工具变量

慧航

2023年5月

内生性问题

- 在线性回归中，我们假设了$\mathbb{E}\left(u|x\right)=0$，从而得到了最小二乘估计。
- 当这一假定被违背时，即$\mathbb{E}\left(u|x\right)\neq 0$，我们称为有内生性问题。
  1. 微观、宏观中的「内生变量」与计量中的内生性的联系区别？
  2. 内生性可能的原因：
     - 遗漏变量
     - 互为因果
     - 度量误差
     - 自选择
     - ……

遗漏变量

如果真实的数据生成过程为:

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_K x_{Ki} + \gamma q_i + v_i$$

而变量q_i是观测不到的，并且假设q_i与x_i之间存在着相关性:

$$q_i = \delta_0 + \delta_1 x_{1i} + \cdots + \delta_K x_{Ki} + e_i$$

那么将以上方程带入结构式，得到:

$$y_i = (\beta_0 + \gamma\delta_0) + (\beta_1 + \gamma\delta_1)\, x_{1i} + \cdots + (\beta_K + \gamma\delta_K)\, x_{Ki} + \gamma e_i + v_i$$

如果令$u_i = \gamma e_i + v_i$，我们有$\mathbb{E}\left(u_i | x_i\right) = 0$，因而那么如果我们忽略了变量$q_i$，那么实际得到的回归系数为

$$\beta_k^* = \beta_k + \gamma\delta_k$$

存在着偏误。

评估遗漏变量的偏误

- 根据以上分析，遗漏变量的误差取决于遗漏变量对于y的影响大小，以及遗漏变量与解释变量之间的相关性。
- 一般而言，以上两种相关性并不容易观测，因而评估遗漏变量对系数造成的偏误大小是比较困难的。
- 一种做法是可以根据可观测变量对可能的由不可观测变量导致的偏误进行评估，比如在加入和不加入控制变量两种情况下，核心解释变量的系数变化可以部分代表估计量的稳定性。
- 这一做法并不严谨，更严谨的推论通常需要更多的假设
- 一般而言，仅仅通过比较加入控制变量前后系数的变化并不足以验证遗漏变量所可能带来偏误的大小。Oster（2019）指出使用以上方法需要额外考虑R^2。

评估遗漏变量的偏误

- 为了说明这一点，我们考虑如下例子：

$$Y = \beta W + X + C$$

- 我们不妨将Y解释为收入，W为教育，而X和C分别为两种不同的能力。简单起见，假设$\beta = 0$。由于W与X和C存在相关性，所以如果我们考虑如下两种回归：
  - 直接使用Y对W回归，得到$\mathring{\beta}$；
  - 使用Y对W和X做回归，得到$\hat{\beta}$
- 那么以上两个估计量都会存在偏差。
- 一种想法是，可以通过比较$\hat{\beta}$和$\mathring{\beta}$，如果系数变化不大，那么意味着结果可能对遗漏变量是相对稳健的。

评估遗漏变量的偏误

- 然而Oster（2019）指出，这种做法是有问题的，关键在于$\mathbb{V}(X)$和$\mathbb{V}(C)$的大小。
- 如果可观测的X的方差远小于不可观测的C的方差，从而加入X的回归的R^2并不会比没有加入时增加太多，此时比较$\hat{\beta}$和$\mathring{\beta}$会发现系数是稳定的，并不会有太大的变化。
- 然而，$\hat{\beta}$和$\mathring{\beta}$实际上都存在着遗漏变量带来的偏差，从而比较$\hat{\beta}$和$\mathring{\beta}$发现系数稳定，如果不考虑R^2的变化，很有可能得到的结论是误导性的。
- Oster（2019）使用模拟的办法展示了这一问题

