超越双重差分：面板数据与处理效应

慧航

2025年10月

Fuzzy DID

基本设定:

- W为实际处理变量，$W=1$为被处理，否则$W=0$
- $G=0/1$为一个分组变量，$G=1$为处理组，$G=0$为对照组
- $T=0/1$：事前、事后
- Y为收入
- 例子：Duflo(2001)
    - W为是否入学
    - G为该地区是否扩建学校：1973-1974的INPRES计划
    - 两个cohorts:
        - 1957-1962出生：计划推行时已经上完学（$T=0$）
        - 1968-1972出生：受到计划的影响（$T=1$）
    - 理想情况：$W=G\times T$：计划施行之前都没上过学，但是计划施行之后有INPRES计划的上学了
    - 然而事实并非如此，$W\neq G\times T$，处理组只是被处理的个体更多了

Fuzzy DID

一些notation:

- W_{gt}、Y_{gt}：$W|G=g,T=t$、$Y|G=g,T=t$
- Y_{wgt}：$Y|W=w,G=g,T=t$

W的假设

W可以是binary处理，也可以是多个组，或者有序变量。方便起见，我们就W为二元变量为例。

数据形式

该方法最适合重复截面（repeated cross-sectional）数据，以及使用cohort作为时间变量的设定（实质也是重复截面）。

Fuzzy DID的基本假设

假设1：模糊设计，第一阶段，即G对W的影响

$\mathbb{E}(W_{11}) > \mathbb{E}(W_{10})$，$\mathbb{E}(W_{11}) - \mathbb{E}(W_{10}) > \mathbb{E}(W_{01}) - \mathbb{E}(W_{00})$

- 该假设定义了处理组：被处理的比例增加最大的组
- 如果处理比率降低，那么重新定义$\tilde{W} = 1 - W$即可。

假设2：参与方程

$W = \mathbb{1}\{V \geq v_{GT}\}$，其中$V \perp\!\!\!\perp T|G$

- 该假设即假设了处理变量W的一个潜在的index结构。这里假设了有个潜在的"参与倾向"V以及一个随着时间、组别变化的临界值v_{GT}。
- 在重复截面中，$W(T)$不可同时观测。

Fuzzy DID的基本假设

接下来我们以一个两期、两组、二元处理为例。

假设3：共同趋势假设

$\mathbb{E}\left(Y\left(0\right)|G,T=1\right)-\mathbb{E}\left(Y\left(0\right)|G,T=0\right)$与$G$无关。

这个假定是在DID中常见的共同趋势假设。

假设4：（时间）同质的处理效应

$\mathbb{E}\left(Y\left(w\right)-Y\left(0\right)|G,T=1,W\left(0\right)=w\right)=\mathbb{E}\left(Y\left(w\right)-Y\left(0\right)|G,T=0,W\left(0\right)=w\right)$

Fuzzy DID中的转换者

可以定义转换者（switcher）：

$$S = \{W(0) < W(1), G = 1\}$$

其中:

$$W(t) = \mathbb{1}\{V \geq v_{Gt}\}$$

从而转换者即$G = 1$组别中从未处理转到被处理的个体。可以定义LATE:

$$\Delta = \mathbb{E}(Y_{11}(1) - Y_{11}(0) | S)$$

即在$G = 1$组别的转换者当$T = 1$时的处理效应。同时，令:

$$S' = \{W(0) \neq W(1), G = 0\}$$

为控制组的转换者，以及:

