\beta + v_{it}$$

- 在大多数应用中，研究者通常使用下列包含时变协变量 X_{it} 的 TWFE 方程：

$$Y_{it} = \theta_t + \eta_i + \alpha D_{it} + X'_{it}\beta + v_{it}$$

- Caetano et al.(2022)、Caetano and Callaway(2024) 指出上述 TWFE 形式可能存在许多潜在的问题：

- 在大多数应用中，研究者通常使用下列包含时变协变量 X_{it} 的 TWFE 方程：

$$Y_{it} = \theta_t + \eta_i + \alpha D_{it} + X'_{it}\beta + v_{it}$$

- Caetano et al.(2022)、Caetano and Callaway(2024) 指出上述 TWFE 形式可能存在许多潜在的问题：
 - 异质性处理效应带来的偏误，实践常使用新 DID 估计量，例如，Zhang and Xu(2025, JEBO)

- 在大多数应用中，研究者通常使用下列包含时变协变量 X_{it} 的 TWFE 方程：

$$Y_{it} = \theta_t + \eta_i + \alpha D_{it} + X'_{it}\beta + v_{it}$$

- Caetano et al.(2022)、Caetano and Callaway(2024) 指出上述 TWFE 形式可能存在许多潜在的问题：
 - 异质性处理效应带来的偏误，实践常使用新 DID 估计量，例如，Zhang and Xu(2025, JEBO)
 - 如果处理变量 D_{it} 会影响时变协变量 X_{it} ，就会带来“bad control”。实践中，常常不包含这类“坏控制变量”，但是 Callaway(2022) 认为，不包含这类时变协变量也不是一个好解决方法。例如，就业培训对劳动者收入的影响，控制职业变量和不控制都会带来问题

- 在大多数应用中，研究者通常使用下列包含时变协变量 X_{it} 的 TWFE 方程：

$$Y_{it} = \theta_t + \eta_i + \alpha D_{it} + X'_{it}\beta + v_{it}$$

- Caetano et al.(2022)、Caetano and Callaway(2024) 指出上述 TWFE 形式可能存在许多潜在的问题：
 - a 异质性处理效应带来的偏误，实践常使用新 DID 估计量，例如，Zhang and Xu(2025, JEBO)
 - b 如果处理变量 D_{it} 会影响时变协变量 X_{it} ，就会带来“bad control”。实践中，常常不包含这类“坏控制变量”，但是 Callaway(2022) 认为，不包含这类时变协变量也不是一个好解决方法。例如，就业培训对劳动者收入的影响，控制职业变量和不控制都会带来问题
 - c 上述 TWFE 对协变量效应的函数形式非常敏感

- 在大多数应用中，研究者通常使用下列包含时变协变量 X_{it} 的 TWFE 方程：

$$Y_{it} = \theta_t + \eta_i + \alpha D_{it} + X'_{it}\beta + v_{it}$$

- Caetano et al.(2022)、Caetano and Callaway(2024) 指出上述 TWFE 形式可能存在许多潜在的问题：
 - a 异质性处理效应带来的偏误，实践常使用新 DID 估计量，例如，Zhang and Xu(2025, JEBO)
 - b 如果处理变量 D_{it} 会影响时变协变量 X_{it} ，就会带来“bad control”。实践中，常常不包含这类“坏控制变量”，但是 Callaway(2022) 认为，不包含这类时变协变量也不是一个好解决方法。例如，就业培训对劳动者收入的影响，控制职业变量和不控制都会带来问题
 - c 上述 TWFE 对协变量效应的函数形式非常敏感
 - d 实践中，还有一种常见的方式，就是控制时间不变的协变量

- 在大多数应用中，研究者通常使用下列包含时变协变量 X_{it} 的 TWFE 方程：

$$Y_{it} = \theta_t + \eta_i + \alpha D_{it} + X'_{it}\beta + v_{it}$$

- Caetano et al.(2022)、Caetano and Callaway(2024) 指出上述 TWFE 形式可能存在许多潜在的问题：
 - 异质性处理效应带来的偏误，实践常使用新 DID 估计量，例如，Zhang and Xu(2025, JEBO)
 - 如果处理变量 D_{it} 会影响时变协变量 X_{it} ，就会带来“bad control”。实践中，常常不包含这类“坏控制变量”，但是 Callaway(2022) 认为，不包含这类时变协变量也不是一个好解决方法。例如，就业培训对劳动者收入的影响，控制职业变量和不控制都会带来问题
 - 上述 TWFE 对协变量效应的函数形式非常敏感
 - 实践中，还有一种常见的方式，就是控制时间不变的协变量
 - 异质性处理效应下对于不同协变量 X_i 的条件 ATTs 会得到更大的权重 (Sloczynski,2022,REStat)

- 如果在没有处理发生时，具有不同特征的个体结果演化存在差异，会怎么样？

- 如果在没有处理发生时，具有不同特征的个体结果演化存在差异，会怎么样？
- 通常，与结果变量演化相关的处理前特征在处理组和控制组之间是非平衡 (unbalanced)，那么，PTA 可能并不可信

- 如果在没有处理发生时，具有不同特征的个体结果演化存在差异，会怎么样？
- 通常，与结果变量演化相关的处理前特征在处理组和控制组之间是非平衡 (unbalanced)，那么，PTA 可能并不可信
- 为了放松 PTA，研究者通常假设只有条件于一系列的可观测处理前协变量时，PTA 才成立

Tips: 对于时变协变量，实践中，通常使用 X_{it} 处理前（样本前）某一时点，或者处理前所有时期的平均值作为处理前协变量，例如，Tang and Xu (2025, JEEM)

- 如果在没有处理发生时，具有不同特征的个体结果演化存在差异，会怎么样？
- 通常，与结果变量演化相关的处理前特征在处理组和控制组之间是非平衡 (unbalanced)，那么，PTA 可能并不可信
- 为了放松 PTA，研究者通常假设只有条件于一系列的可观测处理前协变量时，PTA 才成立