评估遗漏变量的偏误

- Oster（2019）考虑了更一般的设定，对于模型：

$$y_i = \beta w_i + x_i'\psi + q_i + u_i$$

  其中w_i为核心解释变量，x_i为一系列控制变量，而q_i为不可观测的部分
- 记$r_i = x_i'\psi$为可观测部分，不失一般性假设q_i与x_i的每一部分均不相关（即可以将q_i看做是遗漏变量中与x_i无关的部分）。
- 定义

$$\delta = \frac{\sigma_{qw}/\sigma_q^2}{\sigma_{rw}/\sigma_r^2}$$

  其中$\sigma_{rw} = \mathbb{C}\text{ov}\left(r_i, w_i\right), \sigma_{qw} = \mathbb{C}\text{ov}\left(q_i, w_i\right), \sigma_r^2 = \mathbb{V}\left(r_i\right), \sigma_q^2 = \mathbb{V}\left(q_i\right)$，因此$\delta$可以被解释为不可观测变量中与$w$相关的比例与可观测控制中与$w$相关的比例的比值。

评估遗漏变量的偏误

- 现在，如果记w对x的回归系数为η_j，如果额外假设：

$$\frac{\eta_j}{\eta_{j'}} = \frac{\psi_j}{\psi_{j'}} \tag{1}$$

- Oster（2019）得到估计量：

$$\beta^* \approx \hat{\beta} - \delta\left(\mathring{\beta} - \hat{\beta}\right)\frac{R^2_{\max} - R^2}{R^2 - \mathring{R}^2}$$

其中$R^2_{\max}$为理论上可以达到的最大的R^2，$\mathring{R}^2$为使用y对w回归的R^2，而R^2为加入x作为控制变量后的R^2

- 可以得到β^*为一致估计量：

$$\beta^* \xrightarrow{p} \beta$$

评估遗漏变量的偏误

- 由此可见，$\hat{\beta}$的偏差不仅仅取决于$\mathring{\beta} - \hat{\beta}$，还取决于三个不同的$R^2$：
  - $R^2_{\max} - R^2$可以被看做是不可观测部分的部分
  - 而$R^2 - \mathring{R}^2$为可观测的控制变量部分
  - 如果不可观测的部分方差远远小于可观测的控制变量部分的方差，那么遗漏变量的偏差相应的可能会比较小。

评估遗漏变量的偏误

- 注意到如果我们知道了δ和R^2_{max}，我们就可以直接计算出一致估计量β^*。
- 而如果放弃假设式(1)，也有类似的结论。
- Oster（2019）建议可以通过两种办法评估遗漏变量所带来的的偏误可能的大小：
  - 给定δ, R^2_{max}，直接计算出β^*，如果β^*与$\hat{\beta}$的符号相反，意味着$\hat{\beta}$可能不够稳健
  - 计算$\beta^* = 0$时所对应的δ，δ越大则$\hat{\beta}$更加稳健

评估遗漏变量的偏误

- 然而为了使用以上方法，我们必须给出δ和$R^2_{\max}$的具体取值。
  - 对于δ，Oster（2019）建议$\delta = 1$是一个通常来讲比较合适的选择；
  - 而对于$R^2_{\max}$，虽然由于y可能存在天然的误差等原因，实际上能够达到的最大的R^2多数情况下应该是小于1的，不过设定$R^2_{\max} = 1$给出了一个更加稳健的结果。
- 当然，实际应用时也可以根据现实情况对δ和$R^2_{\max}$做出相应调整。

评估遗漏变量的偏误

教育汇报的遗漏变量问题

我们使用WAGE2.dta，使用Mincer方程估计教育回报问题：

```
use datasets/WAGE2.dta
gen exper2=exper^2
reg lwage educ exper exper2, r
```

为了评估遗漏变量可能带来的偏误，可以使用psacalc命令

```
psacalc beta educ
```

评估遗漏变量的偏误

教育汇报的遗漏变量问题

其中beta选项代表需要计算R^*，而educ表示核心解释变量；也可以使用delta和rmax选项修改δ和$R^2_{\max}$的默认值：

```
psacalc beta educ, rmax(0.8) delta(2)
```

如果需要计算\delta，可以使用：

```
psacalc delta educ
```

此外，mcontrol选项可以用于将一些变量在估计$\hat{\beta}$和$\mathring{\beta}$时都用作控制变量，比如我们将变量exper及exper2在所有回归中都控制，仅仅比较IQ加入之前和之后的结果差异：

```
reg lwage educ exper exper2 IQ, r
psacalc beta educ, mcontrol(exper exper2)
```