$$\Delta' = \mathbb{E}(Y_{01}(1) - Y_{01}(0) | S')$$

Fuzzy DID的识别

有了以上准备后，我们可以得到：

$$
\begin{aligned}
\mathbb{E}\left(Y_{11}\right) = \mathbb{E} & \left(Y_{11}\left(1\right) \times \left[\underbrace{\mathbb{1}\left\{v_{11} \leq V_{11} < v_{10}\right\}}_{Switcher} + \underbrace{\mathbb{1}\left\{V_{11} \geq v_{10}\right\}}_{Always\ taker}\right]\right) \\
& + \mathbb{E}\left(Y_{11}\left(0\right) \times \mathbb{1}\left\{V_{11} < v_{11}\right\}\right) \\
& = \mathbb{E}\left(Y_{11}\left(1\right) \times \left[\mathbb{1}\left\{v_{11} \leq V_{11} < v_{10}\right\} + \mathbb{1}\left\{V_{11} \geq v_{10}\right\}\right]\right) \\
& + \mathbb{E}\left(Y_{11}\left(0\right) \times \left(1 - \mathbb{1}\left\{v_{11} \leq V_{11} < v_{10}\right\} - \mathbb{1}\left\{V_{11} \geq v_{10}\right\}\right)\right) \\
& = \mathbb{E}\left(\left[Y_{11}\left(1\right) - Y_{11}\left(0\right)\right] \times \mathbb{1}\left\{v_{11} \leq V_{11} < v_{10}\right\}\right) \\
& + \mathbb{E}\left(\left[Y_{11}\left(1\right) - Y_{11}\left(0\right)\right] \times \mathbb{1}\left\{V_{11} \geq v_{10}\right\}\right) + \mathbb{E}\left(Y_{11}\left(0\right)\right)
\end{aligned}
$$

Fuzzy DID的识别

同理：

$$
\begin{aligned}
\mathbb{E}\left(Y_{10}\right) = & \mathbb{E}\left(Y_{10}(1) \times \underbrace{\mathbb{1}\left\{V_{10} \geq v_{10}\right\}}_{Always\ taker}\right) \\
& + \mathbb{E}\left(Y_{10}(0) \times \underbrace{\mathbb{1}\left\{V_{10} < v_{10}\right\}}_{Switcher\&Never\ taker}\right) \\
& = \mathbb{E}\left(\left[Y_{10}(1) - Y_{10}(0)\right] \times \mathbb{1}\left\{V_{10} \geq v_{10}\right\}\right) + \mathbb{E}\left(Y_{10}(0)\right)
\end{aligned}
$$

从而：

$$
\begin{aligned}
\mathbb{E}\left(Y_{11} - Y_{10}\right) = & \mathbb{E}\left(\left[Y_{11}(1) - Y_{11}(0)\right] \times \mathbb{1}\left\{v_{11} \leq V_{11} < v_{10}\right\}\right) \\
& + \mathbb{E}\left(\left[Y_{11}(1) - Y_{11}(0)\right] \times \mathbb{1}\left\{V_{11} \geq v_{10}\right\}\right) + \mathbb{E}\left(Y_{11}(0)\right) \\
& - \mathbb{E}\left(\left[Y_{10}(1) - Y_{10}(0)\right] \times \mathbb{1}\left\{V_{10} \geq v_{10}\right\}\right) - \mathbb{E}\left(Y_{10}(0)\right)
\end{aligned}
$$

Fuzzy DID的识别

根据假设4，得到：

$$
\begin{aligned}
\mathbb{E}\left(Y_{11}-Y_{10}\right) &= \mathbb{E}\left(\left[Y_{11}(1)-Y_{11}(0)\right]\times \mathbb{1}\left\{v_{11}\leq V_{11}<v_{10}\right\}\right) \\
&\quad + \mathbb{E}\left(Y_{11}(0)\right)-\mathbb{E}\left(Y_{10}(0)\right) \\
&= \mathbb{E}\left(Y_{11}(1)-Y_{11}(0)\,|S,G=1\right)\times \mathrm{P}\left(S|G=1\right) \\
&\quad + \mathbb{E}\left(Y_{11}(0)\right)-\mathbb{E}\left(Y_{10}(0)\right) \\
&= \Delta\times \mathrm{P}\left(S|G=1\right)+\mathbb{E}\left(Y_{11}(0)\right)-\mathbb{E}\left(Y_{10}(0)\right)
\end{aligned}
$$

Fuzzy DID的识别

同理：

$$\mathbb{E}\left(Y_{01}-Y_{00}\right)=\Delta' \times \mathrm{P}\left(S'|G=0\right)+\mathbb{E}\left(Y_{01}(0)\right)-\mathbb{E}\left(Y_{00}(0)\right)$$