Tips: 对于时变协变量，实践中，通常使用 X_{it} 处理前（样本前）某一时点，或者处理前所有时期的平均值作为处理前协变量，例如，Tang and Xu (2025, JEEM)

- 选择的混淆因子应该是基期未处理的协变量，既可以预测处理（可能会导致非平衡性），也会预测潜在结果 Y^0

- 如果在没有处理发生时，具有不同特征的个体结果演化存在差异，会怎么样？
- 通常，与结果变量演化相关的处理前特征在处理组和控制组之间是非平衡 (unbalanced)，那么，PTA 可能并不可信
- 为了放松 PTA，研究者通常假设只有条件于一系列的可观测处理前协变量时，PTA 才成立
Tips: 对于时变协变量，实践中，通常使用 X_{it} 处理前（样本前）某一时点，或者处理前所有时期的平均值作为处理前协变量，例如，Tang and Xu (2025, JEEM)
- 选择的混淆因子应该是基期未处理的协变量，既可以预测处理（可能会导致非平衡性），也会预测潜在结果 Y^0
- 选择方法：常识、逻辑、理论，以及对研究问题的理解等等，应该更加关注结果的决定因素，而不是处理

- 注意，我们缺失的是 $E[Y^0|D=1]$ ，因此，我们想要的协变量是可以高度预测缺失的潜在结果

- 注意，我们缺失的是 $E[Y^0|D=1]$ ，因此，我们想要的协变量是可以高度预测缺失的潜在结果
- 经验方法：

- 注意，我们缺失的是 $E[Y^0|D=1]$ ，因此，我们想要的协变量是可以高度预测缺失的潜在结果
- 经验方法：
 - 1 删掉所有处理后处理组，所有结果都是 Y^0 ，用它们与备选的协变量回归，并进行选择

- 注意，我们缺失的是 $E[Y^0|D=1]$ ，因此，我们想要的协变量是可以高度预测缺失的潜在结果
- 经验方法：
 - 1 删掉所有处理后处理组，所有结果都是 Y^0 ，用它们与备选的协变量回归，并进行选择
 - 2 删掉处理后处理组和控制组，用处理组的 Y^0 来对基期协变量回归，并进行选择

- 注意，我们缺失的是 $E[Y^0|D=1]$ ，因此，我们想要的协变量是可以高度预测缺失的潜在结果
- 经验方法：
 - 1 删掉所有处理后处理组，所有结果都是 Y^0 ，用它们与备选的协变量回归，并进行选择
 - 2 删掉处理后处理组和控制组，用处理组的 Y^0 来对基期协变量回归，并进行选择
 - 3 用 t-2 和 t-1 期，计算 $\Delta E[Y^0|D=1]$ ，并用它对 t-1 期协变量回归，进行选择

- 注意，我们缺失的是 $E[Y^0|D=1]$ ，因此，我们想要的协变量是可以高度预测缺失的潜在结果
- 经验方法：
 - 1 删掉所有处理后处理组，所有结果都是 Y^0 ，用它们与备选的协变量回归，并进行选择
 - 2 删掉处理后处理组和控制组，用处理组的 Y^0 来对基期协变量回归，并进行选择
 - 3 用 t-2 和 t-1 期，计算 $\Delta E[Y^0|D=1]$ ，并用它对 t-1 期协变量回归，进行选择
- 这些经验/数据可用的协变量选择方法仍然要注意检验它们的平衡性 (balance)

- 平行趋势不可检验，但是仍然可以用可观测协变量来评估是否可信

- 平行趋势不可检验，但是仍然可以用可观测协变量来评估是否可信
- 关键协变量的平衡性可以帮助评估处理组和控制组是否有类似的趋势

- 平行趋势不可检验，但是仍然可以用可观测协变量来评估是否可信
- 关键协变量的平衡性可以帮助评估处理组和控制组是否有类似的趋势
- 人口、经济、社会 and 制度条件差异可能意味着不满足平行趋势

- 平行趋势不可检验，但是仍然可以用可观测协变量来评估是否可信
- 关键协变量的平衡性可以帮助评估处理组和控制组是否有类似的趋势
- 人口、经济、社会 and 制度条件差异可能意味着不满足平行趋势
- 用不会受到处理影响的协变量来检验平衡性

- 基期协变量是在处理前测度的，检验它们在处理组和控制组之间的平衡性

- 基期协变量是在处理前测度的，检验它们在处理组和控制组之间的平衡性
- 用表格或者图形报告两组协变量的均值

- 基期协变量是在处理前测度的，检验它们在处理组和控制组之间的平衡性
- 用表格或者图形报告两组协变量的均值
- 计算标准化的差异

$$Norm.Diff = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{(S_T^2 - S_C^2)/2}}$$

- 基期协变量是在处理前测度的，检验它们在处理组和控制组之间的平衡性
- 用表格或者图形报告两组协变量的均值
- 计算标准化的差异

$$Norm.Diff = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{(S_T^2 - S_C^2)/2}}$$

- 上述标准化的差异测度了非平衡性

- 基期协变量是在处理前测度的，检验它们在处理组和控制组之间的平衡性
- 用表格或者图形报告两组协变量的均值
- 计算标准化的差异

$$Norm.Diff = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{(S_T^2 - S_C^2)/2}}$$

