工具变量

工具变量估计法及其局限性

报告人：马兆星

工具变量估计法的理解

- 目标：干扰项e_i的条件均值独立于处置变量D_i，即

$$E(e_i|D_i,\boldsymbol{X}_i)=0$$

- 两种方法来满足
  - “清理”干扰项：添加控制变量(可观测变量或不可观测但随时间变化的变量)
  - “清理”解释变量：不可观测并且随时间变化的变量

↓

工具变量法：将处置变量中与干扰项不相关的部分分离出来

工具变量估计法的理解

- 假设结果方程为

$$Y_i = \alpha + \beta_1 D_i + e_\mathrm{i}$$

其中，D_i和干扰项e_i是相关的，即$Cov(D_i, e_\mathrm{i}) \neq 0$

- OLS回归的系数$\hat{\beta}_1^{OLS}$为

$$\hat{\beta}_1^{OLS} = \frac{\widehat{Cov}(Y_i, D_i)}{\widehat{Var}(D_i)}$$

- $\hat{\beta}_1^{OLS}$的样本概率极限值为

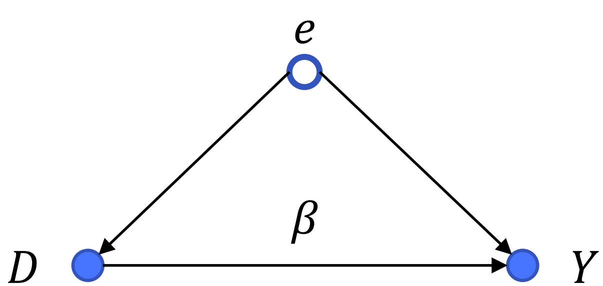
$$plim\hat{\beta}_1^{OLS} = plim\frac{\widehat{Cov}(Y_i,D_i)}{\widehat{Var}(D_i)} = plim\frac{\widehat{Cov}(\alpha+\beta_1 D_i+e_i,D_i)}{\widehat{Var}(D_i)}$$

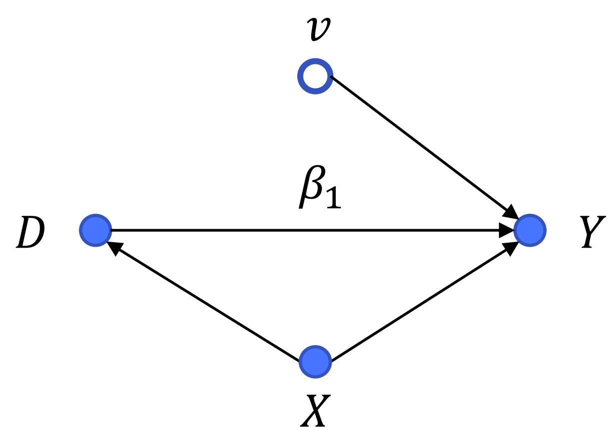
$$= \beta_1 + plim\frac{\widehat{Cov}(e_i,D_i)}{\widehat{Var}(D_i)}$$

$$= \beta_1 + \frac{Cov(e_i,D_i)}{Var(D_i)}$$

工具变量估计法的理解

- 因$Cov(D_i, e_i) \neq 0$，所以$\hat{\beta}_1^{OLS}$不以概率收敛为β_1
- D到Y的因果影响路径有两条：$D \rightarrow Y$(因果路径)和 $D \leftarrow e \rightarrow Y$(混淆路径)



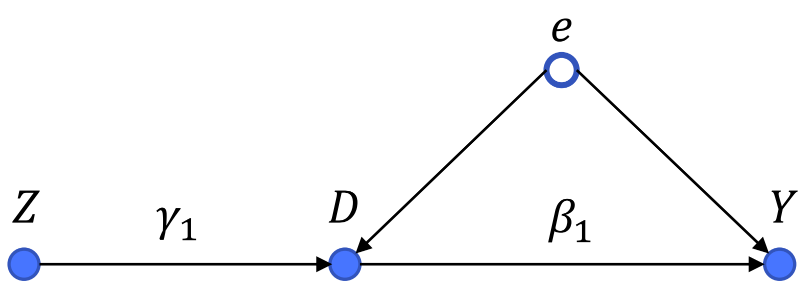


- 增加控制变量：通过观测得到X，将方程分解为

$$Y_i = \alpha + \beta_1 D_i + \beta_2 X_i + v_{\mathrm{i}}$$

工具变量估计法的理解

- 使用工具变量：工具变量Z"清理"掉内生变量中与干扰项相关的变化("坏"的变化)，再用与干扰项不相关的变化("好"的变化)去估计对Y的作用



  - 外生性：Z本身是"干净"的，Z和e_i不相关，即 $Cov(Z_i, e_i)=0$
  - 相关性：Z能够清理内生变量D，Z和D必须相关，即 $Cov(Z_i, D_i) \neq 0$