从而根据共同趋势假设，双重差分：

$$\begin{aligned}
\mathbb{E}\left(Y_{11}-Y_{10}\right)-\mathbb{E}\left(Y_{01}-Y_{00}\right) &= \Delta \times \mathrm{P}(S|G=1) \\
&\quad+\mathbb{E}\left(Y_{11}(0)\right)-\mathbb{E}\left(Y_{10}(0)\right) \\
&\quad-\Delta' \times \mathrm{P}\left(S'|G=0\right) \\
&\quad-\left[\mathbb{E}\left(Y_{01}(0)\right)-\mathbb{E}\left(Y_{00}(0)\right)\right] \\
&= \Delta \times \mathrm{P}(S|G=1)-\Delta' \times \mathrm{P}\left(S'|G=0\right)
\end{aligned}$$

Fuzzy DID的识别

此外：

$$
\begin{aligned}
\mathrm{P}\left(W_{11}=1\right)-\mathrm{P}\left(W_{10}=1\right)&=\mathrm{P}\left(S \mid G=1\right) \\
\mathrm{P}\left(W_{01}=1\right)-\mathrm{P}\left(W_{00}=1\right)&=\mathrm{P}\left(S' \mid G=1\right)
\end{aligned}
$$

从而：

$$
\begin{aligned}
W_{\mathrm{DID}} &= \frac{\mathrm{DID}_Y}{\mathrm{DID}_W} \\
&= \frac{\mathbb{E}\left(Y_{11}-Y_{10}\right)-\mathbb{E}\left(Y_{01}-Y_{00}\right)}{\left[\mathrm{P}\left(W_{11}=1\right)-\mathrm{P}\left(W_{10}=1\right)\right]-\left[\mathrm{P}\left(W_{01}=1\right)-\mathrm{P}\left(W_{00}=1\right)\right]} \\
&= \frac{\Delta \times \mathrm{P}\left(S \mid G=1\right)-\Delta' \times \mathrm{P}\left(S' \mid G=0\right)}{\mathrm{P}\left(S \mid G=1\right)-\mathrm{P}\left(S' \mid G=0\right)} \\
&\triangleq \alpha \Delta+(1-\alpha)\Delta'
\end{aligned}
$$

Fuzzy DID的识别

识别陷阱

值得注意的是，如果$\mathrm{P}(S'|G=0)>0$，那么$\alpha>1$，意味着W_{DID}可能不能识别任何处理效应，甚至也不是任何处理效应的（正的权重）的平均。

所以我们需要一些额外的条件才能够识别Δ:

识别条件1

假设$\Delta=\Delta'$，即处理组、控制组的LATE相同

识别条件2

假设:

$$0<\mathrm{P}(W_{01})=\mathrm{P}(W_{00})<1$$

从而$\mathrm{P}(S'|G=0)=0$，$\alpha=1$。

时间修正的Fuzzy DID

以上我们根据假设1-4找到了Wald类型的Fuzzy DID估计量，现在我们放弃共同趋势假设（假设3）以及时间同质性假设（假设4），转而使用如下假设：

假设5：条件共同趋势假设

假设$\mathbb{E}\left(Y\left(w\right)|G, T=1, W\left(0\right)=w\right)-\mathbb{E}\left(Y\left(w\right)|G, T=0, W\left(0\right)=w\right)$与$G$无关

以上假设相较于假设3，多给定了第0时期的处理状态。如此，在假定1、2、5以及识别条件2的假设下，可以定义：

$$W_{\mathrm{TC}}=\frac{\mathbb{E}\left(Y_{11}\right)-\mathbb{E}\left(Y_{10}+\left[\mathbb{E}\left(Y_{w01}\right)-\mathbb{E}\left(Y_{w00}\right)\right]\right)}{\mathrm{P}\left(W_{11}=1\right)-\mathrm{P}\left(W_{10}=1\right)}$$

就得到了“时间修正”的Wald估计量。本质：使用控制组的$W=w$的样本趋势为处理组做反事实。

双向固定效应估计

- 正如我们在面板数据的分析中，为了消除不可观测异质性，我们通常采取两种办法：
  - 固定效应估计量+时间固定效应（双向固定效应估计量）
  - 一阶差分估计量+时间固定效应
- 在DID中，我们同样也使用了双向固定效应估计DID估计量。
- 尽管双向固定效应估计量如此普遍，但是仍然发现很多奇怪的现象：
  - 固定效应不显著，但是一阶差分显著（Gentzkow et al., 2011）
  - 加权重与否对结果影响很大（Enikolopov et al., 2011）
- de Chaisemartin和D’Haultfoeuille(2020)认为这是异质性处理效应在不同估计下权重不同的结果。