- 上述标准化的差异测度了非平衡性
- Imbens and Rubin(2015) 建议，标准化的差异小于 0.25 可以避免有问题的非平衡性特征

- 标准化的差异

- 标准化的差异
 - 测度组群均值的差异

- 标准化的差异
 - 测度组群均值的差异
 - 表示相对于方差的差异

- 标准化的差异
 - 测度组群均值的差异
 - 表示相对于方差的差异
 - 用于比较组群间平衡性

- 标准化的差异
 - 测度组群均值的差异
 - 表示相对于方差的差异
 - 用于比较组群间平衡性
- Z 得分

- 标准化的差异
 - 测度组群均值的差异
 - 表示相对于方差的差异
 - 用于比较组群间平衡性
- Z 得分
 - 测度单一观测值与均值的距离

- 标准化的差异
 - 测度组群均值的差异
 - 表示相对于方差的差异
 - 用于比较组群间平衡性
- Z 得分
 - 测度单一观测值与均值的距离
 - 表示标准偏离的距离

- 标准化的差异

- 测度组群均值的差异
- 表示相对于方差的差异
- 用于比较组群间平衡性

- Z 得分

- 测度单一观测值与均值的距离
- 表示标准偏离的距离
- 用于评估单个观测值有多不正常

- 标准化的差异

- 测度组群均值的差异
- 表示相对于方差的差异
- 用于比较组群间平衡性

- Z 得分

- 测度单一观测值与均值的距离
- 表示标准偏离的距离
- 用于评估单个观测值有多不正常

- 两种方法都可以标准化差异，并不受个体特征的影响，比较变量。但是，推荐在 DID 中使用标准化差异检验，临界值为 0.25

1 因果推断中的内生性

2 DID 中的协变量

条件平行趋势 PTA

重叠 (overlap) 假设

带协变量的 TWFE

带协变量的 TWFE 为什么有偏误

解决方法

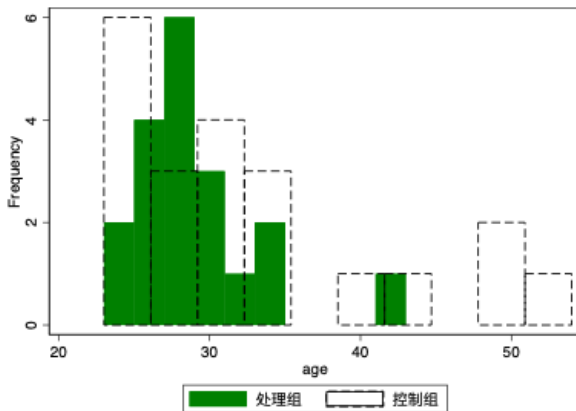
3 DID 中的机器学习

4 内生的处理配置

5 DID-IV 与 IV 型 DID

- 当协变量可用时，还需要一个额外假设：每个个体都有严格为正的概​​率出现在控制组中

- 当协变量可用时，还需要一个额外假设：每个个体都有严格为正的的概率出现在控制组中
- 例如，在职业培训项目评估中，处理组和控制组有相同协变量 X (age) 的人，也就是下图图的处理和控制组重叠 (overlap) 的部分



① 因果推断中的内生性

② DID 中的协变量

条件平行趋势 PTA

重叠 (overlap) 假设

带协变量的 TWFE

带协变量的 TWFE 为什么有偏误

解决方法

③ DID 中的机器学习

④ 内生的处理配置

⑤ DID-IV 与 IV 型 DID

- 在无条件 PTA，用下列 TWFE 来估计 ATT:

$$Y_{i,t} = \alpha_0 + \gamma_0 1\{G_i = 2\} + \lambda_0 1\{T_i = 2\} + \underbrace{\beta_0^{twfe}}_{ATT} (1\{G_i = 2\} \cdot 1\{T_i = 2\}) + \varepsilon_{i,t}$$

- 在无条件 PTA，用下列 TWFE 来估计 ATT：

$$Y_{i,t} = \alpha_0 + \gamma_0 1\{G_i = 2\} + \lambda_0 1\{T_i = 2\} + \underbrace{\beta_0^{twfe}}_{ATT} (1\{G_i = 2\} \cdot 1\{T_i = 2\}) + \varepsilon_{i,t}$$

- 更加常用的 TWFE 方程声明：

$$Y_{i,t} = \alpha_0 + \gamma_0 1\{G_i = 2\} + \lambda_0 1\{T_i = 2\} + \underbrace{\beta_0^{twfe}}_{???} (1\{G_i = 2\} \cdot 1\{T_i = 2\}) + X_i' \alpha_1 + \varepsilon_{i,t}$$

- 在无条件 PTA，用下列 TWFE 来估计 ATT：

$$Y_{i,t} = \alpha_0 + \gamma_0 1\{G_i = 2\} + \lambda_0 1\{T_i = 2\} + \underbrace{\beta_0^{twfe}}_{ATT} (1\{G_i = 2\} \cdot 1\{T_i = 2\}) + \varepsilon_{i,t}$$

- 更加常用的 TWFE 方程声明：

$$Y_{i,t} = \alpha_0 + \gamma_0 1\{G_i = 2\} + \lambda_0 1\{T_i = 2\} + \underbrace{\beta_0^{twfe}}_{???} (1\{G_i = 2\} \cdot 1\{T_i = 2\}) + X_i' \alpha_1 + \varepsilon_{i,t}$$

- 我们使用 Sant' Anna (2025) 的 DGP ($n = 1000$, 蒙特卡洛 10000 次)，真实 $ATT = 0$

- 在无条件 PTA, 用下列 TWFE 来估计 ATT:

$$Y_{i,t} = \alpha_0 + \gamma_0 1\{G_i = 2\} + \lambda_0 1\{T_i = 2\} + \underbrace{\beta_0^{twfe}}_{ATT} (1\{G_i = 2\} \cdot 1\{T_i = 2\}) + \varepsilon_{i,t}$$

- 更加常用的 TWFE 方程声明:

$$Y_{i,t} = \alpha_0 + \gamma_0 1\{G_i = 2\} + \lambda_0 1\{T_i = 2\} + \underbrace{\beta_0^{twfe}}_{???} (1\{G_i = 2\} \cdot 1\{T_i = 2\}) + X_i' \alpha_1 + \varepsilon_{i,t}$$