估计方法——间接最小二乘法

- 假设内生变量D和工具变量Z的回归关系为

$$D_i = \gamma_0 + \gamma_1 Z_i + u_\mathrm{i}$$

- 工具变量Z和被解释变量Y的关系为

$$Y_i = \alpha + \beta_1 D_i + e_i = \alpha + \beta_1(\gamma_0 + \gamma_1 Z_i + u_i) + e_i$$
$$= \alpha + \beta_1\gamma_0 + \beta_1\gamma_1 Z_i + \beta_1 u_i + e_i$$
$$= \pi_0 + \pi_1 Z_i + \xi_i$$

- 通过回归得到$\pi_1 = Cov(Y_i, Z_i)/Var(Z_i)$,又$\pi_1 = \beta_1\gamma_1$,可得

$$\beta_1^{ILS} = \frac{Cov(Y_i, Z_i)/Var(Z_i)}{Cov(D_i, Z_i)/Var(Z_i)} = \frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)}$$

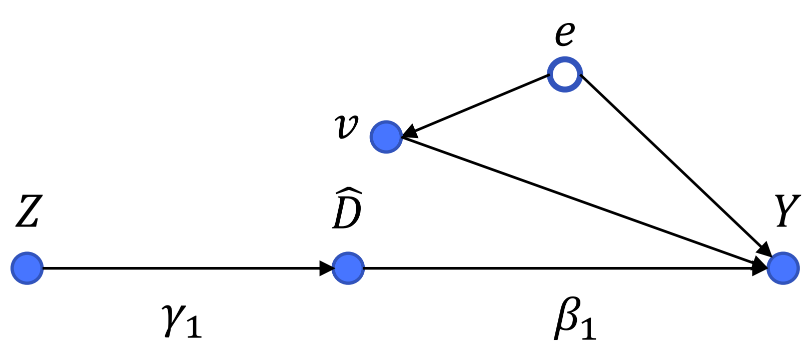
估计方法——两阶段最小二乘法

- 两阶段最小二乘法：通过直接分解出D_i中与干扰项e_i不相关的变化部分$\widehat{D}_i$来进行估计
- 第一阶段：通过工具变量Z_i将D_i分解为两个不相关变量

$$D_i = \widehat{D}_i + v_i = \gamma_0 + \gamma_1 Z_i + v_i$$

- 第二阶段：用D中“好的部分”$\widehat{D}$估计D对Y的影响

$$Y_i = \alpha + \beta_1 D_i + e_i = \alpha + \beta_1\left(\widehat{D}_i + v_i\right) + e_i = \alpha + \beta_1 \widehat{D}_i + \delta_i$$



估计方法——两阶段最小二乘法

- 此时$\widehat{D}_i$与v_i和e_i都不相关，因此$Cov\left(\widehat{D}_i, \delta_i\right) = 0$，对其进行回归，得到正确的$\beta_1$

$$\beta_1^{2SLS} = \frac{Cov(Y_i, \widehat{D}_i)}{Var(\widehat{D}_i)} = \frac{Cov(Y_i, \gamma_0 + \gamma_1 Z_i)}{Var(\gamma_0 + \gamma_1 Z_i)} = \frac{\gamma_1 Cov(Y_i, Z_i)}{\gamma_1^2 Var(Z)}$$

$$= \frac{Cov(Y_i, Z_i)}{\gamma_1 Var(Z_i)} = \frac{Cov(Y_i, Z_i)}{\frac{Cov(D_i, Z_i)}{Var(Z_i)} Var(Z_i)} = \frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)}$$

