局部平均处理效应

慧航

上海对外经贸大学

2025年9月

LATE

令:

$$\begin{aligned} y_i &= y_i(0) + w_i\left(y_i(1) - y_i(0)\right) \\ &= \mathbb{E}\left(y_i(0)\right) + w_i\mathbb{E}\left(y_i(1) - y_i(0)\right) + \zeta_i + \eta_i w_i \end{aligned}$$

其中$\zeta_i = y_i(0) - \mathbb{E}\left(y_i(0)\right)$，$\eta_i = y_i(1) - y_i(0) - \mathbb{E}\left(y_i(1) - y_i(0)\right)$。如果存在一个工具变量$z_i$，使得$\mathbb{C}\left(z_i, w_i\right) \neq 0$，且$z_i \perp\!\!\!\perp \left(y_i(0), y_i(1)\right)$，且$z = 0/1$那么IV估计:

$$\text{plim}\tau_{IV} = \frac{\mathbb{C}(y,z)}{\mathbb{C}(w,z)} = \frac{\mathbb{E}(y|z=1) - \mathbb{E}(y|z=0)}{P(w|z=1) - P(w|z=0)}$$

- 一个简单的例子：y_i为对数收入，w_i为是否上高中，z_i为5km范围内有没有高中

LATE

工具变量识别了什么?

1. 当存在同质的处理效应时，经典的IV估计了同质的处理效应
   - $y_i = y_i(0) + w_i(y_i(1) - y_i(0)) + \epsilon_i + \eta_i w_i = y_i(0) + w_i \tau_{ATE} + v_i$
   - 关键在η_i，如果$\eta_i = 0$，同质的处理效应
2. 当存在异质性的处理效应时，经典IV识别出的参数不可解释
3. 如果加入某些假设条件，经典IV的确可以识别出可以解释的参数

Imbens and Angrist(1994)建立了经典IV的识别条件以及经典IV识别的参数解释。

ITT

如果记$z_i = 0/1$为实际分组变量，w_i为实际被处理变量，我们记：

$$w_i\left(z_i\right) = z_i w_i\left(1\right) + \left(1 - z_i\right) w_i\left(0\right) = \begin{cases} w_i\left(1\right) & z_i = 1 \\ w_i\left(0\right) & z_i = 0 \end{cases}$$

可以看成是关于内生变量的反事实。两个变量将总体分为四类人：

		$w_i(0)$ = 0	$w_i(0)$ = 1
$w_i(1)$	0	never-taker	defier
	1	complier	always-taker

ITT

关键假设:

1. $z_i \perp\!\!\!\perp (y_i(1), y_i(0), w_i(1), w_i(0))$
2. $\mathrm{P}(w_i = 1|z_i)$取决于z_i

ITT

定义intention-to-treat (ITT)，即工具变量估计的分子：

$$\tau_{\text{ITT}} = \mathbb{E}\left(y_i | z_i = 1\right) - \mathbb{E}\left(y_i | z_i = 0\right)$$

将其分解：

$$\begin{aligned}
\tau_{\text{ITT}} &= \mathbb{E}\left(y_i | z_i = 1\right) - \mathbb{E}\left(y_i | z_i = 0\right) \\
&= \mathbb{E}\left(y_i\left(0\right) + w_i\left(y_i\left(1\right) - y_i\left(0\right)\right) | z_i = 1\right) \\
&- \mathbb{E}\left(y_i\left(0\right) + w_i\left(y_i\left(1\right) - y_i\left(0\right)\right) | z_i = 0\right) \\
&= \mathbb{E}\left(w_i\left(y_i\left(1\right) - y_i\left(0\right)\right) | z_i = 1\right) - \mathbb{E}\left(w_i\left(y_i\left(1\right) - y_i\left(0\right)\right) | z_i = 0\right) \\
&= \mathbb{E}\left(w_i\left(1\right)\left(y_i\left(1\right) - y_i\left(0\right)\right) | z_i = 1\right) - \mathbb{E}\left(w_i\left(0\right)\left(y_i\left(1\right) - y_i\left(0\right)\right) | z_i = 0\right) \\
&= \mathbb{E}\left(w_i\left(1\right)\left(y_i\left(1\right) - y_i\left(0\right)\right)\right) - \mathbb{E}\left(w_i\left(0\right)\left(y_i\left(1\right) - y_i\left(0\right)\right)\right) \\
&= \mathbb{E}\left(\left(w_i\left(1\right) - w_i\left(0\right)\right)\left(y_i\left(1\right) - y_i\left(0\right)\right)\right)
\end{aligned}$$