评估遗漏变量的偏误

教育汇报的遗漏变量问题

- 值得注意的是，β^*的计算是通过解一个一元三次方程组实现的，因而可能出现多组解的情况，该命令默认使用与现有估计差距最小的作为beta汇报，同时也会汇报其他的解，在使用时可能需要甄别。
- 如果需要使用reghdfe，可以使用psacalc2命令：https://github.com/ArthurHowardMorris/psacalc_supports_reghdfe

互为因果

如果存在互为因果的两个内生变量y_1, y_2，结构方程：

$$y_1 = \alpha_1 y_2 + x_1\beta_1 + u_1$$

$$y_2 = \alpha_2 y_1 + x_2\beta_2 + u_2$$

联立两个方程，可以得到：

$$\begin{aligned}
y_2 &= \alpha_2\left(\alpha_1 y_2 + x_1\beta_1 + u_1\right) + x_2\beta_2 + u_2 \\
&= \alpha_1\alpha_2 y_2 + \alpha_2 x_1\beta_1 + x_2\beta_2 + \alpha_2 u_1 + u_2 \\
&= \frac{\alpha_2}{1-\alpha_1\alpha_2}x_1\beta_1 + \frac{1}{1-\alpha_1\alpha_2}x_2\beta_2 + \frac{\alpha_2 u_1 + u_2}{1-\alpha_1\alpha_2} \\
&\triangleq x'\delta + v
\end{aligned}$$

使用决定y_2、但同时不决定y_1的外生的扰动，即包含在x_2而不包含在x_1的变量(z)进行识别。

度量误差

考虑一个一元线性回归，假设数据的真实生成过程为：

$$y_i = \beta_0^* + \beta^* x_i^* + v_i$$

其中x_i^*为真实值。

- 观察不到x_i^*
- 只能观察到有误差的$x_i = x_i^* + e_i$

如果我们直接用y_i对x_i做回归，即：

$$y_i = \beta_0 + \beta x_i + u_i$$

那么$u_i = v_i - \beta e_i$。如果假设$\mathbb{E}\left(e_i | x_i^*\right) = 0$，那么

$$\mathbb{C}\text{ov}\left(x_i, e_i\right) = \mathbb{C}\text{ov}\left(x_i^* + e_i, e_i\right) = \mathbb{V}\left(e_i\right)$$

因而

$$\mathbb{C}\text{ov}\left(x_i, u_i\right) = \mathbb{C}\text{ov}\left(x_i, v_i - \beta e_i\right) = -\beta \mathbb{V}\left(e_i\right) \neq 0$$

因而导致了内生性问题。

度量误差

如果我们直接使用带有度量误差的x_i进行回归，那么我们将得到：

$$
\begin{aligned}
\text{plim}\hat{\beta} &= \frac{\mathbb{C}\text{ov}\left(x_i, y_i\right)}{\mathbb{V}\left(x_i\right)} \\
&= \frac{\mathbb{C}\text{ov}\left(x_i^* + e_i, \beta_0^* + \beta^* x_i^* + v_i\right)}{\mathbb{V}\left(x_i^*\right) + \mathbb{V}\left(e_i\right)} \\
&= \beta^* \cdot \frac{\mathbb{V}\left(x_i^*\right)}{\mathbb{V}\left(x_i^*\right) + \mathbb{V}\left(e_i\right)}
\end{aligned}
$$

得到的$\left|\hat{\beta}\right| < |\beta^*|$，即估计的系数的绝对值总是小于真实值的绝对值，存在着向中性偏误（attenuation bias）。

自选择

- 在经济学中，需要研究的很多变量通常并不是随机给定，而是行为人最大化效用进而做出选择的结果。
- 当存在这种自选择（self-selection）问题时，如果将这些行为人自己选择的变量作为自变量，经常会存在内生性问题。

