2017「应用微观计量经济学」暑期学校

政策评估与处理效应

司继春

上海对外经贸大学

2017年8月

Rubin因果模型

思想来源：实验

1. Neyman (1923) 提出了潜在结果（potential outcomes）
2. Fisher (1925)：推断

Rubin (1974)：在观察数据（observational study）中使用潜在因果的语言。

Rubin因果模型

- 对于同一个个体，不同的处理水平对应着不同的潜在结果（potential outcomes）
- 研究者只能观察到对应于实际接收的处理水平的结果。

例如，对于$i = 1, ..., N$个个体选择是否参与一个培训项目，有两种处理水平：

1. 参与项目，记$W_i = 1$
2. 不参与项目，记$W_i = 0$

相应的，潜在结果$Y_i(W_i)$为：

1. $Y_i(0)$：如果不参与的结果（如工资水平）
2. $Y_i(1)$：如果参与的结果

Rubin因果模型

由于个体i只能选择参与或者不参与，因而只能观察到$Y_i(0)$和$Y_i(1)$其中的一个。记观察到的结果为Y_i，那么：

$$
\begin{aligned}
Y_i &= W_i Y_i(1) + (1 - W_i) Y_i(0) \\
&= \begin{cases} Y_i(0) & W_i = 0 \\ Y_i(1) & W_i = 1 \end{cases}
\end{aligned}
$$

模型设定优点：

1. 允许异质性（相比较于一般计量模型中的结构方程）
2. 在设定分配机制（assignment mechanism）之前即定义了因果（在潜在结果方程中没有内生性等）
3. 将潜在结果的建模与分配机制建模分开

因而分析分配机制非常重要！！！

Rubin因果模型

分配机制：

1. 随机实验：
   1.1 处理的分配概率不随着潜在结果而改变
   1.2 处理的分配概率为协变量（covariates）的已知函数
2. 无混淆分配（unconfounded assignment）
   - 要求给定协变量，潜在结果与分配独立：

$$W_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) | X_i$$

   - 与随机实验差别：$P(W_i|X_i)$为未知函数
   - 又称为：
     - selection-on-observable
     - exogeneity
     - conditional independence assumption (CIA)

- 其他
  - selection-on-unobservable
  - 例：Roy Model：$W_i = 1(Y_i(1) \geq Y_i(0))$

Rubin因果模型

关键假设：**SUTVA（Stable Unit Treatment Value Assumption）**：

> 任何个体的潜在结果不会随着其他个体的分配而改变，
> 且对于每个个体，没有其他不同形式的、可以导致不同
> 潜在结果的处理水平。

含义：排除了：

1. 一般均衡效应（general equilibrium effects）
2. 伙伴效应（peer effects）
3. ...

如果违背：

1. 重新定义研究个体
2. 直接设定个体之间的交互

Rubin因果模型

不同形式的处理效应：

- 个体处理效应：对于个体i，个体处理效应为$\Delta_i = Y_i(1) - Y_i(0)$

异质性（heterogeneous effects）：Δ_i随着i的变化而变化。

- 同质处理效应（Homogeneous treatment effects）：$Y_i(1) - Y_i(0) = \Delta$
  - 例：$Y_i = g(X_i) + \alpha W_i + u_i$
- 条件同质处理效应：$\Delta_i = \Delta(X_i)$
  - 例：$Y_i = g(X_i, D_i) + u_i \Rightarrow \Delta_i = g(X_i, 1) - g(X_i, 0) = \Delta(X_i)$
  - 例：$Y_i = X_i'\beta + W_i X_i'\alpha + u_i$
- 异质处理效应
  - $Y_i = g(X_i, D_i, u_i)$

Rubin因果模型

感兴趣的处理效应:

1. 平均处理效应（Average Treatment Effects, ATE）：$\mathrm{ATE} = \mathbb{E}\left(\Delta_\mathrm{i}\right)$
2. 处理组平均处理效应（Average Treatment Effects on the Treated, TT）：$\mathrm{ATT} = \mathbb{E}\left(\Delta_\mathrm{i} | \mathrm{W}_\mathrm{i} = 1\right)$
3. 未处理组平均处理效应（Average Treatment Effects on the Untreated, TUT）：$\mathrm{ATUT} = \mathbb{E}\left(\Delta_\mathrm{i} | \mathrm{W}_\mathrm{i} = 0\right)$

Rubin因果模型

以上定义的是总体处理效应，然而给定样本，我们通常关注给定协变量X_i时的处理效应：

1. $\text{CATE}(X_i) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}(\Delta_i | X_i)$
2. $\text{CATT}(X_i) = \frac{1}{N}\sum_{i|W_i=1} \mathbb{E}(\Delta_i | X_i)$
3. $\text{CATUT}(X_i) = \frac{1}{N}\sum_{i|W_i=0} \mathbb{E}(\Delta_i | X_i)$

Rubin因果模型

几种处理效应关系：

- 同质处理效应：

$$\mathrm{ATE} = \mathrm{ATT} = \mathrm{ATUT} = \mathrm{CATE}\left(\mathrm{X_i}\right) = \mathrm{CATT}\left(\mathrm{X_i}\right) = \mathrm{CATUT}\left(\mathrm{X_i}\right)$$

- 条件同质处理效应：
  - $\mathrm{CATE}\left(\mathrm{X_i}\right) = \mathrm{CATT}\left(\mathrm{X_i}\right) = \mathrm{CATUT}\left(\mathrm{X_i}\right)$
  - 可能$\mathrm{ATE} \neq \mathrm{ATT} \neq \mathrm{ATUT}$，由于$\mathrm{X_i}$的分布随$\mathrm{W_i}$不同
- 异质性处理效应：
  - 如果$\mathrm{Y_i}(1) - \mathrm{Y_i}(0) \perp\!\!\!\perp \mathrm{W_i}|\mathrm{X_i}$：结论同条件同质处理效应
  - 否则$\mathrm{CATE}\left(\mathrm{X_i}\right) \neq \mathrm{CATT}\left(\mathrm{X_i}\right) \neq \mathrm{CATUT}\left(\mathrm{X_i}\right)$

Rubin因果模型

其他感兴趣的处理效应：

- 分位数处理效应（quantile treament effects）：

$$\tau_q = F^{-1}_{Y(1)}(q) - F^{-1}_{Y(0)}(q)$$

- 处理效应的分位数：

$$\tilde{\tau}_q = F^{-1}_{Y(1)-Y(0)}(q)$$

- 最难的参数：$(Y_i(1), Y_i(0))$的联合分布：

$$P(Y_i(1) \leq y_1, Y_i(0) \leq y_0)$$

Rubin因果模型

可能的偏误：由于$Y = WY(1) + (1 - W)Y(0) = Y(0) + W(Y(1) - Y(0))$，从而：

$$
\begin{aligned}
\mathbb{E}(Y|W=1) - \mathbb{E}(Y|W=0) = & \\
(\text{ATT}) \quad & \mathbb{E}(Y_1 - Y_0|W=1) + \\
(\text{Selection bias}) \quad & \mathbb{E}(Y_0|W=1) - \mathbb{E}(Y_0|W=0)
\end{aligned}
$$

如果我们关心平均处理效应：

$$
\begin{aligned}
\mathbb{E}(Y|W=1) - \mathbb{E}(Y|W=0) = & \\
(\text{ATE}) \quad & \mathbb{E}(Y_1 - Y_0) + \\
(\text{Sorting Gain}) \quad & \mathbb{E}(Y_1 - Y_0|W=1) - \mathbb{E}(Y_1 - Y_0) + \\
(\text{Selection bias}) \quad & \mathbb{E}(Y_0|W=1) - \mathbb{E}(Y_0|W=0)
\end{aligned}
$$

