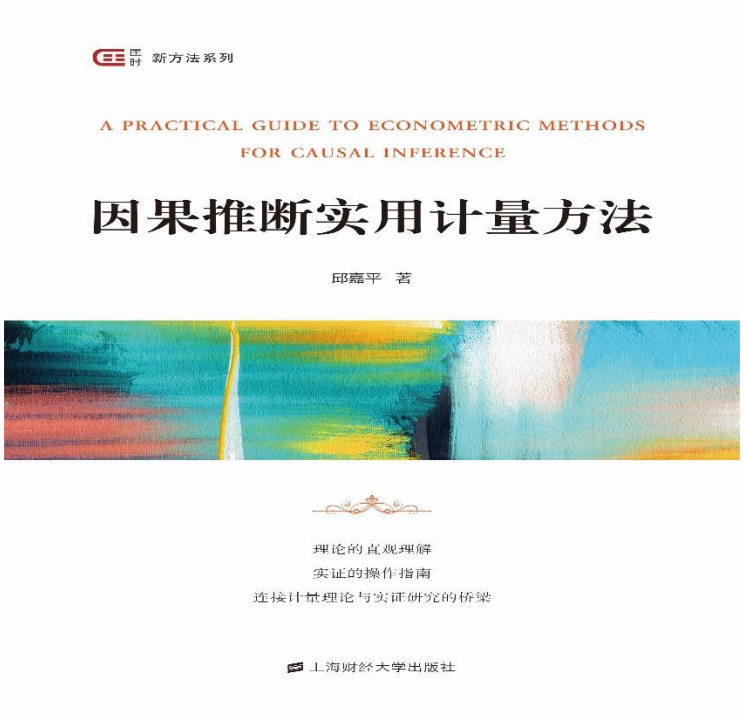
第10章：工具变量



邱嘉平

大纲

1. 工具变量估计法的直观理解
2. 工具变量两阶段估计法
3. 工具变量估计法的局限性
4. 工具变量估计的检验
5. 工具变量使用步骤和常见问题

工具变量估计法的直观理解

工具变量的作用

- 如果我们直观地认为处置变量D_i变化中和干扰项相关的部分为“坏的变化”, 和干扰项不相关的部分为“好的变化”。

- 工具变量的方法就是将处置变量中“好的变化”分离出来，并只用“好的变化“部分去估计处置变量对被解释变量的因果影响。

通过工具变量清理内生变量以解决内生性问题

- 工具变量，顾名思义，不同于解释变量，它是帮助我们找到因果关系的工具。

- 工具变量方法解决内生性走的是另一条途径：它先“清理”掉内生处置变量中和干扰项相关的变化（“坏”的变化），再用和干扰项不相关的变化（“好”的变化）去估计对Y的作用。

工具变量的要求

一个能达到清理内生变量的工具变量需要“有用”并且“干净”。

- 要清理出D好的部分，Z本身必须是**干净**的，$Cov(Z_i, e_i) = 0$。我们称这个条件为“外生性”（Exogeneity）

- Z能够清理内生变量D，Z必须和D相关，即$Cov(Z_i, D_i) \neq 0$，如果二者不相关，它的变化信息没法帮助我们分离出D “好”的变化，Z是“没用的”工具变量。

估计方法1：间接最小二乘法（Indirect Least Squares）

- 假设结果方程为：

$$Y_i = \alpha + \beta_1 D_i + e_i \quad (1)$$

D_i是内生变量，$Cov(D_i, e_i) \neq 0$，

- 内生变量D和工具变量Z的回归关系如下：

$$D_i = \gamma_0 + \gamma_1 Z_i + u_i \quad (2)$$

将方程（2）带入方程（1），得到：

$$\begin{aligned} Y_i &= \alpha + \beta_1(\gamma_0 + \gamma_1 Z_i + u_i) + e_i \\ &= \underbrace{\alpha\beta_1\gamma_0}_{\pi_0} + \underbrace{\beta_1\gamma_1}_{\pi_1} Z_i + \underbrace{\beta_1 u_i + e_i}_{\zeta_i} \end{aligned}$$