- 我们使用 Sant' Anna (2025) 的 DGP ($n = 1000$, 蒙特卡洛 10000 次), 真实 $ATT = 0$
- 用上述 TWFE 估计的系数 $\beta_0^{twfe} = -16.36$

① 因果推断中的内生性

② DID 中的协变量

条件平行趋势 PTA

重叠 (overlap) 假设

带协变量的 TWFE

带协变量的 TWFE 为什么有偏误

解决方法

③ DID 中的机器学习

④ 内生的处理配置

⑤ DID-IV 与 IV 型 DID

- TWFE 声明为

$$Y_{i,t} = \alpha_0 + \gamma_0 1\{G_i = 2\} + \lambda_0 1\{T_i = 2\} \\ + \underbrace{\beta_0^{twfe}}_{???} (1\{G_i = 2\} \cdot 1\{T_i = 2\}) + X_i' \alpha_1 + \varepsilon_{i,t}$$

- TWFE 声明为

$$Y_{i,t} = \alpha_0 + \gamma_0 1\{G_i = 2\} + \lambda_0 1\{T_i = 2\} + \underbrace{\beta_0^{twfe}}_{???} (1\{G_i = 2\} \cdot 1\{T_i = 2\}) + X_i' \alpha_1 + \varepsilon_{i,t}$$

- 注意，我们来看看：

$$\begin{aligned} \mathbb{E}[Y_{i,t} | G_i = \infty, T_i = 1, X_i] &= \tilde{\alpha}_0 + X_i' \tilde{\alpha}_1 \\ \mathbb{E}[Y_{i,t} | G_i = \infty, T_i = 2, X_i] &= \tilde{\alpha}_0 + \tilde{\lambda}_0 + X_i' \tilde{\alpha}_1 \\ \mathbb{E}[Y_{i,t} | G_i = 2, T_i = 1, X_i] &= \tilde{\alpha}_0 + \tilde{\gamma}_0 + X_i' \tilde{\alpha}_1 \\ \mathbb{E}[Y_{i,t} | G_i = 2, T_i = 2, X_i] &= \tilde{\alpha}_0 + \tilde{\gamma}_0 + \tilde{\lambda}_0 + \tilde{\beta}_0^{twfe} + X_i' \tilde{\alpha}_1 \end{aligned}$$

- 首先,

$$\mathbb{E}[Y_{i,t}|G_i = \infty, T_i = 2, X_i] - \mathbb{E}[Y_{i,t}|G_i = \infty, T_i = 1, X_i] = \tilde{\lambda}_0$$

控制组个体的结果变量演化并不依赖于 X_i

- 首先,

$$\mathbb{E}[Y_{i,t}|G_i = \infty, T_i = 2, X_i] - \mathbb{E}[Y_{i,t}|G_i = \infty, T_i = 1, X_i] = \tilde{\lambda}_0$$

控制组个体的结果变量演化并不依赖于 X_i

- 第二,

$$\mathbb{E}[Y_{i,t}|G_i = 2, T_i = 2, X_i] - \mathbb{E}[Y_{i,t}|G_i = 2, T_i = 1, X_i] = \tilde{\lambda}_0 + \tilde{\beta}_0^{\text{twfe}}$$

处理组个体结果变量演化不依赖于 X_i

- 首先,

$$\mathbb{E}[Y_{i,t}|G_i = \infty, T_i = 2, X_i] - \mathbb{E}[Y_{i,t}|G_i = \infty, T_i = 1, X_i] = \tilde{\lambda}_0$$

控制组个体的结果变量演化并不依赖于 X_i

- 第二,

$$\mathbb{E}[Y_{i,t}|G_i = 2, T_i = 2, X_i] - \mathbb{E}[Y_{i,t}|G_i = 2, T_i = 1, X_i] = \tilde{\lambda}_0 + \tilde{\beta}_0^{twfe}$$

处理组个体结果变量演化不依赖于 X_i

- 最后, 在 CPTA、NA、Overlap 假设下,

$$\begin{aligned} ATT(X) &= (\mathbb{E}[Y_{i,t}|G_i = 2, T_i = 2, X_i] - \mathbb{E}[Y_{i,t}|G_i = 2, T_i = 1, X_i]) \\ &\quad - (\mathbb{E}[Y_{i,t}|G_i = \infty, T_i = 2, X_i] - \mathbb{E}[Y_{i,t}|G_i = \infty, T_i = 1, X_i]) \\ &= \tilde{\beta}_0^{twfe} \end{aligned}$$

ATT 在协变量 X 的子群体之间是同质的

- 下面，证明偏误的来源【可跳过】：对反事实 $\mathbb{E}[Y_{i,t}|G_i = \infty, X_i]$ 引入协变量 X_i 的效应：

$$Y_{it}^{\infty} = \alpha_0 + \lambda_t + \gamma_i + X_i' \alpha_{1t} + \epsilon_{it}$$

其中， X_i 对 Y_{it}^{∞} 的效应可能随时间变化，因此， α_{1t} 带有时间下标。

- 下面，证明偏误的来源【可跳过】：对反事实 $\mathbb{E}[Y_{i,t}|G_i = \infty, X_i]$ 引入协变量 X_i 的效应：

$$Y_{it}^{\infty} = \alpha_0 + \lambda_t + \gamma_i + X_i' \alpha_{1t} + \epsilon_{it}$$