一般设定

- 假设一个处理变量W以及一个结果变量Y，从而反事实为$Y(W)$。
  - 我们暂且考虑$W=0/1$
  - 后续可以扩展到有序的处理变量
- 假设一个分组变量$G\in\{0,1,2,...,\bar{G}\}$以及时间变量$T\in\{0,1,2,...,\bar{T}\}$
- 注意每一个g,t的组合里面可能有一个观测，也可能有不止一个观测
- 在此基础上，我们引入一个重要的假定：

假设1：平衡面板

假设对于所有的$(g,t)\in\{0,1,2,...,\bar{G}\}\times\{0,1,2,...,\bar{T}\}$都有观测（$\mathrm{P}(G=g,T=t)>0$）

重要假定

仿照双重差分模型的假设:

假设2：共同趋势假设

对于$t \in \{0, 1, 2, ..., \bar{T}\}$，有:

$$\mathbb{E}\left(Y\left(0\right)|G, T = t\right) - \mathbb{E}\left(Y\left(0\right)|G, T = t - 1\right)$$

与G无关。

该假设与双重差分模型中的共同趋势假设相同。

重要假定

此外，还需要对于处理变量的假定：

假设3：处理的单调性

对于每一期$t \in \{0, 1, 2, ..., \bar{T}\}$,$W = W(T)$且：

1. $W(t) \perp\!\!\!\perp T|G$
2. $\mathrm{P}(W(t) \geq W(t-1)|G) = 1$或者$\mathrm{P}(W(t) \leq W(t-1)|G) = 1$

以上假定处理变量在每个组内，处理变量$W(t)$对每一期都是稳定的，且每个组别内部，相邻两期的处理要么单调递增，要么单调递减。

- 如果每个g, t内部只有一个观测，以上假设自动满足
- 如果每个g, t内部有不止一个观测，以上假设不一定满足了
  - 对于面板数据，$W(t)$可以直接看到
  - 对于重复截面数据，以上假设就不可检验了

重要假定

此外，还要额外假定处理效应的稳定性：

假设4：稳定的处理效应

对于所有的$(g,t) \in \{0,1,2,...,\bar{G}\} \times \{0,1,2,...,\bar{T}\}$：

$$
\begin{aligned}
&\mathbb{E}\left(Y(1)-Y(0) \,|G=g, T=t, W(t-1)=1\right) \\
&= \mathbb{E}\left(Y(1)-Y(0) \,|G=g, T=t-1, W(t-1)=1\right)
\end{aligned}
$$

关心的处理效应

在以下介绍中，我们主要关心两种处理效应：

- 处理组平均处理效应：

$$\Delta^{TR} = \mathbb{E}\left(Y\left(1\right) - Y\left(0\right)|W = 1\right)$$

- LATE：

$$\Delta^{S} = \mathbb{E}\left(Y\left(1\right) - Y\left(0\right)|S\right)$$

其中$S = \{W\left(t-1\right) \neq W\left(t\right)\}$，即相邻两期处理状态不相等的转换者。

额外记：

- $\Delta_{gt}^{TR} = \mathbb{E}\left(Y\left(1\right) - Y\left(0\right)|W = 1, G = g, T = t\right)$
- $\Delta_{gt}^{S} = \mathbb{E}\left(Y\left(1\right) - Y\left(0\right)|S, G = g, T = t\right)$

两组、两期的特例

如果我们有两组、两期，即$\bar{G}=\bar{T}=1$，当我们加入组别G的固定效应以及时间固定效应时，那么：

$$
\begin{aligned}
\beta_{fe}=\beta_{fd} &= \frac{\left[\mathbb{E}\left(Y_{11}\right)-\mathbb{E}\left(Y_{10}\right)\right]-\left[\mathbb{E}\left(Y_{01}\right)-\mathbb{E}\left(Y_{01}\right)\right]}{\left[\mathbb{E}\left(W_{11}\right)-\mathbb{E}\left(W_{10}\right)\right]-\left[\mathbb{E}\left(W_{01}\right)-\mathbb{E}\left(W_{01}\right)\right]} \\
&= \frac{\mathrm{DID}_Y}{\mathrm{DID}_W}
\end{aligned}
$$