可见工具变量Z和被解释变量Y的的关系如下

$$Y_i = \pi_0 + \pi_1 Z_i + \zeta_i,$$

估计方法1：间接最小二乘法（ILS）

$$\pi_1 = \beta_1 \gamma_1,$$

它的意思是

$$\underbrace{(Z\text{对}Y\text{的作用})}_{\pi_1} = \underbrace{(Z\text{对}D\text{的作用})}_{\gamma_1} \times \underbrace{(D\text{对}Y\text{的作用})}_{\beta_1}。$$

因此，要得到D对Y的作用的系数β_1，我们可以用下面关系得到：

$$(D\text{对}Y\text{的作用}) = \frac{(Z\text{对}Y\text{的作用})}{(Z\text{对}D\text{的作用})}$$

$$\beta_1 = \frac{\pi_1}{\gamma_1}$$

估计方法1：间接最小二乘法（ILS）

- 从回归方程：$Y_i = \pi_0 + \pi_1 Z_i + \zeta_i$
  得到：$\pi_1 = Cov(Z_i, Y_i)/Var(Z_i)$

- 从回归方程：$D_i = \gamma_0 + \gamma_1 Z_i + u_i$
  得到：$\gamma_1 = Cov(Z_i, D_i)/Var(Z_i)$

- 代入$\beta_1 = \frac{\pi_1}{\gamma_1}$，我们得到间接最小二乘法的系数估计量$\beta^{ILS}$。

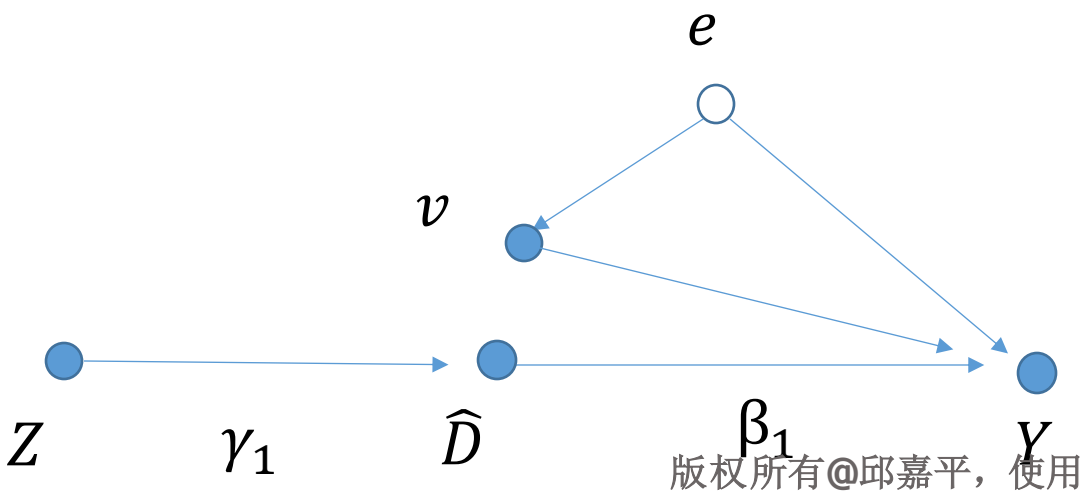
$$\beta_1^{ILS} = \frac{Cov(Y_i, Z_i)/Var(Z_i)}{Cov(D_i, Z_i)/Var(Z_i)} = \frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)}$$

估计方法2：两阶段最小二乘法(Two Stages Least Square, 2SLS)

- 第一阶段通过工具变量Z_i将D_i分解为两个不相关的变量，$D_i = \widehat{D}_i + v_i$，其中$\widehat{D}_i = \gamma_0 + \gamma_1 Z_i$，分解可以通过将$D_i$对$Z_i$回归，得到：

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

- 回归分解后$Cov(Z_i, v_i) = 0$，因此$Cov(\widehat{D}_i, v_i) = 0$，$\widehat{D}_i$也和$e$不相关，是$D_i$中“好的部分”。