其中， X_i 对 Y_{it}^{∞} 的效应可能随时间变化，因此， α_{1t} 带有时间下标。



$$\begin{aligned} ATT(X) &= (\mathbb{E}[Y_{i,t}|G_i = 2, T_i = 2, X_i] - \mathbb{E}[Y_{i,t}|G_i = 2, T_i = 1, X_i]) \\ &\quad - (\mathbb{E}[Y_{i,t}|G_i = \infty, T_i = 2, X_i] - \mathbb{E}[Y_{i,t}|G_i = \infty, T_i = 1, X_i]) \\ &= \tilde{\beta}_0^{twfe} \\ &\quad + \underbrace{[E\{X_i'|G_i = 2\}(\alpha_{1t} - \alpha_{1t-1}) - E\{X_i'|G_i = \infty\}(\alpha_{1t} - \alpha_{1t-1})]}_{TWFE \text{ bias}} \end{aligned}$$

这意味着，如果协变量 X_i 对结果的效应为常数 ($\alpha_{it} = \alpha_{it-1}$)，或者协变量在处理组和控制组的均值相等 ($E\{X_i'|G_i = 2\} = E\{X_i'|G_i = \infty\}$)，上述 DID 估计量才能识别 ATT。

1 因果推断中的内生性

2 DID 中的协变量

条件平行趋势 PTA

重叠 (overlap) 假设

带协变量的 TWFE

带协变量的 TWFE 为什么有偏误

解决方法

3 DID 中的机器学习

4 内生的处理配置

5 DID-IV 与 IV 型 DID

回归调整法 RA

- 最熟悉的的就是回归调整法，由 Heckman et al.(1997)、Heckman et al.(1998) 提出，直接建模结果变量的演化

回归调整法 RA

- 最熟悉的就是回归调整法，由 Heckman et al.(1997)、Heckman et al.(1998) 提出，直接建模结果变量的演化
- 我们已经看到，在一般性假设下，

$$\begin{aligned}
 ATT &= \mathbb{E} \left[\underbrace{\mathbb{E}[Y_{t=2} | G = 2, X]}_{=m_{t=2}^{G=2}(X)} - \underbrace{\mathbb{E}[Y_{t=1} | G = 2, X]}_{=m_{t=1}^{G=2}(X)} \right] - \left(\underbrace{\mathbb{E}[Y_{t=2} | G = \infty, X]}_{=m_{t=2}^{G=\infty}(X)} - \underbrace{\mathbb{E}[Y_{t=1} | G = \infty, X]}_{=m_{t=1}^{G=\infty}(X)} \right) \\
 &= \mathbb{E} \left[\left(m_{t=2}^{G=2}(X) - m_{t=1}^{G=2}(X) \right) - \left(m_{t=2}^{G=\infty}(X) - m_{t=1}^{G=\infty}(X) \right) \middle| G = 2 \right]
 \end{aligned}$$

回归调整法 RA

- 最熟悉的就是回归调整法，由 Heckman et al.(1997)、Heckman et al.(1998) 提出，直接建模结果变量的演化
- 我们已经看到，在一般性假设下，

$$\begin{aligned}
 ATT &= \mathbb{E} \left[\left(\underbrace{\mathbb{E}[Y_{t=2} | G=2, X]}_{=m_{t=2}^{G=2}(X)} - \underbrace{\mathbb{E}[Y_{t=1} | G=2, X]}_{=m_{t=1}^{G=2}(X)} \right) - \left(\underbrace{\mathbb{E}[Y_{t=2} | G=\infty, X]}_{=m_{t=2}^{G=\infty}(X)} - \underbrace{\mathbb{E}[Y_{t=1} | G=\infty, X]}_{=m_{t=1}^{G=\infty}(X)} \right) \right] \\
 &= \mathbb{E} \left[\left(m_{t=2}^{G=2}(X) - m_{t=1}^{G=2}(X) \right) - \left(m_{t=2}^{G=\infty}(X) - m_{t=1}^{G=\infty}(X) \right) \middle| G=2 \right]
 \end{aligned}$$

- 关键是用合适的方法估计未知函数 $m_t^G(X)$: 参数、半参数、非参数

回归调整法 RA

- 例如，我们将结果的演化函数 $m_t^G(X)$ 建模为：
$$\mu_{t=s}^{G=g}(X) = X' \beta_{0,t=s}^{G=g}$$

回归调整法 RA

- 例如，我们将结果的演化函数 $m_t^G(X)$ 建模为：

$$\mu_{t=s}^{G=g}(X) = X' \beta_{0,t=s}^{G=g}$$
- 然后，我们可以利用 OLS 估计每个 X 子样本的 β_s ，用处理组个体的所有协变量值来拟合 $\hat{\mu}_{t=s}^{G=g}(X)$ ，并将这些拟合值结合起来：

$$\widehat{ATT}_n^{reg} = \mathbb{E}_n \left[\left(\hat{\mu}_{t=2}^{G=2}(X) - \hat{\mu}_{t=1}^{G=2}(X) \right) - \left(\hat{\mu}_{t=2}^{G=\infty}(X) - \hat{\mu}_{t=1}^{G=\infty}(X) \right) \middle| G=2 \right]$$

回归调整法 RA

- 例如，我们将结果的演化函数 $m_t^G(X)$ 建模为：

$$\mu_{t=s}^{G=g}(X) = X' \beta_{0,t=s}^{G=g}$$
- 然后，我们可以利用 OLS 估计每个 X 子样本的 β_s ，用处理组个体的所有协变量值来拟合 $\hat{\mu}_{t=s}^{G=g}(X)$ ，并将这些拟合值结合起来：

$$\widehat{ATT}_n^{reg} = \mathbb{E}_n \left[\left(\hat{\mu}_{t=2}^{G=2}(X) - \hat{\mu}_{t=1}^{G=2}(X) \right) - \left(\hat{\mu}_{t=2}^{G=\infty}(X) - \hat{\mu}_{t=1}^{G=\infty}(X) \right) \right] \Big| G=2$$