两组、两期的特例

由于：

$$\begin{aligned}\mathbb{E}\left(Y_{11}\right) &= \mathbb{E}\left(Y_{11}\left(0\right)\right) + \mathbb{E}\left(Y_{11}\left(W_{11}\right) - Y_{11}\left(0\right)\right) \\ &= \mathrm{P}\left(W_{11}=1\right)\left[\mathbb{E}\left(Y_{11}\left(0\right)\right) + \Delta_{gt}^{TR}\right] \\ &+ \left(1-\mathrm{P}\left(W_{11}=1\right)\right)\mathbb{E}\left(Y_{11}\left(0\right)\right) \\ &= \mathrm{P}\left(W_{11}=1\right)\Delta_{11}^{TR} + \mathbb{E}\left(Y_{11}\left(0\right)\right)\end{aligned}$$

从而：

$$\begin{aligned}\mathbb{E}\left(Y_{11}\right) - \mathbb{E}\left(Y_{10}\right) &= \mathrm{P}\left(W_{11}=1\right)\Delta_{11}^{TR} + \mathbb{E}\left(Y_{11}\left(0\right)\right) \\ &- \mathrm{P}\left(W_{10}=1\right)\Delta_{10}^{TR} - \mathbb{E}\left(Y_{10}\left(0\right)\right) \\ &= \mathrm{P}\left(W_{11}=1\right)\Delta_{11}^{TR} - \mathrm{P}\left(W_{10}=1\right)\Delta_{10}^{TR} \\ &+ \mathbb{E}\left(Y_{11}\left(0\right)\right) - \mathbb{E}\left(Y_{10}\left(0\right)\right)\end{aligned}$$

两组、两期的特例

同理:

$$\begin{aligned}\mathbb{E}\left(Y_{01}\right)-\mathbb{E}\left(Y_{00}\right)=&\mathrm{P}\left(W_{01}=1\right)\Delta_{01}^{TR}-\mathrm{P}\left(W_{00}=1\right)\Delta_{00}^{TR}\\&+\mathbb{E}\left(Y_{01}(0)\right)-\mathbb{E}\left(Y_{00}(0)\right)\end{aligned}$$

根据共同趋势假设（假设2）：

$$\begin{aligned}&\left[\mathbb{E}\left(Y_{11}\right)-\mathbb{E}\left(Y_{10}\right)\right]-\left[\mathbb{E}\left(Y_{01}\right)-\mathbb{E}\left(Y_{01}\right)\right]\\=&\left[\mathrm{P}\left(W_{11}=1\right)\Delta_{11}^{TR}-\mathrm{P}\left(W_{10}=1\right)\Delta_{10}^{TR}\right]-\left[\mathrm{P}\left(W_{01}=1\right)\Delta_{01}^{TR}-\mathrm{P}\left(W_{00}=1\right)\Delta_{00}^{TR}\right]\end{aligned}$$

两组、两期的特例

从而

$$
\begin{aligned}
\beta_{fe} = \beta_{fd} &= \frac{[\mathbb{E}(Y_{11}) - \mathbb{E}(Y_{10})] - [\mathbb{E}(Y_{01}) - \mathbb{E}(Y_{01})]}{[\mathbb{E}(W_{11}) - \mathbb{E}(W_{10})] - [\mathbb{E}(W_{01}) - \mathbb{E}(W_{01})]} \\
&= \frac{[\mathrm{P}(W_{11}=1)\Delta_{11}^{TR} - \mathrm{P}(W_{10}=1)\Delta_{10}^{TR}] - [\mathrm{P}(W_{01}=1)\Delta_{01}^{TR} - \mathrm{P}(W_{00}=1)\Delta_{00}^{TR}]}{[\mathrm{P}(W_{11}=1) - \mathrm{P}(W_{10}=1)] - [\mathrm{P}(W_{01}=1) - \mathrm{P}(W_{00}=1)]} \\
&= \frac{\mathrm{P}(W_{11}=1)}{\mathrm{DID}_W}\Delta_{11}^{TR} - \frac{\mathrm{P}(W_{10}=1)}{\mathrm{DID}_W}\Delta_{10}^{TR} - \frac{\mathrm{P}(W_{01}=1)}{\mathrm{DID}_W}\Delta_{01}^{TR} + \frac{\mathrm{P}(W_{00}=1)}{\mathrm{DID}_W}\Delta_{00}^{TR}
\end{aligned}
$$