估计方法2：两阶段最小二乘法（2SLS）

- 第二阶段用D中“好的部分”$\widehat{D}$估计D对Y的影响，将分解后的D代入公式，得到：

$$
\begin{aligned}
Y_i &= \alpha + \beta_1 D_i + e_i \\
&= \alpha + \beta_1 \left(\widehat{D}_i + v_i\right) + e_i \\
&= \alpha + \beta_1 \widehat{D}_i + \underbrace{\delta_i}_{\beta v_i + e_i}
\end{aligned}
$$

- 由于$\widehat{D}_i = \gamma_0 + \gamma_1 Z_i$与$v_i$和$e_i$都不相关，因此$Cov(\widehat{D}_i, \delta_i) = 0$，对回归可以得到正确的$\beta_1$

估计方法2：两阶段最小二乘法（2SLS）

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

我们看到两阶段最小二乘法β_1^{2SLS}得到的系数和间接最小二乘法得到的系数β^{ILS}是一样的，

$$\beta_1^{2SLS} = \beta_1^{ILS} = \frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)}$$

工具变量数量

- 识别不足：工具变量的数量<内生变量的数量。这两种方法都没法估计出模型中内生变量的系数β_1。

- 刚好识别：工具变量数量=内生变量数量。模型中内生变量的系数β可以得到唯一的估计值，比如模型只有一个内生变量和一个工具变量，ISL法和2LSL法将得到相同的唯一解。

工具变量数量

- 过度识别：工具变量数量>内生变量数量。如果分别使用不同工具变量，会得到不同的β_1估计值。

- 2SLS提供了一个多工具变量的“最佳组合“方法。它通过第一阶段的回归找到两个工具变量的线性组合$\gamma_1 Z_1 + \gamma_1 Z_2$来最佳地拟合$D$，即通过回归方程：

$$D_i = \gamma_0 + \gamma_1 Z_i^1 + \gamma_2 Z_i^2 + v_i。$$

第二阶段将Y_i对$\widehat{D}_i$回归：

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

由于$\widehat{D}_i$和δ_i不相关，得到$\hat{\beta}_1^{2SLS}$是β_1一致估计量。

工具变量两阶段估计法

模型设置

我们要估计的结果模型是:

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

其中D_{1i}是一个内生变量，$\mathrm{Cov}(D_{1i}, e_i) \neq 0$,

$X_{2i}, \ldots, X_{ki}$是其他外生变量。D_{1i}有一个工具变量Z_{1i}。

合理的工具变量需要满足的条件

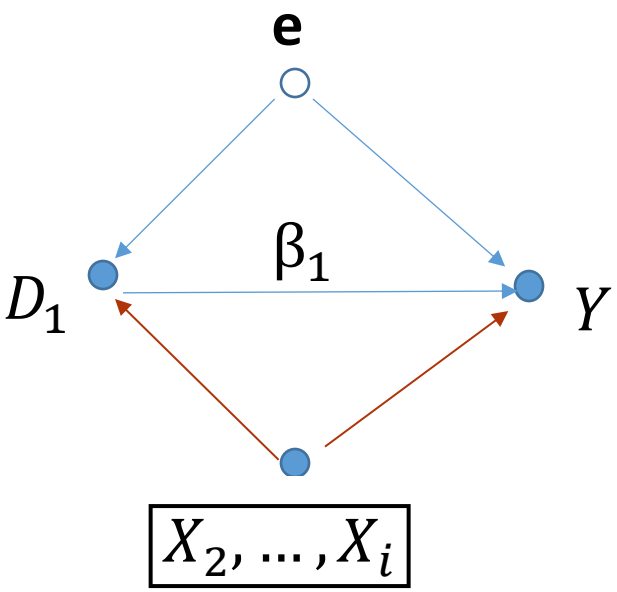
- **外生性（干净）**： Z_{1i}和结果模型中的干扰项e_i不相关，$Cov(Z_{1i}, e_i) = 0$，**即当控制了模型里所有解释变量（包括内生变量和其他外生变量变量）对Y_i的作用后，Z_{1i}对Y_i没有作用。**