- 回归调整法 DID 估计量依赖于研究者正确建模结果演化

逆概率加权法 IPW

- 我们其实很难知道结果的演化函数

逆概率加权法 IPW

- 我们其实很难知道结果的演化函数
- 换一种方法——建模倾向得分，例如，属于处理组的概率， $p(X) \equiv \mathbb{P}(G = 2|X) = \mathbb{P}(D = 1|X)$

逆概率加权法 IPW

- 我们其实很难知道结果的演化函数
- 换一种方法——建模倾向得分，例如，属于处理组的概率， $p(X) \equiv \mathbb{P}(G = 2|X) = \mathbb{P}(D = 1|X)$
- 这种方法由 Adabie(2005) 提出：

$$ATT^{ipw} = \frac{\mathbb{E} \left[\left(D - \frac{(1-D)p(X)}{1-p(X)} \right) (Y_{t=2} - Y_{t=1}) \right]}{\mathbb{E}[D]}$$

逆概率加权法 IPW

- 我们其实很难知道结果的演化函数
- 换一种方法——建模倾向得分，例如，属于处理组的概率， $p(X) \equiv \mathbb{P}(G = 2|X) = \mathbb{P}(D = 1|X)$
- 这种方法由 Adadie(2005) 提出：

$$ATT^{ipw} = \frac{\mathbb{E} \left[\left(D - \frac{(1-D)p(X)}{1-p(X)} \right) (Y_{t=2} - Y_{t=1}) \right]}{\mathbb{E}[D]}$$

- 这些公式意味着，可以使用两步估计法：

逆概率加权法 IPW

- 我们其实很难知道结果的演化函数
- 换一种方法——建模倾向得分，例如，属于处理组的概率， $p(X) \equiv \mathbb{P}(G = 2|X) = \mathbb{P}(D = 1|X)$
- 这种方法由 Adadie(2005) 提出：

$$ATT^{ipw} = \frac{\mathbb{E} \left[\left(D - \frac{(1-D)p(X)}{1-p(X)} \right) (Y_{t=2} - Y_{t=1}) \right]}{\mathbb{E}[D]}$$

- 这些公式意味着，可以使用两步估计法：
 - 1 选择合适的方法来估计未知的倾向得分 $p(X)$

逆概率加权法 IPW

- 我们其实很难知道结果的演化函数
- 换一种方法——建模倾向得分，例如，属于处理组的概率， $p(X) \equiv \mathbb{P}(G = 2|X) = \mathbb{P}(D = 1|X)$
- 这种方法由 Adadie(2005) 提出：

$$ATT^{ipw} = \frac{\mathbb{E} \left[\left(D - \frac{(1-D)p(X)}{1-p(X)} \right) (Y_{t=2} - Y_{t=1}) \right]}{\mathbb{E}[D]}$$

- 这些公式意味着，可以使用两步估计法：
 - 1 选择合适的方法来估计未知的倾向得分 $p(X)$
 - 2 将估计的倾向得分代入 **ATT 公式**，用样本期望取代总体期望

逆概率加权法 IPW

- 例如，用 $\pi X = \frac{e^{X'\gamma_0}}{1+e^{X'\gamma_0}}$ 来作为倾向得分的模型

逆概率加权法 IPW

- 例如，用 $\pi X = \frac{e^{X'\gamma_0}}{1+e^{X'\gamma_0}}$ 来作为倾向得分的模型
- 可以用 logit ML 估计 γ_0

逆概率加权法 IPW

- 例如，用 $\pi X = \frac{e^{X'\gamma_0}}{1+e^{X'\gamma_0}}$ 来作为倾向得分的模型
- 可以用 logit ML 估计 γ_0
- 将估计的 $\hat{\pi}(X)$ 代入 Abadie 的 ATT 估计量中：

$$\hat{ATT}^{ipw} = \frac{\mathbb{E} \left[\left(D - \frac{(1-D)\hat{\pi}(X)}{1-\hat{\pi}(X)} \right) (Y_{t=2} - Y_{t=1}) \right]}{\mathbb{E}[D]}$$

逆概率加权法 IPW

- 例如，用 $\pi X = \frac{e^{X'\gamma_0}}{1+e^{X'\gamma_0}}$ 来作为倾向得分的模型
- 可以用 logit ML 估计 γ_0
- 将估计的 $\hat{\pi}(X)$ 代入 Abadie 的 ATT 估计量中：

$$\hat{ATT}^{ipw} = \frac{\mathbb{E} \left[\left(D - \frac{(1-D)\hat{\pi}(X)}{1-\hat{\pi}(X)} \right) (Y_{t=2} - Y_{t=1}) \right]}{\mathbb{E}[D]}$$

- Abadie 的 IPW DID 估计量的一个潜在缺陷是：权重非正态，例如，之和并非为 1

逆概率加权法 IPW

- 例如，用 $\pi X = \frac{e^{X'\gamma_0}}{1+e^{X'\gamma_0}}$ 来作为倾向得分的模型
- 可以用 logit ML 估计 γ_0
- 将估计的 $\hat{\pi}(X)$ 代入 Abadie 的 ATT 估计量中：

$$\hat{ATT}^{ipw} = \frac{\mathbb{E} \left[\left(D - \frac{(1-D)\hat{\pi}(X)}{1-\hat{\pi}(X)} \right) (Y_{t=2} - Y_{t=1}) \right]}{\mathbb{E}[D]}$$

- Abadie 的 IPW DID 估计量的一个潜在缺陷是：权重非正态，例如，之和并非为 1
- 而 Hajek(1971) 类型的估计量更加稳定，因为他们使用“正态”权重

逆概率加权法 IPW

- 例如，用 $\pi X = \frac{e^{X'\gamma_0}}{1+e^{X'\gamma_0}}$ 来作为倾向得分的模型
- 可以用 logit ML 估计 γ_0
- 将估计的 $\hat{\pi}(X)$ 代入 Abadie 的 ATT 估计量中：

$$\hat{ATT}^{ipw} = \frac{\mathbb{E} \left[\left(D - \frac{(1-D)\hat{\pi}(X)}{1-\hat{\pi}(X)} \right) (Y_{t=2} - Y_{t=1}) \right]}{\mathbb{E}[D]}$$