偏识别

在最宽松的假设条件下，我们可以放弃点识别（point identification），转而使用偏识别（Partial identification）

- 点识别识别具体的点
- 偏识别仅仅能够给出上下界

偏识别

如果潜在因果$\mathrm{Y_i}(\mathrm{W_i})$是有界的，比如，$\mathrm{Y_i}(\mathrm{W_i}) = 0/1$，由于:

$$\begin{aligned}
\mathbb{E}\left(\mathrm{Y_i}(1) - \mathrm{Y_i}(0)\right) = & \mathbb{E}\left(\mathrm{Y_i}(1) \mid \mathrm{W_i} = 1\right) \mathrm{P}\left(\mathrm{W_i} = 1\right) \\
& + \mathbb{E}\left(\mathrm{Y_i}(1) \mid \mathrm{W_i} = 0\right) \mathrm{P}\left(\mathrm{W_i} = 0\right) \\
& - \mathbb{E}\left(\mathrm{Y_i}(0) \mid \mathrm{W_i} = 1\right) \mathrm{P}\left(\mathrm{W_i} = 1\right) \\
& - \mathbb{E}\left(\mathrm{Y_i}(0) \mid \mathrm{W_i} = 0\right) \mathrm{P}\left(\mathrm{W_i} = 0\right)
\end{aligned}$$

因而其下界为:

$$\begin{aligned}
\tau_{\mathrm{l}} = & \mathbb{E}\left(\mathrm{Y_i}(1) \mid \mathrm{W_i} = 1\right) \mathrm{P}\left(\mathrm{W_i} = 1\right) \\
& - \mathrm{P}\left(\mathrm{W_i} = 1\right) - \mathbb{E}\left(\mathrm{Y_i}(0) \mid \mathrm{W_i} = 0\right) \mathrm{P}\left(\mathrm{W_i} = 0\right)
\end{aligned}$$

上界为:

$$\begin{aligned}
\tau_{\mathrm{u}} = & \mathbb{E}\left(\mathrm{Y_i}(1) \mid \mathrm{W_i} = 1\right) \mathrm{P}\left(\mathrm{W_i} = 1\right) \\
& + \mathrm{P}\left(\mathrm{W_i} = 0\right) - \mathbb{E}\left(\mathrm{Y_i}(0) \mid \mathrm{W_i} = 0\right) \mathrm{P}\left(\mathrm{W_i} = 0\right)
\end{aligned}$$

偏识别

如果我们关心潜在结果的分布，由于:

$$\begin{aligned} F_{Y_1}(y) &= P\left(Y_i(1) \le y\right) \\ &= P\left(Y_i(1) \le y | W_i = 1\right) P\left(W_i = 1\right) \\ &+ P\left(Y_i(1) \le y | W_i = 0\right) P\left(W_i = 0\right) \end{aligned}$$

因而：$F_{Y_1}(y)$的下界为:

$$P\left(Y_i(1) \le y | W_i = 1\right) P\left(W_i = 1\right)$$

而上界为:

$$P\left(Y_i(1) \le y | W_i = 1\right) P\left(W_i = 1\right) + P\left(W_i = 0\right)$$

偏识别

其他如：

1. Manski and Pepper(2000)
2. Jun, Lee and Shin (2016)

Matching

在Unconfoundedness假设条件下，可以方便的得到处理效应的识别。关键假设：

1. Unconfoundedness假设（CIA假设）：

$$W_i \amalg (Y_i(1), Y_i(0)) | X$$

2. 共同支撑假设（Common support assumption, CSA）：

$$0 < P(W_i = 1 | X_i) < 1$$

Matching

CIA意味着均值独立，即：

$$\mathbb{E}\left(Y_i\left(1\right)\left|X_i, W_i\right.\right) = \mathbb{E}\left(Y_i\left(1\right)\left|X_i\right.\right)$$

$$\mathbb{E}\left(Y_i\left(0\right)\left|X_i, W_i\right.\right) = \mathbb{E}\left(Y_i\left(0\right)\left|X_i\right.\right)$$