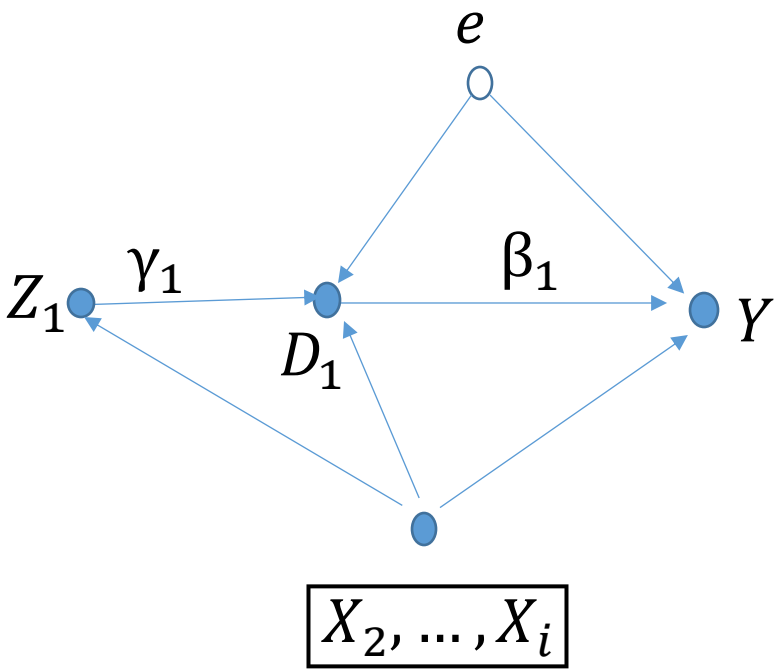
- **相关性（有用）**： 在内生变量D_{1i}和工具变量Z_{1i}和其它外生变量的回归关系中，

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

工具变量的系数$\gamma_1 \neq 0$。即当控制了模型里所有外生变量和D_{1i}的相关性后，Z_{1i}和D_{1i}仍然存在相关性。具体而言，

合理的工具变量需要满足的条件





模型估计：一个内生变量和一个工具变量

第一阶段：将内生变量对工具变量和所有外生变量做回归

$$\underbrace{D_{1i}}_{\text{内生变量}} = \gamma_0 + \underbrace{\gamma_1 Z_{1i}}_{\text{工具变量}} + \underbrace{\gamma_2 X_{2i} + \cdots + \gamma_k X_{ki}}_{\text{所有的外生变量}} + v_i$$

用得到的估计系数$(\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_i)$计算内生变量的预测值：

$$\widehat{D}_{1i} = \hat{\gamma}_0 + \hat{\gamma}_1 Z_{1i} + \hat{\gamma}_2 X_2 i + \cdots + \hat{\gamma}_k X_k 。$$

这个预测值$\widehat{D}_{1i}$不和e_i相关，它是D_{1i} 中“好”的变化部分。

模型估计：一个内生变量和一个工具变量

第二阶段：将预测值$\widehat{D}_{1i}$替代结果模型中的内生变量D_{1i}并进行回归

$$Y_i = \alpha + \underbrace{\beta_1 \widehat{D}_{1i}}_{\text{预测值}} + \underbrace{\beta_2 X_{2i} + \cdots + \beta_k X_k}_{\text{所有的外生变量}} + \delta_i$$

得到$\widehat{D}_{1i}$的系数$\hat{\beta}_1^{2SLS}$称为样本两阶段最小二乘法（2SLS）系数。

由于在第二阶段回归中只使用了D_{1i} 中和干扰项不相关的“好”的变化部分$\widehat{D}_{1i}$，因此得到的$\hat{\beta}_1^{2SLS}$是系数β_1的一致估计量$plim\hat{\beta}_1^{2LS} = \beta_1$。

模型估计：多个内生变量和多个工具变量

2SLS可以很方便地处理有多个内生变量和多个工具变量的情况。

假设我们要估计的模型是:

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

其中D_{1i}和D_{2i}是内生变量。D_{1i}有个工具变量Z_{1i}，D_{2i}有两个工具变量Z_{2i}和Z_{3i}。

模型估计:多个内生变量和多个工具变量

第一阶段：将每个内生变量单独对所有工具变量和所有其它外生变量做回归：

$$\underbrace{D_{1i}}_{\text{内生变量}} = \gamma_0 + \underbrace{\gamma_1 Z_{1i} + \gamma_2 Z_{2i} + \gamma_3 Z_{3i}}_{\text{所有工具变量}} + \underbrace{\gamma_4 X_{3i} + \cdots + \gamma_{k+1} X_k}_{\text{所有其他外生变量}} + v_{1i}$$

$$\underbrace{D_{2i}}_{\text{内生变量}} = \theta_0 + \underbrace{\theta_1 Z_{1i} + \theta_2 Z_{2i} + \theta_3 Z_{3i}}_{\text{所有的工具变量}} + \underbrace{\theta_4 X_{3i} + \cdots + \theta_{k+1} X_k}_{\text{所有其他外生变量}} + v_{2i}$$

用得到的系数计算每个内生变量的预测值：

$$\widehat{D}_{1i} = \hat{\gamma}_0 + \hat{\gamma}_1 Z_{1i} + \hat{\gamma}_2 Z_{2i} + \hat{\gamma}_3 Z_{3i} + \hat{\gamma}_4 X_{3i} + \cdots + \hat{\gamma}_{k+1} X_{ki}$$

$$\widehat{D}_{2i} = \hat{\theta}_0 + \hat{\theta}_1 Z_{1i} + \hat{\theta}_2 Z_{2i} + \hat{\theta}_3 Z_{3i} + \hat{\theta}_4 X_{3i} + \cdots + \hat{\theta}_{k+1} X_{ki}$$

模型估计:多个内生变量和多个工具变量

第二阶段：用内生变量的预测值替代结果模型中的内生变量并进行回归

$$Y_i = \alpha + \underbrace{\beta_1 \widehat{D}_{1i} + \beta_2 \widehat{D}_{2i}}_{\text{预测值}} + \underbrace{\beta_3 X_{3i} + \cdots + \beta_k X_{ki}}_{\text{所有的外生变量}} + \delta_i$$

得到X_1和X_2的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

理解工具变量估计法的局限性

为方便我们理解工具变量估计法的局限性，我们考虑简单的单变量两阶段模型

$$Y_i = \alpha + \beta_1 D_i + e_i$$

$$D_i = \gamma_0 + \gamma_1 Z_i + u_i$$

其中D_i是内生变量，Z_i是工具变量。

系数β_1的样本两阶段估计量 为

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

大样本下的局限性 （偏差性）

$\hat{\beta}_1^{2SLS}$的大样本概率极限值为：

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

$$= \beta_1 + \underbrace{\frac{Cov(Z_i, e_i)}{Cov(D_i, Z_i)}}_{\text{2SLS大样本偏差项}}$$

- 当工具变量是完全外生，即$Cov(Z_i, e_i) = 0$，2SLS大样本偏差项为0，$\hat{\beta}_1^{2SLS}$是一致估计量（当样本足够大，估计量趋近真实值）。

大样本下的局限性 （偏差性）

$$plim\hat{\beta}_1^{2LS} = \beta_1 + \underbrace{\frac{Cov(Z_i, e_i)}{Cov(D_i, Z_i)}}_{\text{2SLS大样本偏差项}}$$

- 但如果工具变量不是完全是外生的，$Cov(Z_i, e_i) \neq 0$，即使偏差项的分子$Cov(Z_i, e_i)$很小，如果工具变量和内生变量的相关性很小，即偏差项里的分母$Cov(D_i, Z_i)$很小，这个偏差也会被放得很大。