- 与间接最小二乘法得到的系数相同，方法处理方式稍有不同，本质上是一样的

工具变量数量问题

- 工具变量数量＜内生变量数量
  “识别不足”：两种方法都无法估计出内生变量的系数
- 工具变量数量＝内生变量数量
  “刚好识别”：模型中内生变量系数可得到唯一估计值
- 工具变量数量＞内生变量数量
  - 分别使用不同的工具变量，会得到不同的β_1估计值
  - 使用两个工具变量的线性组合，使用2SLS可以得到一个多工具变量的“最佳组合”，来最佳拟合D，及通过回归方程

$$D_i = \gamma_0 + \gamma_1 Z_{1i} + \gamma_2 Z_{2i} + v_i$$

得到$\widehat{D}_i = \hat{\gamma}_0 + \hat{\gamma}_1 Z_{1i} + \hat{\gamma}_2 Z_{2i}$

工具变量数量问题

- 此时两阶段最小二乘法用两个工具变量的线性组合最大限度地分离出D_i中的外生部分$\widehat{D}_i$，再使用"最佳"外生部分$\widehat{D}_i$，将Y_i对$\widehat{D}_i$回归

$$Y_i = \alpha + \beta_1 \widehat{D}_i + \delta_i$$

- 因此更常使用两阶段最小二乘法

两阶段最小二乘法

- 假设结果模型是

$$Y_i = \alpha + \beta_1 D_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + e_i$$

- 工具变量Z_{1i}满足两个条件
  - 外生性

  Z_{1i}与干扰项e_i不相关，$Cov(Z_i, e_i) = 0$，"干净"地分离D_{1i}
  - 相关性

  控制外生变量与D_{1i}的相关性后，Z_{1i}与D_{1i}仍存在相关性

$$D_i = \gamma_0 + \gamma_1 Z_{1i} + \gamma_2 X_{2i} + \cdots + \gamma_k X_{ki} + v_i$$

其中，工具变量的系数$\gamma_1 \neq 0$

两阶段最小二乘法——模型估计

- 第一阶段

将内生变量对工具变量和所有外生变量进行回归

$$D_{1i} = \gamma_0 + \gamma_1 Z_{1i} + \gamma_2 X_{2i} + \cdots + \gamma_k X_{ki} + v_i$$

用所得到的系数计算内生变量的预测值

$$\widehat{D}_i = \hat{\gamma}_0 + \hat{\gamma}_1 Z_{1i} + \hat{\gamma}_2 X_{2i} + \cdots + \hat{\gamma}_k Z_{ki}$$

- 第二阶段

用预测值$\widehat{D}_i$替代结果模型中的内生变量D_{1i}，进行回归

$$Y_i = \alpha + \beta_1 \widehat{D}_i + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \delta_i$$

多个内生变量和多个工具变量

- 假设要估计的模型是

$$Y_i = \alpha + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + e_i$$

D_{1i}的有一个工具变量Z_{1i}，D_{2i}的有两个工具变量Z_{2i}和Z_{3i}

- 第一阶段

将每个内生变量单独对所有工具变量和所有其他外生变量回归：

$$D_{1i} = \gamma_0 + \gamma_1 Z_{1i} + \gamma_2 Z_{2i} + \gamma_3 Z_{3i} + \gamma_4 X_{3i} + \cdots + \gamma_{k+1} X_{ki} + v_{1i}$$

$$D_{2i} = \theta_0 + \theta_1 Z_{1i} + \theta_2 Z_{2i} + \theta_3 Z_{3i} + \theta_4 X_{3i} + \cdots + \theta_{k+1} X_{ki} + v_{2i}$$

多个内生变量和多个工具变量

- 第二阶段
  用内生变量预测值替代模型中的内生变量并进行回归
  $Y_i = \alpha + \beta_1 \widehat{D}_{1i} + \beta_2 \widehat{D}_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + \delta_i$
- 得到D_1和$D_2$2SLS估计系数$\hat{\beta}_1^{2SLS}$和$\hat{\beta}_2^{2SLS}$是β_1和β_2的一致估计量