LATE

进一步使用全概率公式:

$$
\begin{aligned}
\tau_{\mathrm{ITT}} &= \mathbb{E}\left(\left(w_i\left(1\right)-w_i\left(0\right)\right)\left(y_i\left(1\right)-y_i\left(0\right)\right)\right) \\
&= \mathbb{E}\left(\left(y_i\left(1\right)-y_i\left(0\right)\right)|w_i\left(1\right)-w_i\left(0\right)=1\right)P\left(w_i\left(1\right)-w_i\left(0\right)=1\right) \\
&- \mathbb{E}\left(\left(y_i\left(1\right)-y_i\left(0\right)\right)|w_i\left(1\right)-w_i\left(0\right)=-1\right)P\left(w_i\left(1\right)-w_i\left(0\right)=-1\right) \\
&= \mathbb{E}\left(\left(y_i\left(1\right)-y_i\left(0\right)\right)|w_i\left(1\right)=1,w_i\left(0\right)=0\right)P\left(w_i\left(1\right)=1,w_i\left(0\right)=0\right) \\
&- \mathbb{E}\left(\left(y_i\left(1\right)-y_i\left(0\right)\right)|w_i\left(1\right)=0,w_i\left(0\right)=1\right)P\left(w_i\left(1\right)=0,w_i\left(0\right)=1\right)
\end{aligned}
$$

如果额外额外假设: $\mathrm{P}\left(w_i\left(1\right)=0,w_i\left(0\right)=1\right)=0$ (单调性)
那么:

$$
\tau_{\mathrm{ITT}} = \mathbb{E}\left(\left(t_i\left(1\right)-t_i\left(0\right)\right)|w_i\left(1\right)=1,w_i\left(0\right)=0\right)P\left(w_i\left(1\right)=1,w_i\left(0\right)=0\right)
$$

LATE

由于:

$$\mathrm{P}\left(w_i=1|z_i=1\right)-P\left(w_i=1|z_i=0\right)=\mathrm{P}\left(w_i\left(1\right)=1\right)-P\left(x_i\left(0\right)=1\right)$$
$$=\mathrm{P}\left(x_i\left(1\right)=1,w_i\left(0\right)=0\right)$$

进而，工具变量估计了:

$$\mathrm{LATE}=\tau_{\mathrm{IV}}=\frac{\tau_{\mathrm{ITT}}}{P\left(w=1|z=1\right)-P\left(w=1|z=0\right)}$$
$$=\mathbb{E}\left(\left(y_i\left(1\right)-y_i\left(0\right)\right)|w_i\left(1\right)=1,w_i\left(0\right)=0\right)$$

即如果假设$\mathrm{P}\left(w_i\left(1\right)=0,w_i\left(0\right)=1\right)=0$（不存在defier），那么工具变量识别了complier的平均处理效应，即**Local Average Treatment Effects, LATE**。

随机实验中的非遵从

- 理想情况下，个体实际是否被处理应该由随机分组完全决定
- 然而现实情况可能不一样：非遵从（noncompliance）
    - 有的被分到处理组可能（自己）选择不被处理（单边非遵从，one-sided noncompliance）
    - 有的被分到控制组可能自己选择被处理
    - 上面两种都有：双边非遵从（two-sided noncompliance）
- 除此之外，还有些实验是这样设计的：
    - 通过疫苗广告宣传打疫苗，疫苗广告是随机分组的，打不打疫苗是自己决定的

LATE实例：OHIE数据

OHIE实验

在美国，Medicaid是真对穷人的健康保险计划。在2008年时，俄勒冈州计划回复Medicaid中的OHP Standard计划。由于预计申请人数非常多，因而州政府推出了一个按照抽签分配名额的方法。个人一旦被抽中，整个家庭都可以享受该计划。在个人被抽中后，州政府会联系申请人参加计划，然而由于种种原因，并非所有抽中的人最终都参加了该计划。

LATE实例：OHIE数据

OHIE实验

Finkelstein等人（2012）根据这个计划研究了健康保险对医疗资源使用、健康等方面的影响。其主要的估计方程为：

$$y_{ih} = \beta_0 + \beta_1 \times \text{Insurance}_{ih} + x'_{ih}\eta + u_{ih}$$

而第一阶段方程为：

$$\text{Insurance}_{ih} = \delta_0 + \delta_1 \times \text{Lottery}_{ih} + x'_{ih}\zeta + \mu_{ih}$$

而ITT为：

$$y_{ih} = \gamma_0 + \gamma_1 \times \text{Lottery}_{ih} + x'_{ih}\xi + \epsilon_{ih}$$

从而$\gamma_1 = \beta_1 \times \delta_1$。代码：ohie_qje.do

工具变量、LATE及边际处理效应

回忆：

$$Y_i = Y_i(0) + W_i(Y_i(1) - Y_i(0)) + \epsilon_i + \eta_i W_i = Y_i(0) + W_i \tau_{ATE} + v_i$$

其中$\epsilon_i = Y_i(0) - \mathbb{E}(Y_i(0))$，$\eta_i = Y_i(1) - Y_i(0) - \mathbb{E}(Y_i(1) - Y_i(0))$:

- 异质性在处理效应中非常非常重要：工具变量估计不可用
- LATE解决了工具变量估计的可解释性问题，但是只能得到局部解释
    - 外部有效性不足
- 如何使用工具变量帮助识别ATE?