自选择

教育回报

如果我们关心是否上大学对未来收入的影响，记$d_i=1$为上过大学，$d_i=0$为没上过大学，记$y_i(0)$为假设该个体没上过大学的收入，$y_i(1)$为假设该个体上过大学的收入，且数据生成过程为：

$$\begin{aligned} y_i(1) &= \alpha + x_i'\beta + u_{1i} \\ y_i(0) &= x_i'\beta + u_{0i} \end{aligned}$$

那么观察到的收入为：

$$\begin{aligned} y_i &= d_i y_i(1) + (1-d_i)\, y_i(0) \\ &= \alpha \cdot d_i + x_i'\beta + d_i u_{1i} + (1-d_i)\, u_{0i} \\ &\triangleq \alpha \cdot d_i + x_i'\beta + v_i \end{aligned}$$

自选择

教育回报

若d_i是完全随机分配的，即$d_i \amalg (u_{1i}, u_{0i})$，那么：

$$\begin{aligned}\mathbb{C}\mathrm{ov}(d_i, v_i) &= \mathbb{C}\mathrm{ov}(d_i, d_i u_{1i} + (1-d_i) u_{0i}) \\ &= \mathbb{C}\mathrm{ov}(d_i, d_i u_{1i}) + \mathbb{C}\mathrm{ov}(d_i, (1-d_i) u_{0i}) \\ &= \mathbb{E}\left(d_i^2 u_{1i}\right) - \mathbb{E}(d_i)\,\mathbb{E}(d_i u_{1i}) - \mathbb{E}(d_i (1-d_i) u_{0i}) \\ &= 0\end{aligned}$$

然而，如果d_i不是随机分配的，比如个体通过如下过程选择是否上大学：

$$d_i = 1\{y_i(1) \geq y_i(0)\} = 1\{\alpha + u_{1i} \geq u_{0i}\}$$

那么d_i与(u_{1i}, u_{0i})不可能独立，因而$\mathbb{C}\mathrm{ov}(d_i, v_i) \neq 0$。

工具变量

不管是遗漏变量、反响因果、度量误差还是自选择问题，都会导致核心解释变量与误差项的相关性，即在回归方程：

$$y = \beta \times y_1 + x'\delta + u$$

中y_1与误差项u之间相关。比如：

- 如果y_1为国家的GDP，y为一个国家内战爆发的次数
- 如果y_1为孩子的数量，y为孩子的教育经费（quantity-quality tradeoff）

解决方案：找到外生的影响x但是不会直接影响y的变量，即工具变量（instrument variable）z，比如：

- 内战例子中，可以使用天气（z）作为国家GDP的工具变量
- 生育问题中，计划生育政策作为孩子数量的工具变量

工具变量

如果我们要估计的结构方程（structural equation）为：

$$y = \alpha + \beta y_1 + u$$

然而$\mathbb{C}\mathrm{ov}\left(y_1, u\right) \neq 0$，但是我们可以找到一个$z$，使得$\mathbb{C}\mathrm{ov}\left(z, u\right) = 0$，且：

$$y_1 = \gamma + \delta z + e$$

我们把以上方程带入结构式，得到：

$$\begin{aligned} y &= \alpha + \beta\left(\gamma + \delta z + e\right) + u \\ &= \left(\alpha + \beta\gamma\right) + \beta\delta z + u + \beta e \end{aligned}$$

我们称上式为简约式（reduced form）。

工具变量

现在两个式子：

$$y_1 = \gamma + \delta z + e$$

以及

$$\begin{aligned} y &= (\alpha + \beta\gamma) + \beta\delta z + u + \beta e \\ &= \gamma^* + \delta^* z + v \end{aligned}$$

都不存在内生性问题，因而我们可以一致估计$\gamma, \delta, \gamma^*, \delta^*$。进而，可以得到$\beta$：

$$\beta = \frac{\delta^*}{\delta} = \frac{\beta\delta}{\delta}$$

假设：

1. z与u不相关
2. 分母（δ）不为0。

工具变量

一般情况下，如果估计方程：

$$y = \alpha y_1 + x'\beta + u$$

其中y_1为关心的内生变量，即$\mathbb{C}\text{ov}\left(y_1, u\right) \neq 0$，但是存在工具变量$z$，使得$\mathbb{E}\left(u|z\right) = 0$，且$\mathbb{E}\left(y_1|x, z\right) \neq \mathbb{E}\left(y_1|x\right)$，那么$y_1$的简约式可以写为：

$$y_1 = z'\gamma + x'\delta + v$$

那么可以通过工具变量达到对α的识别。

两阶段最小二乘

估计：两阶段最小二乘法（2SLS）

1. 第一阶段回归，对y_2的简约式进行回归：

$$y_1 = z'\gamma + x'\delta + v$$
$$\hat{y}_1 = z'\hat{\gamma} + x'\hat{\delta}$$