而CSA需要对于相同的X_i，都有处理组和非处理组：经常需要trimming。

Matching

由于:

$$\begin{aligned}\mathbb{E}\left(Y_i|W_i=1,X_i\right)-\mathbb{E}\left(Y_i|W_i=0,X_i\right)\\ =&\mathbb{E}\left(Y_i\left(1\right)|X_i,W_i=1\right)-\mathbb{E}\left(Y_i\left(0\right)|X_i,W_i=0\right)\\ =&\mathbb{E}\left(Y_i\left(1\right)|X_i\right)-\mathbb{E}\left(Y_i\left(0\right)|X_i\right)\\ =&\mathbb{E}\left(Y_i\left(1\right)-Y_i\left(0\right)|X_i\right)\end{aligned}$$

因而

$$\mathbb{E}\left[\mathbb{E}\left(Y_i|W_i=1,X_i\right)-\mathbb{E}\left(Y_i|W_i=0,X_i\right)\right]=\mathbb{E}\left(Y_i\left(1\right)-Y_i\left(0\right)\right)$$

同理:

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left(Y_i|W_i=1,X_i\right)-\mathbb{E}\left(Y_i|W_i=0,X_i\right)|W_i=1\right]=\\ \mathbb{E}\left(Y_i\left(1\right)-Y_i\left(0\right)|W_i=1\right)\end{aligned}$$

Matching

方法一：回归

设定

$$\mathbb{E}\left(Y_i|W_i=1,X_i\right)=\mathbb{E}\left(Y_i\left(1\right)|X_i\right)=\mu_1\left(X_i\right)$$

$$\mathbb{E}\left(Y_i|W_i=0,X_i\right)=\mathbb{E}\left(Y_i\left(0\right)|X_i\right)=\mu_0\left(X_i\right)$$

平均处理效应：

$$\hat{\tau}_{reg}=\frac{1}{N}\sum_{i=1}^{N}\left[\hat{\mu}_1\left(X_i\right)-\hat{\mu}_0\left(X_i\right)\right]$$

1. 简单线性回归：

$$Y_i=\alpha+X_i'\beta+\tau\cdot W_i+\epsilon_i$$

   或者：

$$Y_i=\alpha+X_i'\beta+\tau\cdot W_i+W_i\left(X_i-\bar{X}\right)'\delta+\epsilon_i$$

2. 非参数回归

Matching

或者：Nearest Neighbor Matching：

1. 给定一个正的常数M，比如$M = 1$
2. 令$d(\cdot,\cdot)$为一个距离函数，比如欧几里得距离：

$$d(X_i, X_j) = (X_i - X_j)'(X_i - X_j)$$

或者Mahalanobis距离：

$$d(X_i, X_j) = (X_i - X_j)' \Sigma_X^{-1} (X_i - X_j)$$

Matching

Nearest-neighbor matching: 对于任意处理组的i，从控制组中找到最近的M个控制组个体，记:

$$J_M(i) = \{l_1(i), \ldots, l_M(i)\}$$

定义

$$\hat{Y}_i(0) = \frac{1}{M} \sum_{m \in J_M(i)} Y_m$$

可以使用:

$$\frac{1}{N_1} \sum_{i|W_i=1} \left[Y_i(1) - \hat{Y}_i(0) \right]$$

Matching

实践中，有不同的匹配方案:

1. 选择M，一般而言如果控制组数量远远大于实验组数量，可以使用较多的M
2. 序贯/非序贯
3. 贪婪/非贪婪
4. 放回/无放回
5. 先进行分组，组内进行匹配
6. 使用propensity score进行排序，进而匹配

Matching

倾向得分匹配的步骤：

1. 定义距离
2. 匹配
3. 评估匹配结果：
   3.1 balancing: t-test、Standardised Bias
   3.2 unconfoundedness：使用明显无效应的其他的Y
4. 评估政策效应
5. 敏感性分析

Matching

其他匹配方法:

1. Kernel matching
2. Radius matching
3. Stratification or interval matching
4. Propensity score matching

Matching

倾向得分（Propensity score）方法：Rosenbaum and Rubin(1983)
证明：

$$W_i \perp\!\!\!\perp \left(Y_i\left(1\right), Y_i\left(0\right)\right)|X_i \Longleftrightarrow W_i \perp\!\!\!\perp \left(Y_i\left(1\right), Y_i\left(0\right)\right)|P\left(X_i\right)$$

因而控制倾向得分匹配就足够了。

Matching

Rosenbaum and Rubin(1983)提出了三阶段的方法:

1. 估计倾向得分:
$$P(W_i|X_i)$$
2. 用Y_i对W_i和P_i做回归，得到$\hat{\mathbb{E}}(Y_i|W_i, P(X_i))$
3. 估计ATT:
$$\frac{1}{N_1}\sum_{i|W_i=1}\left[Y_i(1)-\hat{Y}_i(0)\right]$$

注意在使用Propensity Score时，一定要注意Common support假设：trimming！

Matching

其他方法:

1. blocking/subclassification/stratification
2. nearest-neighbor matching
3. kernel matching
4. ...

Matching

注意到，由于：

$$
\begin{aligned}
\mathbb{E}\left(\frac{W_iY_i}{P(X_i)}\right) &= \mathbb{E}\left(\frac{W_iY_i(1)}{P(X_i)}\right) \\
&= \mathbb{E}\left[\mathbb{E}\left(\frac{W_iY_i(1)}{P(X_i)}|X_i\right)\right] \\
&= \mathbb{E}\left[\frac{\mathbb{E}(W_iY_i(1)|X_i)}{P(X_i)}\right] \\
&= \mathbb{E}\left[\frac{\mathbb{E}(W_i|X_i)\,\mathbb{E}(Y_i(1)|X_i)}{P(X_i)}\right] \\
&= \mathbb{E}\left[\mathbb{E}(Y_i(1)|X_i)\right] = \mathbb{E}(Y_i(1))
\end{aligned}
$$

同理：

$$
\mathbb{E}\left(\frac{(1-W_i)Y_i}{1-P(X_i)}\right) = \mathbb{E}(Y_i(0))
$$

Matching

因而平均处理效应：

$$\tau_{\mathrm{ATE}} = \mathbb{E}\left[\frac{\mathrm{W_i Y_i}}{\mathrm{P}\left(\mathrm{X_i}\right)} - \frac{\left(1-\mathrm{W_i}\right)\mathrm{Y_i}}{1-\mathrm{P}\left(\mathrm{X_i}\right)}\right]$$

可以使用：

$$\hat{\tau}_{\mathrm{ATE}} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\left[\frac{\mathrm{W_i Y_i}}{\mathrm{P}\left(\mathrm{X_i}\right)} - \frac{\left(1-\mathrm{W_i}\right)\mathrm{Y_i}}{1-\mathrm{P}\left(\mathrm{X_i}\right)}\right]$$

进行估计，称为Inverse Propensity Weighting(IPW)。其中$\mathrm{P}\left(\mathrm{X_i}\right)$可以使用Logistic sieve估计量（Hirano, Imbense and Ridder, 2003）。

Matching

然而IPW方法对倾向得分非常敏感。可以考虑使用Robins等人提出的双向稳健（Double robustness）方法：

1. 结合了回归方法和IPW
2. 只需要$\mathrm{P}(\mathrm{X_i})$或者结果方程至少有一个设定正确（双向稳健）

最小化：

$$\min_{\alpha_0,\beta_0} \sum_{i|W_i=0} \frac{\left[Y_i - \alpha_0 - \beta_0'\left(X_i - \bar{X}_i\right)\right]}{P\left(X_i;\hat{\gamma}\right)}$$

$$\min_{\alpha_1,\beta_1} \sum_{i|W_i=1} \frac{\left[Y_i - \alpha_1 - \beta_1'\left(X_i - \bar{X}_i\right)\right]}{1 - P\left(X_i;\hat{\gamma}\right)}$$