- Abadie 的 IPW DID 估计量的一个潜在缺陷是：权重非正态，例如，之和并非为 1
- 而 Hajek(1971) 类型的估计量更加稳定，因为他们使用“正态”权重
- Sant' Anna and Zhao(2020) 基于 Abadie(2005)，引入了 Hajek(1971) 类型的 IPW DID 估计量

逆概率加权法 IPW

- Sant' Anna and Zhao(2020) 的 IPW DID 估计量公式:

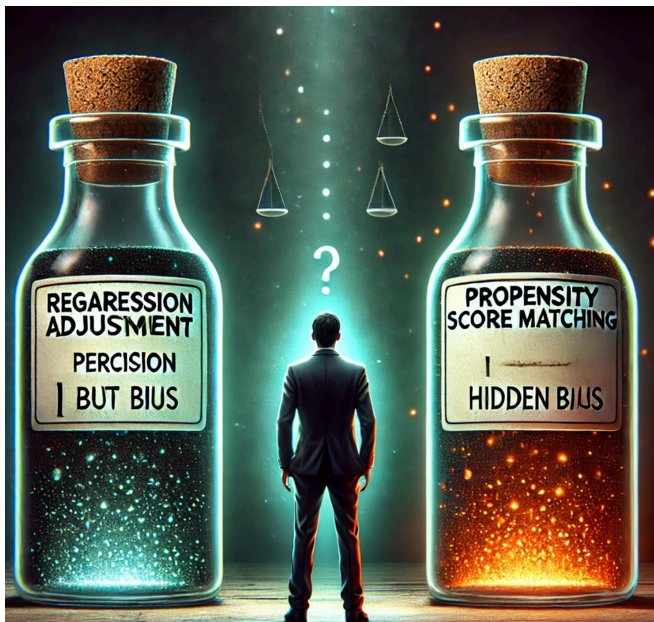
$$\begin{aligned}
 ATT_{sz}^{ipw} &= \mathbb{E}[(w_{G=2}^p(D) - w_{G=\infty}^p(D, X; p))(Y_{t=2} - Y_{t=1})] \\
 &= \mathbb{E}\left[\left(\frac{D}{\mathbb{E}[D]} - \frac{\frac{p(X)(1-D)}{1-p(X)}}{\mathbb{E}\left[\frac{p(X)(1-D)}{1-p(X)}\right]}\right)(Y_{t=2} - Y_{t=1})\right]
 \end{aligned}$$

逆概率加权法 IPW

- Sant' Anna and Zhao(2020) 的 IPW DID 估计量公式:

$$\begin{aligned} ATT_{sz}^{ipw} &= \mathbb{E}[(w_{G=2}^p(D) - w_{G=\infty}^p(D, X; p))(Y_{t=2} - Y_{t=1})] \\ &= \mathbb{E}\left[\left(\frac{D}{\mathbb{E}[D]} - \frac{\frac{p(X)(1-D)}{1-p(X)}}{\mathbb{E}\left[\frac{p(X)(1-D)}{1-p(X)}\right]}\right)(Y_{t=2} - Y_{t=1})\right] \end{aligned}$$

- IPW DID 估计量依赖于研究者建模倾向得分的能力



双重稳健 DID 估计量

- 结合 RA 和 IPW 的 DID 估计量，由 Sant' Anna and Zhao(2020) 提出

双重稳健 DID 估计量

- 结合 RA 和 IPW 的 DID 估计量，由 Sant' Anna and Zhao(2020) 提出
- 只要回归调整和倾向得分之一被正确建模出来，这个估计量就是一致的

双重稳健 DID 估计量

- 结合 RA 和 IPW 的 DID 估计量，由 Sant' Anna and Zhao(2020) 提出
- 只要回归调整 and 倾向得分之一被正确建模出来，这个估计量就是一致的
- 估计量的公式：

$$\begin{aligned}
 ATT^{dr,p} &= \mathbb{E} \left[\left(w_{G=2}^p(D) - w_{G=\infty}^p(D, X; p) \right) \left((Y_{t=2} - Y_{t=1}) - (m_{t=2}^{C=\infty}(X) - m_{t=1}^{C=\infty}(X)) \right) \right] \\
 &= \mathbb{E} \left[\left(\frac{D}{\mathbb{E}[D]} - \frac{\frac{p(X)(1-D)}{1-p(X)}}{\mathbb{E} \left[\frac{p(X)(1-D)}{1-p(X)} \right]} \right) \left((Y_{t=2} - Y_{t=1}) - (m_{t=2}^{G=\infty}(X) - m_{t=1}^{G=\infty}(X)) \right) \right]
 \end{aligned}$$

双重稳健 DID 估计量

- 结合 RA 和 IPW 的 DID 估计量，由 Sant' Anna and Zhao(2020) 提出
- 只要回归调整 and 倾向得分之一被正确建模出来，这个估计量就是一致的
- 估计量的公式：

$$\begin{aligned}
 ATT^{dr,p} &= \mathbb{E} \left[\left(w_{G=2}^p(D) - w_{G=\infty}^p(D, X; p) \right) \left((Y_{t=2} - Y_{t=1}) - (m_{t=2}^{C=\infty}(X) - m_{t=1}^{C=\infty}(X)) \right) \right] \\
 &= \mathbb{E} \left[\left(\frac{D}{\mathbb{E}[D]} - \frac{\frac{p(X)(1-D)}{1-p(X)}}{\mathbb{E} \left[\frac{p(X)(1-D)}{1-p(X)} \right]} \right) \left((Y_{t=2} - Y_{t=1}) - (m_{t=2}^{G=\infty}(X) - m_{t=1}^{G=\infty}(X)) \right) \right]
 \end{aligned}$$