或者$\hat{Y}_1 = P_Z Y_1$，其中$Z = (Z, X)$，$P_Z = Z\left(Z'Z\right)^{-1} Z'$。

2. 第二阶段回归，用$\hat{y}_1$代替y_1：

$$y = \hat{\alpha}\hat{y}_1 + x'\hat{\beta} + \hat{e}$$

或者：$\left(\hat{\alpha}', \hat{\beta}'\right)' = \left(XP_Z Z\right)^{-1}\left(XP_Z Y\right)$，其中$X = (Y_1, X)$。

注意：不要手动算两阶段最小二乘，差别：在算标准误的时候用$\hat{e}$还是$\hat{u} = y - \hat{\alpha}y_1 + x'\hat{\beta}$?

两阶段最小二乘

工具变量的理解：我们使用了那些variation?

1. 第一阶段回归：
$$\hat{y}_1 = z'\hat{\gamma} + x'\hat{\delta}$$
2. 第二阶段回归：
$$y = \hat{\alpha}\hat{y}_1 + x'\hat{\beta} + \hat{e}$$

考虑使用分步回归，即先用$\hat{y}_1$对x做回归得到残差，等价于用$z'\hat{\gamma}$对x做回归，得到残差，再用y对以上残差做回归。
只使用了单纯因为工具变量而导致的y_1的变化的部分variation（见LATE）。

GMM

1. 两阶段最小二乘即在特定矩条件下的广义矩估计（Generalized method of moments， GMM）
   - 矩条件为：$\mathbb{E}(z_i u_i) = 0$，则GMM目标函数：

$$\min \left[\sum_i z_i \left(y_i - \tilde{x}_i' \tilde{\beta} \right) \right]' W \left[\sum_i z_i \left(y_i - \tilde{x}_i' \tilde{\beta} \right) \right]$$

   如果取$W = Z'Z$，则得到了两阶段最小二乘（2SLS）。
2. 在同方差假定下，以上的权重矩阵为最优权重矩阵。
3. **当同方差假定不满足时，以上的权重矩阵不是最优权重矩阵。**

过度识别检验

- 在$L > K$的情形下，我们可以检验矩条件（工具变量）是否有效。
- 可以证明，**当使用最优权重矩阵时**，GMM目标函数渐进服从χ^2分布，自由度为$L-K$：

$$\left[\sum_i m\left(w_i,\hat{\theta}\right)\right]'\left(\sum_{i=1}^{N}\left[m\left(w_i,\hat{\theta}\right)m\left(w_i,\hat{\theta}\right)'\right]\right)^{-1}\left[\sum_i m\left(w_i,\hat{\theta}\right)\right] \overset{a}{\sim} \chi^2\left(L-K\right)$$

- Hansen's J-statistics
- Sargan test
- 思想：使用K个工具做估计，用剩下的$L-K$个工具做检验

控制函数法

控制函数法：如果估计方程：

$$y = \alpha y_1 + x'\beta + u$$

以及y_1的简约式：

$$y_1 = z'\gamma + x'\delta + v$$

在简约式中，y_1被分解为两部分：

1. $z'\gamma + x'\delta$是与u不相关的部分
2. 如果y_1与u相关，则相关性必在v里面。

控制函数法

因而，如果假定：

$$u = \rho v + e$$

那么：

$$\begin{aligned}\mathbb{E}\left(y|x,z,v\right) &= \alpha\mathbb{E}\left(y_1|x,z,v\right) + x'\beta + \mathbb{E}\left(u|x,z,v\right)\\ &= \alpha y_1 + x'\beta + \rho v + e\end{aligned}$$