平均处理效应为：

$$\hat{\tau}_{\mathrm{ATE}} = \hat{\alpha}_1 - \hat{\alpha}_0$$

Matching

其他方法：Imai and Ratkovic (2014)：解：

$$\frac{1}{N}\sum_{i=1}^{N} g_{\gamma}\left(W_i, X_i\right) = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{W_i}{P\left(X_i;\gamma\right)} - \frac{1-W_i}{1-P\left(X_i;\gamma\right)}\right) f\left(X_i\right)\right] = 0$$

仍然是双向稳健的。此外，Fan et al. (2016)做了推广。

DID

违背Unconfoundedness假设，但是有面板数据的条件下：双重差分模型（Difference-in-differences）。

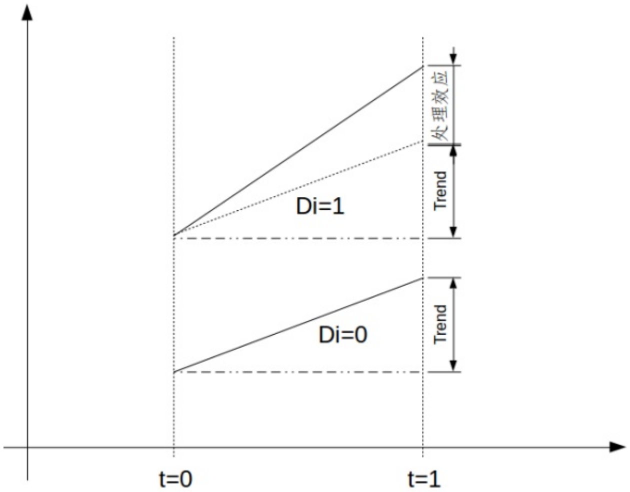
一般设定：

1. 分组变量：$G_{ig} = 1$：处理组；$G_{ig} = 0$：实验组
2. 时间变量：$d_t = 0/1$
3. 处理$W_{it} = d_t \cdot G_i$
4. Outcome：Y_{igt}

	$G_{ig} = 0$	$G_{ig} = 1$
$d_t = 0$	0	0
$d_t = 1$	0	1

DID

关键假设：共同趋势（Common trend）



DID

双重差分模型:

1. 第一次差分：对于两个不同的分组，分别计算:

$$\Delta_1 = \mathbb{E}\left(\mathrm{Y}_{\mathrm{igt}} | \mathrm{G}_{\mathrm{gi}} = 1, \mathrm{d}_{\mathrm{t}} = 1\right) - \mathbb{E}\left(\mathrm{Y}_{\mathrm{igt}} | \mathrm{G}_{\mathrm{gi}} = 1, \mathrm{d}_{\mathrm{t}} = 0\right)$$

$$\Delta_0 = \mathbb{E}\left(\mathrm{Y}_{\mathrm{igt}} | \mathrm{G}_{\mathrm{gi}} = 0, \mathrm{d}_{\mathrm{t}} = 1\right) - \mathbb{E}\left(\mathrm{Y}_{\mathrm{igt}} | \mathrm{G}_{\mathrm{gi}} = 0, \mathrm{d}_{\mathrm{t}} = 0\right)$$

2. 第二次差分：处理效应:

$$\tau = \Delta_1 - \Delta_0$$

DID

实践中，等价于使用回归：

$$\mathrm{Y}_{\mathrm{igt}} = \mathrm{c} + \lambda \cdot \mathrm{d}_{\mathrm{t}} + \gamma \cdot \mathrm{G}_{\mathrm{ig}} + \beta \cdot \mathrm{d}_{\mathrm{t}} \cdot \mathrm{G}_{\mathrm{ig}} + \mathrm{u}_{\mathrm{it}}$$