- **和内生变量相关性很小的工具变量称为弱工具变量**。

大样本下的局限性（精准性）

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

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

在同方差情况下，其渐进方差为：

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

其中σ_e^2和σ_D^2分别是干扰项e和内生变量D的方差，ρ_{DZ}是工具变量和内生变量的相关系数$\rho_{DZ} = \frac{\mathrm{Cov}(\mathrm{D}_i, Z_i)}{Var(Z_i)}$，

大样本下的局限性（精准性）

如果不考虑内生性而直接使用OLS估计，得到的系数的渐进方差为：

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

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

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

比较$\hat{\beta}_1^{2SLS}$和$\hat{\beta}_1^{OLS}$在大样本下的精准性

- **工具变量估计量的方差总是大于OLS估计值的方差**

- 在弱工具的情况下，工具变量和内生变量的相关系数ρ_{XZ}很小，因此能分解出的内生变量“好”的信息很少，工具变量估计系数$\hat{\beta}_1^{2SLS}$的估计精确度很低，造成方差$\mathrm{Avar}\left(\hat{\beta}_1^{2SLS}\right)$很大。

有限样本下的局限性 （偏差性）

- 在有限样本里，2SLS估计量β^{2SLS}是有偏的估计量（估计量的期望值不等于真实值），$\mathrm{E}(\hat{\beta}_1^{2SLS}) \neq \beta_1$。

- Hahn和Hausman（2002）给出了在有限样本里，$\hat{\beta}_1^{2SLS}$的偏差

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

有限样本下的局限性 （精确性）

- 在有限样本里，我们不知道$\hat{\beta}_1^{2SLS}$的分布。

- 样本方差$\widehat{Var}(\hat{\beta}_1^{2SLS})$在有限样本通常偏小，尤其是当工具变量是弱工具变量。这意味着如果使用$\widehat{\mathrm{Var}}\left(\hat{\beta}_1^{2SLS}\right)$进行$t$检验我们会过于容易拒绝原假设而得到“显著”结果。因此通常使用的t检验在小样本里不适用。

工具变量的局限性

- **要注意样本数量**。 2SLS的估计系数$\hat{\beta}_1^{2SLS}$只有在大样本里是一致的，在小样本里是有偏的。
- **除非确定有内生性，否则应避免使用工具变量估计**
- **避免使用弱工具变量**。

工具变量估计的检验

是否需要使用工具变量？（内生性检验）

- **Durbin-Wu-Hausman χ_J^2检验**

通过比较工具变量估计值$\hat{\beta}_1^{2SLS}$和OLS估计值$\hat{\beta}_1^{OLS}$来检验变量外生性的方法。

- H_0: D_i是外生的，**检验统计量：**

$$\mathrm{H} = \left(\hat{\beta}_1^{2SLS} - \hat{\beta}_1^{OLS}\right)' \left[Avar\left(\hat{\beta}_1^{2SLS}\right) - Avar\left(\hat{\beta}_1^{OLS}\right)\right]^{-1} \left(\hat{\beta}_1^{2SLS} - \hat{\beta}_1^{OLS}\right) \sim \chi_J^2$$

在原假设下OLS和工具变量得到的参数估计量是一致的，因此$\hat{\beta}_1^{2SLS} - \hat{\beta}_1^{OLS}$和H值应该接近零。

内生性检验的局限性

检验一个变量是否是外生的前提条件是我们有一个有效的工具变量（满足相关性和外生性），如果工具变量本身不是有效的，检验出来的结果是无效的。

工具变量是否满足相关性? (弱工具变量检验)

- 如果只有一个内生变量，一个简单的检验方法就是观察2SLS中第一阶段的关于所有工具变量的系数同时为0的F检验。当工具变量数量为1，2，3，5，10时，Stock，Wright，and Yogo (2002) 建议第一阶段的F统计量的关键值设为8.96，11.59，12.83，15.09和22.88。如果第一阶段的F检验值低于这些关键值，则可能存在弱工具变量问题。