待估计方程变为：

$$y = \alpha y_1 + x'\beta + \rho v + e$$

因而如果观察到v，就可以直接控制「内生性」v从而达到识别。

控制函数法

估计步骤：

1. 估计简约式$y_1 = z'\gamma + x'\delta + v$得到$\hat{v}$
2. 将$\hat{v}$带入主方程，使用回归：

$$y = \alpha y_1 + x'\beta + \rho\hat{v} + e$$

   进行估计。

检验内生性：$H_0 : \rho = 0$（或者Hausman检验）。

控制函数法

- 或者，可以使用极大似然估计，即假设(u_i, v_i)的联合（正态）分布，进而：

$$f\left(y_i, y_{1i} | x_i, z_i\right) = f\left(y_i | y_{1i}, x_i, z_i\right) \cdot f\left(y_{1i} | x_i, z_i\right)$$

  如果假定(u_i, v_i)为联合正态分布，即得到有限信息极大似然估计（LIML）。
- 其他估计方法：$k-$class估计量

工具变量

工具变量的两个假定：

1. 与误差项不相关——Hansen’s test
2. 与内生变量高度相关

如果第二项假定不满足？

弱工具

以上得知，对于要估计的方程：

$$y = \alpha + \beta y_1 + u$$

以及第一阶段：

$$y_1 = \gamma + \delta z + e$$

估计可以写为：

$$\hat{\beta}_{2SLS} = \frac{\mathbb{C}\text{ov}\left(y, z\right)}{\mathbb{C}\text{ov}\left(y_1, z\right)}$$

如果分母趋向于0，IV估计的结果会非常不稳定，并偏向OLS估计量——弱工具。

工具变量

- Angrist and Krueger (1991)研究教育回报时，使用了出生季节与州、年份的乘积作为上学时间的工具变量
- Bound, Jaeger and Baker (1996)对此研究做出了批评，发现随机生成一组出生季节仍然能得到相似的结论。
- 工具变量太多导致估计结果偏向OLS的估计结果

工具变量

- 对于少量弱工具:
  - 诊断弱工具
  - 使用LIML，并调整置信区间。
- 对于Many instruments:
  - 2SLS表现非常差
  - 使用LIML，或者REQML(Chamberlain and Imbense, 2004)
  - 使用Lasso对第一阶段工具进行挑选

欠识别检验

- 对于第一阶段：
$$y_1 = z'\gamma + x'\delta + v$$
我们要求工具变量至少要和内生变量相关
- 如果不相关：欠识别（underidentification）
- 检验：
  - 原假设：$H_0 : \gamma = 0$ 使用F检验——第一阶段F值
  - Stata：Kleibergen and Paap rk Wald F

弱工具诊断

弱工具诊断：

- 第一阶段F值，一般在一个内生变量一个工具变量的情况下，为了使得2SLS估计量偏差不大于10%，F要大于10
- Cragg-Donald 统计量（$F - value$推广，当只有一个内生变量时就是$F - value$）
- Stock and Yogo (2002)提出了针对Cragg-Donald 统计量的临界值

置信区间调整：

- Anderson and Rubin (1949)
- Lee等人（2022）：根据第一阶段F值进行调整
- Valid t-ratio inference, Lee et al.(2022)

tF标准误

- 诊断与置信区间调整方法的结合
- Stock and Yogo (2005)的诊断方法可以理解为:

$$P\left\{t^2 > c^*, F > F^*\right\} \leq \alpha$$

  其中F为第一阶段的F值。以上步骤即首先判断第一阶段F值是否大于某个临界值，进而使用传统的t检验
- Lee等人（2022）：

$$P\left\{t^2 > c_\alpha\left(F\right)\right\} \leq \alpha$$

  即根据第一阶段F值调整t统计量的临界值
  - 缺点：目前只支持一个内生变量、一个工具变量

tF标准误

Panel A. Selected values of tF critical values, $\sqrt{c_{0.05}(F)}$, *and tF standard error adjustments,* $\sqrt{c_{0.05}(F)}/1.96$