即第一次差分：

$$\Delta_1 = \mathbb{E}\left(\mathrm{Y}_{\mathrm{igt}} | \mathrm{G}_{\mathrm{gi}} = 1, \mathrm{d}_{\mathrm{t}} = 1\right) - \mathbb{E}\left(\mathrm{Y}_{\mathrm{igt}} | \mathrm{G}_{\mathrm{gi}} = 1, \mathrm{d}_{\mathrm{t}} = 0\right) = \lambda + \beta$$

$$\Delta_0 = \mathbb{E}\left(\mathrm{Y}_{\mathrm{igt}} | \mathrm{G}_{\mathrm{gi}} = 0, \mathrm{d}_{\mathrm{t}} = 1\right) - \mathbb{E}\left(\mathrm{Y}_{\mathrm{igt}} | \mathrm{G}_{\mathrm{gi}} = 0, \mathrm{d}_{\mathrm{t}} = 0\right) = \lambda$$

第二次差分：

$$\tau = \Delta_1 - \Delta_0 = \beta$$

DID

实践中的设定:

1. 加入控制，即可以假定给定控制x_{it}的情况下，共同趋势假设成立，使用回归:

$$Y_{igt} = c + \lambda \cdot d_t + \gamma \cdot G_{ig} + \beta \cdot d_t \cdot G_{ig} + x_{it}'\beta + u_{it}$$

2. 直接控制个体和时间效应，而不是分组变量:

$$Y_{igt} = c_i + \lambda_t + \beta \cdot d_t \cdot G_{ig} + x_{it}'\beta + u_{it}$$

3. 直接加入个体趋势:

$$Y_{igt} = c_i + \lambda_t + \gamma_{1i} t + \gamma_{2i} t^2 + \beta \cdot d_t \cdot G_{ig} + x_{it}'\beta + u_{it}$$

4. 交互固定效应（Bai, 2009）

$$Y_{igt} = c_i'\lambda_t + \beta \cdot d_t \cdot G_{ig} + x_{it}'\beta + u_{it}$$

DID

其他设定问题:

1. 多期DID：将$d_t \cdot G_{ig}$替换为d_{it}，其中d_{it}指第t期第i个个体是否接受处理。
2. 标准误的问题：（广义的）自相关严重影响推断，cluster是必须的

DID

检验：

1. Placebo test：提前冲击发生的时间
2. 设定：

$$Y_{igt} = c_i + \lambda_t + \sum_{\tau=1}^{T} \beta_\tau \cdot 1\{t = \tau\} \cdot G_{ig} + x_{it}'\beta + u_{it}$$

2.1 冲击发生之前不显著
2.2 冲击发生之后显著

DID

其他方法:

1. Triple differences: e.g.，某个地区g的某个特定组别r收到处理：加入两两交互项，三个变量的交互项的系数为处理效应。
2. DID+Matching
3. 非线性DID：Changes-in-Changes

DID

非线性DID：CIC（Athey and Imbens, 2006）
假设：

1. $Y_i(0) = h(U_i, T_i)$，其中$h(\cdot, \cdot)$对U_i为单调的函数，T_i为时间
2. $U \amalg T|G$，即给定分组，U的分布与时间独立
3. support假设

关键结论：

$$F_{Y^N,11}(y) = F_{Y,10}\left(F^{-1}_{Y,00}\left(F_{Y,01}(y)\right)\right)$$

其中$F_{Y,gt}$为分布函数。进而处理效应：

$$\tau_{CIC} = \mathbb{E}(Y_{11}) - \mathbb{E}\left(F^{-1}_{01}\left(F_{00}\left(Y_{10}\right)\right)\right)$$

RDD

断点回归（Regression discontinuity）用于政策变量仅仅取决于一个连续变量x，且在某个点$x = c$处，参与概率有跳跃的情况。

1. Sharp RD：

$$W_i = 1\left(X_i \geq c\right)$$

2. Fuzzy RD

$$\lim_{x \downarrow c} P\left(W_i | X_i = c\right) \neq \lim_{x \uparrow c} P\left(W_i | X_i = c\right)$$

- 优点：识别干净
- 缺点：外部有效性

RDD

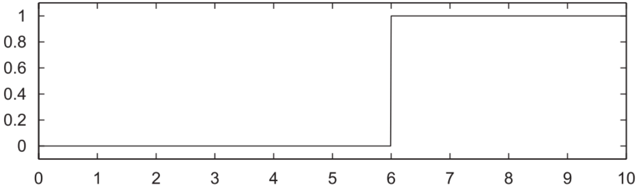


Fig. 1. Assignment probabilities (SRD).