工具变量估计法的局限性——大样本

- 偏差性

$$plim\hat{\beta}_1^{2SLS} = plim\frac{\widehat{Cov}(Y_i,Z_i)}{\widehat{Cov}(D_i,Z_i)} = \beta_1 + plim\frac{\widehat{Cov}(Z_i,e_i)}{\widehat{Cov}(D_i,Z_i)}$$

$$= \beta_1 + \frac{Cov(Z_i,e_i)}{Cov(D_i,Z_i)}$$

- 当工具变量完全外生时，即$Cov(Z_i,e_i)=0$，2SLS大样本偏差项=0， $plim\hat{\beta}_1^{2SLS}=\beta_1$， $\hat{\beta}_1^{2SLS}$是一致估计量
- 当$Cov(Z_i,e_i)\neq 0$时，即使分子很小，如果分母$Cov(D_i,Z_i)$很小，偏差会被放得很大
- 与内生变量相关性很小的工具变量被称为弱工具变量

工具变量估计法的局限性

- 比较$\hat{\beta}_1^{2SLS}$和$\hat{\beta}_1^{OLS}$在大样本下的偏差

$$\frac{plim\hat{\beta}_1^{2SLS}-\beta_1}{plim\hat{\beta}_1^{OLS}-\beta_1}=\frac{Cov(Z_i,e_i)}{Cov(D_i,Z_i)}\frac{Var(D_i)}{Cov(D_i,e_i)}$$

  - 如果工具变量不完全外生，且工具变量很弱，那么上式的比率可能远大于1，即使是在大样本中，两阶段工具变量估计量$\hat{\beta}_1^{2SLS}$也可能比OLS估计量的偏差还大。

工具变量估计法的局限性

- 有效性

工具变量两阶段估计量$\hat{\beta}_1^{2SLS}$是渐近正态分布的

$$\hat{\beta}_1^{2SLS} \xrightarrow{\mathrm{d}} N\left(\beta_1, Avar\left(\hat{\beta}_1^{2SLS}\right)\right)$$

在同方差情况下，其渐进方差为

$$Avar\left(\hat{\beta}_1^{2SLS}\right) = \frac{\sigma_e^2}{N\sigma_D^2\rho_{DZ}^2}$$

不考虑内生性使用OLS估计模型，得到系数渐进方差

$$Avar\left(\hat{\beta}_1^{OLS}\right) = \frac{\sigma_e^2}{N\sigma_D^2}$$

工具变量估计法的局限性

- 对比$\hat{\beta}_1^{2SLS}$和$\hat{\beta}_1^{OLS}$的方差，得到

$$\frac{Avar\left(\hat{\beta}_1^{2SLS}\right)}{Avar\left(\hat{\beta}_1^{OLS}\right)} = \frac{1}{\rho_{DZ}^2}$$

  - ρ_{DZ}小于1，故工具变量估计值的方差总是大于OLS估计值的方差，即$Avar\left(\hat{\beta}_1^{2SLS}\right) > Avar\left(\hat{\beta}_1^{OLS}\right)$因为工具变量估计值只使用了$D_i$中与工具变量相关的一部分信息，而OLS使用了全部信息
  - 在弱工具变量情况下，ρ_{DZ}很小，能分解出的内生变量"好"的信息很少，造成$Avar\left(\hat{\beta}_1^{2SLS}\right)$很大，此时通常的显著性检验并不可靠

工具变量估计法的局限性——有限样本

- 偏差性

  有限样本中，模型的$\hat{\beta}_1^{2SLS}$的偏差为

  $$\hat{\beta}_1^{2SLS}\text{偏差} = E\left(\hat{\beta}_1^{2SLS}\right) - \beta_1 = \frac{K\rho}{N}\left(\frac{1}{R^2} - 1\right)$$

  - 其中，K为工具变量数量，N为样本数量，ρ为内生变量与干扰项的相关系数及R^2为工具变量与内生变量的相关系数
  - 则弱工具变量会造成较大偏差，尤其样本数量较小的时候

工具变量估计法的局限性

- 有限样本中，$\hat{\beta}_1^{2SLS}$与$\hat{\beta}_1^{OLS}$偏差的比率

$$\frac{E\left(\hat{\beta}_1^{2SLS}\right)-\beta_1}{E\left(\hat{\beta}_1^{OLS}\right)-\beta_1}=\frac{K}{NR^2}$$

  - 即使工具变量处理了内生性问题，如果工具变量太弱，在有限样本里的偏差有可能比OLS的还糟糕

- 有效性

  在有限样本里，$\hat{\beta}_1^{2SLS}$的分布未知，因此通常使用的t检验在小样本里并不适用