F	4.000	4.008	4.015	4.023	4.031	4.040	4.049	4.059	4.068	4.079
$\sqrt{c_{0.05}(F)}$	18.656	18.236	17.826	17.425	17.033	16.649	16.275	15.909	15.551	15.201
$\sqrt{c_{0.05}(F)}/1.96$	9.519	9.305	9.095	8.891	8.691	8.495	8.304	8.117	7.934	7.756
	4.090	4.101	4.113	4.125	4.138	4.151	4.166	4.180	4.196	4.212
	14.859	14.524	14.197	13.878	13.566	13.260	12.962	12.670	12.385	12.107
	7.581	7.411	7.244	7.081	6.922	6.766	6.614	6.465	6.319	6.177
	4.229	4.247	4.265	4.285	4.305	4.326	4.349	4.372	4.396	4.422
	11.834	11.568	11.308	11.053	10.804	10.561	10.324	10.091	9.864	9.642
	6.038	5.902	5.770	5.640	5.513	5.389	5.268	5.149	5.033	4.920
	4.449	4.477	4.507	4.538	4.570	4.604	4.640	4.678	4.717	4.759
	9.425	9.213	9.006	8.803	8.605	8.412	8.222	8.037	7.856	7.680
	4.809	4.701	4.595	4.492	4.391	4.292	4.195	4.101	4.009	3.919
	4.803	4.849	4.897	4.948	5.002	5.059	5.119	5.182	5.248	5.319
	7.507	7.338	7.173	7.011	6.854	6.699	6.549	6.401	6.257	6.117
	3.830	3.744	3.660	3.578	3.497	3.418	3.341	3.266	3.193	3.121
	5.393	5.472	5.556	5.644	5.738	5.838	5.944	6.056	6.176	6.304
	5.979	5.844	5.713	5.584	5.459	5.336	5.216	5.098	4.984	4.872
	3.051	2.982	2.915	2.849	2.785	2.723	2.661	2.602	2.543	2.486
	6.440	6.585	6.741	6.907	7.085	7.276	7.482	7.702	7.940	8.196
	4.762	4.655	4.550	4.448	4.348	4.250	4.154	4.061	3.969	3.880
	2.430	2.375	2.322	2.270	2.218	2.169	2.120	2.072	2.025	1.980
	8.473	8.773	9.098	9.451	9.835	10.253	10.711	11.214	11.766	12.374
	3.793	3.707	3.624	3.542	3.463	3.385	3.309	3.234	3.161	3.090
	1.935	1.892	1.849	1.808	1.767	1.727	1.688	1.650	1.613	1.577
	13.048	13.796	14.631	15.566	16.618	17.810	19.167	20.721	22.516	24.605
	3.021	2.953	2.886	2.821	2.758	2.696	2.635	2.576	2.518	2.461
	1.542	1.507	1.473	1.440	1.407	1.376	1.345	1.315	1.285	1.256
	27.058	29.967	33.457	37.699	42.930	49.495	57.902	68.930	83.823	104.67
	2.406	2.352	2.299	2.247	2.197	2.147	2.099	2.052	2.006	1.96
	1.228	1.200	1.173	1.147	1.121	1.096	1.071	1.047	1.024	1.00

tF标准误

- 或者可以使用软件包：
  - net install tf, force from(http://www.princeton.edu/~davidlee/wp/)
  - 需要首先安装ivreg2、ranktest
  - 用法与ivreg2一致

LATE

令:

$$
\begin{aligned}
Y_i &= Y_i(0) + W_i(Y_i(1) - Y_i(0)) \\
&= \mathbb{E}(Y_i(0)) + W_i\mathbb{E}(Y_i(1) - Y_i(0)) + \zeta_i + \eta_i W_i
\end{aligned}
$$