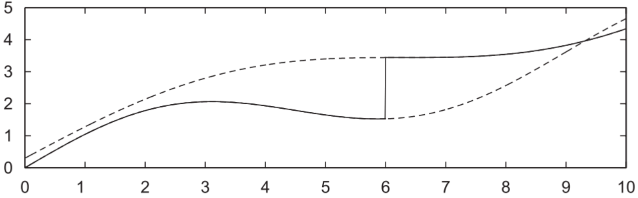


Fig. 2. Potential and observed outcome regression functions.

RDD

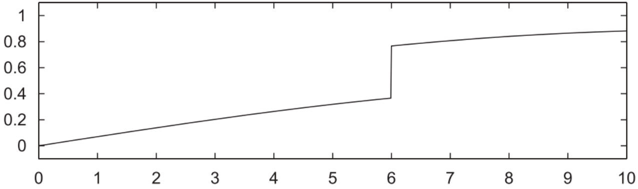


Fig. 3. Assignment probabilities (FRD).

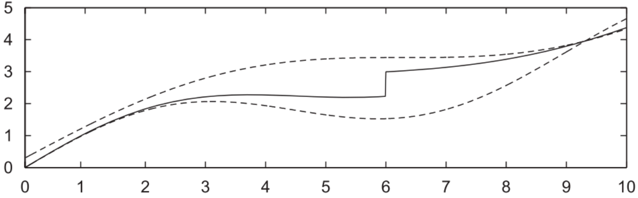


Fig. 4. Potential and observed outcome regression (FRD).

Sharp RD

政策效应：

$$\tau_{\mathrm{SRD}} = \lim_{x \downarrow c} \mathbb{E}\left(Y_i | X_i = c\right) - \lim_{x \uparrow c} \mathbb{E}\left(Y_i | X_i = c\right)$$

估计：

1. 参数方法
2. 非参数方法：
   2.1 在$X = c$两边取两个小邻域计算均值
   2.2 Local constant
   2.3 Local polynomial (Local linear)：使用多项式在$X = c$两边（$(c - h, c + h)$）对outcome进行拟合：

$$Y_i = \alpha + \tau W_i + 1\left\{X_i \geq c\right\} f_r\left(X_i - c\right) + 1\left\{X_i < c\right\} f_l\left(X_i - c\right) + u_i$$

Fuzzy RD

令$S_i = 1\{X_i \geq c\}$

- Intention-to-treat (ITT):

$$Y_i = \alpha_1 + \tau_{ITT} S_i + S_i \cdot f_r(X_i - c) + [1 - S_i] f_l(X_i - c) + u_i$$

- 政策概率:

$$W_i = \alpha_2 + \delta \cdot S_i + S_i \cdot g_r(X_i - c) + [1 - S_i] g_l(X_i - c) + v_i$$

- 结构方程:

$$Y_i = \alpha + \tau \cdot W_i + S_i \cdot h_r(X_i - c) + [1 - S_i] h_l(X_i - c) + \epsilon_i$$

Fuzzy RD

估计:

1. 根据以上可知:
$$\tau = \frac{\tau_{\mathrm{ITT}}}{\delta}$$
其中τ_{ITT}和δ都可以通过Local polynomial估计得到
2. 2SLS

问题:

1. 窗宽选取?
2. 多项式阶数

RDD

一般步骤:

1. 画图
2. 检验manipulation
3. 推断
4. 稳健性检验

RDD

检验manipulation：McCrary(2008)