- Sant' Anna and Zhao(2020) 现实了 $ATT^{dr,p}$ 是半参数有效的

双重稳健 DID 估计量

- 结合 RA 和 IPW 的 DID 估计量，由 Sant' Anna and Zhao(2020) 提出
- 只要回归调整 and 倾向得分之一被正确建模出来，这个估计量就是一致的
- 估计量的公式：

$$\begin{aligned}
 ATT^{dr,p} &= \mathbb{E} \left[\left(w_{G=2}^p(D) - w_{G=\infty}^p(D, X; p) \right) \left((Y_{t=2} - Y_{t=1}) - (m_{t=2}^{C=\infty}(X) - m_{t=1}^{C=\infty}(X)) \right) \right] \\
 &= \mathbb{E} \left[\left(\frac{D}{\mathbb{E}[D]} - \frac{\frac{p(X)(1-D)}{1-p(X)}}{\mathbb{E} \left[\frac{p(X)(1-D)}{1-p(X)} \right]} \right) \left((Y_{t=2} - Y_{t=1}) - (m_{t=2}^{G=\infty}(X) - m_{t=1}^{G=\infty}(X)) \right) \right]
 \end{aligned}$$

- Sant' Anna and Zhao(2020) 现实了 $ATT^{dr,p}$ 是半参数有效的
- 如果所有的函数形式正确声明，DRDID 估计量是最优效率的

软件命令与应用文献

- 上述方法均可以使用 R 命令 (drdid/did/fect)、stata 命令 (drdid、csdid/csdid2/xtevent)、python 命令 (drdid/csdid)
- CuadrosMenaca, A., Thomsen, M. R., & Nayga Jr, R. M. (2023). School breakfast and student behavior. American Journal of Agricultural Economics, 105(1), 99-121.

其它的应用方法

- Caetano et al.(2022)、Caetano and Callaway(2024) 还给出了一些其它的实践中可用的方法：

其它的应用方法

- Caetano et al.(2022)、Caetano and Callaway(2024) 还给出了一些其它的实践中可用的方法：
 - a 只要满足下列条件，仍然可以控制处理前协变量来消除受处理影响的时变协变量带来的偏误：

$$X_{t^*}(0) \perp D | X_{t^*-1}$$

这个额外的条件意味着，协变量在有相同处理前特征 X 和时间不变特征 Z 的处理组和控制组之间有类似的演化路径

Tips: 实践中，我们可以检验处理前时期的样本，来看看是否有证据来支持这个条件

其它的应用方法

- b 我们对时变协变量的外生性非常有信心，这个时候要从理论、相关文献、制度背景、经济社会现实等方面来阐述这种外生性的可信度

其它的应用方法

- b 我们对时变协变量的外生性非常有信心，这个时候要从理论、相关文献、制度背景、经济社会现实等方面来阐述这种外生性的可信度
- c 施加额外的协变量 X 效应线性假设，然后使用新 DID 中的 imputation 方法，例如，stata 命令 `fect/jwddid/wooldid/did_imputation/did2s` 等等

其它的应用方法

- b 我们对时变协变量的外生性非常有信心，这个时候要从理论、相关文献、制度背景、经济社会现实等方面来阐述这种外生性的可信度
- c 施加额外的协变量 X 效应线性假设，然后使用新 DID 中的 imputation 方法，例如，stata 命令 `fect/jwddid/wooldid/did_imputation/did2s` 等等
- d Karim and Webb(2024) 施加了另一个假设（双向 CCC）：协变量的效应在个体和时间维度是相同的。在这个假设下，他们提出了三种类型 DID-INT 估计量，用的是结果变量余值的平行趋势（调整协变量）

其它的应用方法

- b 我们队时变协变量的外生性非常有信心，这个时候要从理论、相关文献、制度背景、经济社会现实等方面来阐述这种外生性的可信度
- c 施加额外的协变量 X 效应线性假设，然后使用新 DID 中的 imputation 方法，例如，stata 命令 `fect/jwddid/wooldid/did_imputation/did2s` 等等
- d Karim and Webb(2024) 施加了另一个假设（双向 CCC）：协变量的效应在个体和时间维度是相同的。在这个假设下，他们提出了三种类型 DID-INT 估计量，用的是结果变量余值的平行趋势（调整协变量）
- e Viviano and Bradic(2024) 提出一种方法用 LP 来允许时变协变量依赖处理路径，R 命令 `DynBalancing`

- 1 因果推断中的内生性
- 2 DID 中的协变量
- 3 DID 中的机器学习
- 4 内生的处理配置
- 5 DID-IV 与 IV 型 DID
- 6 参考文献

- 1 因果推断中的内生性
- 2 DID 中的协变量
- 3 DID 中的机器学习
- 4 内生的处理配置
- 5 DID-IV 与 IV 型 DID
- 6 参考文献

- ◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ↺ 🔍 ↻

- ◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ◻ ↺ 🔍 ↻

- 1 Jing Li, Difference-in-Difference Method, 2025 spring
- 2 Sant' Anna, DiD Resources, 2025
- 3 Callaway, Five Minute Summary: Difference in differences with time-varying covariates, 2022
- 4 Soczyski, Tymon. "Interpreting OLS estimands when treatment effects are heterogeneous: Smaller groups get larger weights." Review of Economics and Statistics 104.3 (2022): 501-509.
- 5 Zhang L, Xu W. Television exposure in early childhood and subsequent cognitive outcomes: Evidence from rural China[J]. Journal of Economic Behavior & Organization, 2025, 230: 106878.
- 6 Tang L, Xu W. Patronage and pollution[J]. Journal of Environmental Economics and Management, 2025, 130: 103115.

Thanks!