其中$\zeta_i = Y_i(0) - \mathbb{E}(Y_i(0))$，$\eta_i = Y_i(1) - Y_i(0) - \mathbb{E}(Y_i(1) - Y_i(0))$。
如果存在一个工具变量Z_i，使得$\mathbb{C}\mathrm{ov}(Z_i, W_i) \neq 0$，
且$Z_i \amalg (Y_i(0), Y_i(1))$，且$Z = 0/1$那么IV估计:

$$
\mathrm{plim}\tau_{IV} = \frac{\mathbb{C}\mathrm{ov}(Y, Z)}{\mathbb{C}\mathrm{ov}(W, Z)} = \frac{\mathbb{E}(Y|Z=1) - \mathbb{E}(Y|Z=0)}{P(W|Z=1) - P(W|Z=0)}
$$

- 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。

LATE

如果给定一个工具变量Z_i为0/1变量，我们记：

$$W_i\left(Z_i\right)=Z_iW_i\left(1\right)+\left(1-Z_i\right)W_i\left(0\right)=\begin{cases}W_i\left(1\right) & Z_i=1\\ W_i\left(0\right) & Z_i=0\end{cases}$$

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

LATE

关键假设：

1. $Z_i \amalg (Y_i(1), Y_i(0), W_i(1), W_i(0))$
2. $P(W_i = 1|Z_i)$取决于Z_i

ITT

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

$$\tau_{ITT}=\mathbb{E}\left(Y_i|Z_i=1\right)-\mathbb{E}\left(Y_i|Z_i=0\right)$$

将其分解：

$$\begin{aligned}
\tau_{ITT}&=\mathbb{E}\left(Y_i|Z_i=1\right)-\mathbb{E}\left(Y_i|Z_i=0\right)\\
&=\mathbb{E}\left(Y_i\left(0\right)+W_i\left(Y_i\left(1\right)-Y_i\left(0\right)\right)|Z_i=1\right)\\
&-\mathbb{E}\left(Y_i\left(0\right)+W_i\left(Y_i\left(1\right)-Y_i\left(0\right)\right)|Z_i=0\right)\\
&=\mathbb{E}\left(W_i\left(Y_i\left(1\right)-Y_i\left(0\right)\right)|Z_i=1\right)-\mathbb{E}\left(W_i\left(Y_i\left(1\right)-Y_i\left(0\right)\right)|Z_i=0\right)\\
&=\mathbb{E}\left(W_i\left(1\right)\left(Y_i\left(1\right)-Y_i\left(0\right)\right)|Z_i=1\right)-\mathbb{E}\left(W_i\left(0\right)\left(Y_i\left(1\right)-Y_i\left(0\right)\right)|Z_i=0\right)\\
&=\mathbb{E}\left(W_i\left(1\right)\left(Y_i\left(1\right)-Y_i\left(0\right)\right)\right)-\mathbb{E}\left(W_i\left(0\right)\left(Y_i\left(1\right)-Y_i\left(0\right)\right)\right)\\
&=\mathbb{E}\left(\left(W_i\left(1\right)-W_i\left(0\right)\right)\left(Y_i\left(1\right)-Y_i\left(0\right)\right)\right)
\end{aligned}$$

LATE

进一步使用全概率公式:

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

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

$$
\tau_{ITT} = \mathbb{E}\left(\left(Y_i\left(1\right)-Y_i\left(0\right)\right)|W_i\left(1\right)=1,W_i\left(0\right)=0\right)P\left(W_i\left(1\right)=1,W_i\left(0\right)=0\right)
$$

LATE

由于:

$$
\begin{aligned}
P(W_i=1|Z_i=1)-P(W_i=1|Z_i=0) &= P(W_i(1)=1)-P(W_i(0)= \\
&= P(W_i(1)=1, W_i(0)=0)
\end{aligned}
$$

进而，工具变量估计了:

$$
\begin{aligned}
LATE=\tau_{IV} &= \frac{\tau_{ITT}}{P(W=1|Z=1)-P(W=1|Z=0)} \\
&= \mathbb{E}\left(\left(Y_i(1)-Y_i(0)\right)|W_i(1)=1, W_i(0)=0\right)
\end{aligned}
$$

即如果假设$P(W_i(1)=0, W_i(0)=1)=0$（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**。

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：

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

而第一阶段方程为：

$$Insurance_{ih} = \delta_0 + \delta_1 \times Lottery_{ih} + x'_{ih}\zeta + \mu_{ih}$$

而ITT为：

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

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