LATE

以上讨论的工具变量取值范围为$Z=0/1$，如果Z可以取多个值，比如$Z=z_0,...,z_K$，需要重新排列使得:

$$\mathbb{E}\left(W_i|Z_i=z_{k-1}\right)\leq\mathbb{E}\left(W_i|Z_i=z_k\right),k=1,...,K$$

假设:

$$\mathbb{E}\left[g\left(Z_i\right)W_i\right]=0$$

那么工具变量估计:

$$\tau_{IV}=\frac{\mathbb{C}\left(g\left(Z_i\right),Y_i\right)}{\mathbb{C}\left(g\left(Z_i\right),W_i\right)}=\sum_{k=0}^{K}\lambda_k\tau_{z_k,z_{k-1}}$$

其中$\tau_{z_k,z_{k-1}}=\mathbb{E}\left[Y_i\left(1\right)-Y_i\left(0\right)|W_i\left(z_k\right)=1,W_i\left(z_{k-1}\right)=0\right]$。其中$g\left(\cdot\right)$函数决定了权重。

LATE

如果工具是连续的，那么可以定义：

$$\tau_z = \lim_{z' \downarrow z, z'' \uparrow z} \tau_{z', z''}$$

Vytlacil (2002)证明，如果工具变量的假设对于所有的z', z''成立，那么实际上只需要假设一个潜在的index结构：

$$W_i = \mathbb{1}\left\{h\left(Z_i\right) \geq U_i\right\}$$

我们可以将U_i标准化为均匀分布$U_i \sim \mathcal{U}\left[0, 1\right]$。

边际处理效应

定义边际处理效应$\tau(u)$:

$$\tau(u)=\mathbb{E}\left(Y_i(1)-Y_i(0)\,|U_i=u\right)$$

实际上边际处理效应可以看成LATE，即由于:

$$\begin{aligned}\tau_z &= \lim_{z'\downarrow z, z''\uparrow z}\mathbb{E}\left[Y_i(1)-Y_i(0)\,|W_i\left(z'\right)=1, W_i\left(z''\right)=0\right]\\ &= \lim_{z'\downarrow z, z''\uparrow z}\mathbb{E}\left[Y_i(1)-Y_i(0)\,|h\left(z'\right)\geq U_i, h\left(z''\right)<U_i\right]\\ &= \mathbb{E}\left(Y_i(1)-Y_i(0)\,|h(z)=U_i\right)\end{aligned}$$

因而:

$$\tau(u)=\tau_z, u=h(z)$$

而平均处理效应:

$$\tau_{ATE}=\int_0^1\tau(u)\,du$$

边际处理效应的含义

一个例子。Mincer方程：

$$\ln(wage_i) = \alpha + \beta W_i + \gamma_1 exp_i + \gamma_2 exp_i^2 + u_i$$

然而，方程中我们忽略了一些因素，比如动力（motivation）：

$$\ln(wage_i) = \alpha + \beta W_i + \gamma_1 exp_i + \gamma_2 exp_i^2 + mot_i + u_i$$

进而，教育回报对于motivation可能是异质性的：

$$\ln(wage_i) = \alpha + \beta W_i + \gamma_1 exp_i + \gamma_2 exp_i^2 + mot_i + \theta \cdot W_i \cdot mot_i + u_i$$

动力越高，W_i越容易偏向于1，而潜在工资可能越高：selection-on-unobservalbe。

边际处理效应的含义

倾向得分:

$$p_i = P\left(W_i | Z_i\right)$$

假设Z_i为连续的。对Mincer方程求期望:

$$\mathbb{E}\left(\ln\left(wage_i\right) | exp_i, Z_i\right) = K\left(p_i, mot_i\right) = K\left(p_i\right)$$

其中第二个等号由于Z_i的外生性。MTE:

$$MTE = \frac{\partial \mathbb{E}\left(\ln\left(wage_i\right) | exp_i, Z_i\right)}{\partial p_i} = K'\left(p\right)$$

如果motivation能观察到:

$$MTE = \beta + \theta \cdot mot_i$$

MTE设定

一般设定：

$$Y(1) = \mu_1(X, U_1)$$

$$Y(0) = \mu_0(X, U_0)$$

注意，这里允许X与U_0, U_1相关。处理效应：

$$\Delta = Y(1) - Y(0) = \mu_1(X, U_1) - \mu_0(X, U_0)$$

MTE设定

假定：

$$W_i = \mathbb{1}\left\{P\left(Z_i\right) \geq U_{W,i}\right\}$$

不失一般性假设$U_W \sim \mathcal{U}\left[0, 1\right]$，重要假设：$Z_i \perp\!\!\!\perp U_{W,i}$