从而固定效应或者一阶差分的系数是四个处理组处理效应的一个加权平均:

- 如果处理效应同质性，即$\Delta_{11}^{TR} = \Delta_{10}^{TR} = \Delta_{01}^{TR} = \Delta_{00}^{TR}$，那么可以识别处理组处理效应
- 如果异质性处理效应，**注意有的权重的负的**！可能不能识别处理组处理效应
- 在DID设计中，$W = G \times T$，从而$\mathrm{P}(W_{10}=1) = \mathrm{P}(W_{01}=1) = \mathrm{P}(W_{00}=1) = 0$，从而可以识别出$\Delta_{11}^{TR}$。

两组、两期的特例

如果考虑LATE，额外使用假设3和假设4，有：

$$\beta_{fe} = \beta_{fd} = \frac{\mathrm{P}(W_{11}=1) - \mathrm{P}(W_{10}=1)}{\mathrm{DID}_W}\Delta_{11}^S - \frac{\mathrm{P}(W_{01}=1) - \mathrm{P}(W_{00}=1)}{\mathrm{DID}_W}\Delta_{01}^S$$

也是一个加权平均：

- 再次，权重有可能为负
- Fuzzy DID通过假设$\mathrm{P}(W_{01}=1) - \mathrm{P}(W_{00}=1) = 0$得到了识别。

一个更加一般的结果

如果有多组、多期，那么可以证明，对于$k \in \{fe, fd\}$，即双向固定效应和一阶差分的估计结果可以写为：

- 如果关注处理组处理效应，如果假设1和假设2成立，那么：

$$\beta_k = \mathbb{E}\left(v_{k,GT}^{TR}\Delta_{GT}^{TR}|W=1\right)$$

- 如果关注LATE，并且假设1、2、3、4成立，那么：

$$\beta_k = \mathbb{E}\left(v_{k,GT}^{S}\Delta_{GT}^{S}|S\right)$$

值得注意的是，**权重$v_{k,GT}^{TR}, v_{k,GT}^{S}$都有可能是负的**！

- 在异质性处理效应的条件下，识别可能失败。

何时可以成功识别?

一个比较简单的结果是，如果：

- （Staggered adoption）$W = \mathbb{1}\{T \geq t_G\}$，即每个组都是在前面一段时间不被处理，在时间$t_G$之后被处理

结合假设1，可以证明P$\left(v_{fd,GT}^S \geq 0|S\right) = 1$，即如果使用一阶差分识别LATE，那么所有权重都大于0。
如果除Staggered adoption之外额外假设：

- 每组每期只有一个观测

那么结合假设1，有：P$\left(v_{fe,GT}^S \geq 0|S\right) = 1$，即使用固定效应识别LATE的权重都是大于0的。

权重的计算

为了诊断权重有没有可能为负，或者权重为负会不会导致（符号）识别的失败，作者同时推导了能够使得所求得的$\beta_k(k = fd, fe)$与Δ^{TR}, Δ^S符号相反的最小的$\Delta^{TR}_{gt}, \Delta^S_{gt}$的异质性程度度量。令

$$\sigma^{TR} = \sqrt{\mathbb{V}\left(\Delta^{TR}_{gt} | W = 1\right)}, \sigma^S = \sqrt{\mathbb{V}\left(\Delta^S_{gt} | S\right)}$$

那么其下界为：

- 如果假设2成立，为了获得处理组处理效应，下界为：

$$\underline{\sigma}^{TR}_k = \frac{|\beta_k|}{\sqrt{\mathbb{V}\left(v^{TR}_{k,GT} | D = 1\right)}}$$

- 如果假设2、3、4成立，为了获得LATE，下界为：

$$\underline{\sigma}^S_k = \frac{|\beta_k|}{\sqrt{\mathbb{V}\left(v^S_{k,GT} | S = 1\right)}}$$

异质性下界的使用

- 如果计算得到的$\underline{\sigma}_k^{TR}, \underline{\sigma}_k^{S}$很小，意味着模型对于异质性处理效应的稳健性较低，很容易出现错误的结果
- 可以用来查看模型对于异质性处理效应的稳健性
- 也可以用来比较固定效应与一阶差分，在两个模型中进行选择。
- Stata：twowayfeweights