MTE与自选择

特例：广义Roy模型：

$$Y_1 = \mu_1(X) + U_1$$

$$Y_0 = \mu_0(X) + U_0$$

$$W = \mathbb{1}\{Y_1 - Y_0 - C \geq 0\}$$

其中成本：

$$C = \mu_c(Z) + U_C$$

因而：

$$\begin{aligned} W &= \mathbb{1}\{\mu_1(X) - \mu_0(X) - \mu_C(Z) \geq U_C + U_0 - U_1\} \\ &= \mathbb{1}\{\mu(Z) \geq V\} \\ &= \mathbb{1}\left\{F_V^{-1}(\mu(Z)) \geq U_W\right\} \end{aligned}$$

MTE

MTE被定义为:

$$MTE(u) = \mathbb{E}(Y_1 - Y_0 | U_W = u)$$

- 如前所述，与LATE关系密切
- 提供了U_W的不同分位数上的平均效应
    - 也就是不同的Propensity Score上的平均效应
- 反映了选择与处理效应异质性之间的关系

SHANGHAI UNIVERSITY OF INTERNATIONAL BUSINESS AN
1960

MTE与其他处理效应的关系

其他的处理效应可以写为MTE的函数:

$$ATE = \mathbb{E}(Y_1 - Y_0) = \int_0^1 MTE(u_W)\, du_W$$

$$TT = \mathbb{E}(Y_1 - Y_0 | W = 1) = \int_0^1 MTE(u_W)\, \omega_{TT}(u_W)\, du_W$$

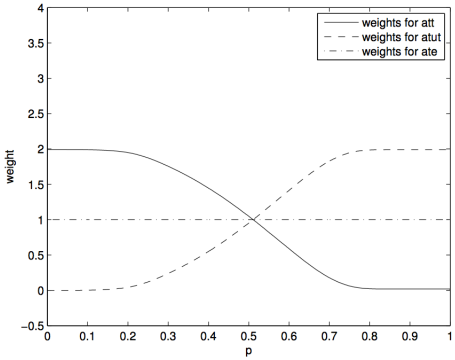
$$TUT(x) = \mathbb{E}(Y_1 - Y_0 | W = 0) = \int_0^1 MTE(u_W)\, \omega_{TUT}(u_W)\, du_W$$

$$PRTE = \mathbb{E}(Y_{a'}) - \mathbb{E}(Y_a) = \int_0^1 MTE(u_W)\, \omega_{PRTE}(u_W)\, du_W$$

$$IV_J = \int_0^1 \Delta^{MTE}(u_W)\, \omega_{IV}^J(u_W)\, du_W \quad given\ instruments\ J(Z)$$

MTE

Weights on MTE for Alternative Parameters



(a) ATT, ATUT, and ATE

Source: "Beyond LATE with a Discrete Instrument: Heterogeneity in the Quantity-Quality Interaction of Children", by Brinch, Mogstad, and Wiswall (JPE, 2016, forthcoming).

MTE的识别

识别:

$$
\begin{aligned}
\mathbb{E}\left(Y|P\left(Z\right)=p\right) &= \mathbb{E}\left(WY\left(1\right)+\left(1-W\right)Y\left(0\right)|P\left(Z\right)=p\right) \\
&= \mathbb{E}\left(Y\left(0\right)\right)+\mathbb{E}\left(W\left(Y\left(1\right)-Y\left(0\right)\right)|P\left(Z\right)=p\right) \\
&= \mathbb{E}\left(Y\left(0\right)\right) \\
&+\mathbb{E}\left(W\left(Y\left(1\right)-Y\left(0\right)\right)|W=1,P\left(Z\right)=p\right)P\left(W=1|P\left(Z\right)=p\right) \\
&= \mathbb{E}\left(Y\left(0\right)\right)+\int_0^p \mathbb{E}\left(W\left(Y\left(1\right)-Y\left(0\right)\right)|U_W=u\right)du
\end{aligned}
$$

从而:

$$
\frac{\partial\mathbb{E}\left(Y|P\left(Z\right)=p\right)}{\partial p}=\mathbb{E}\left(W\left(Y\left(1\right)-Y\left(0\right)\right)|U_W=p\right)
$$

MTE的估计

- 估计方法:
  1. 计算propensity score：$p = p(x, z)$
  2. 使用y对x和p做回归
     - 注意Y对p应该是非线性函数：多项式、半参数估计
     - 检查common support
  3. 计算ATE、ATT等处理效应
- stata命令：margte