工具变量是否是外生的？（过度识别检验）

- **恰当识别（工具变量的数量=内生变量的数量）**情况下，我们无法对外生性假设进行检验。

- 因为要检验$\mathrm{Cov}(Z_i, \mathrm{e}_i) = 0$，由于$\mathrm{e}_i$观测不到，我们必须先估计$\mathrm{e}_\mathrm{i}$，再用的它估计值$\hat{e}_i$检验$\mathrm{Cov}(Z_i, \hat{\mathrm{e}}_i) = 0$。但是当我们去估计$\hat{e}_i = Y_i - \hat{\beta}_1 D_i$时，我们需要满足外生条件的工具变量才能得到一致的估计值$\hat{\beta}_1$，从而得到一致的估计量$\hat{e}_i$。因此通过$\mathrm{Cov}(Z_i, \hat{\mathrm{e}}_i) = 0$检验工具变量是否外生是没有意义的，因为要得到一致的估计量$\hat{\beta}_1$，我们已经假设了工具是外生的。

工具变量是否是外生的？（过度识别检验）

- **过度识别(工具变量的数量>内生变量的数量）：**假设有一个内生变量，两个工具变量$(Z_1\ Z_2)$，我们可以采用下面的分步检验方法：

- 先假设第一个工具变量Z_1满足外生性$\mathrm{Cov}(\mathrm{Z_{1i}}, e_i) = 0$，并只用$Z_1$作为工具变量得到系数的估计值$\hat{\beta}_1^{Z_1}$，$\hat{\beta}_1^{Z_1}$是$\beta_1$一致估计量，因此残差$\hat{e}_i^{Z_1} = Y_i - \hat{\beta}_1^{Z_1} D_i$也是$\mathrm{e}_i$一致估计量。接着使用$\hat{e}_i^{Z_1}$对工具变量$Z_2$的外生性进行检验，即检验 $\mathrm{Cov}\left(Z_{2i}, \hat{e}_i^{Z_i}\right) = 0$是否成立。**注意这个检验是建立在第一个工具变量$Z_1$满足外生性的假设成立的前提下**。

工具变量是否是外生的？（过度识别检验）

- 类似地，我们也可以先假设第二个工具变量满足外生条件$\mathrm{Cov}(\mathrm{Z}_{2\mathrm{i}}, e_i) = 0$。并只用$\mathrm{Z}_2$作为工具变量得到系数的估计值$\hat{\beta}_1^{Z_2}$，如果$Z_2$满足外生性的假设成立，$\hat{\beta}_1^{Z_2}$是$\beta_1$一致估计量，因此残差$\hat{e}_i^{Z_2} = Y_i - \hat{\beta}_1^{Z_2} D_i$也是$\mathrm{e}_i$一致估计量。接着使用$\hat{e}_i^{Z_2}$对工具变量$\mathrm{Z}_1$的外生性进行检验，即检验 $\mathrm{Cov}\left(Z_{1i}, \hat{e}_i^{Z_2}\right) = 0$是否成立。**注意这个检验是建立在第二个工具变量$Z_2$满足外生性的假设成立的前提下。**

过度识别检验的局限性

- 如果一个工具变量Z_1通过了外生性检验，并不能说明Z_1一定是外生的，因为可能是另一个工具Z_2的外生性假设错误导致了Z_1通过了外生性检验。
- 同理，当Z_2通过检验是，我们也不能得出Z_2是外生的结论。

- **因此，当过度识别检验通过时，我们并不能得出所有工具变量都是外生的结论。**

常用的过度识别检验

原假设为H_0: 所有的工具变量都是外生的。

- 先使用所有工具变量用2SLS进行回归，得到残差（干扰项的估计值）$\hat{e}_i$ 。如果所有工具变量是外生的，残差$\hat{e}_i$是干扰项e_i的一致估计量。

- 将$\hat{e}_i$作为被解释变量，所有的工具变量（和模型里的其他外生变量）作为解释变量，用OLS进行回归，得到R^2。如果所有的工具变量都是外生的，那么它们与残差$\hat{e}_i$是无关的，所以R^2会较小。

- 进行假设检验，原假设 H_0: 所有工具变量都是外生下，统计量$NR^2 \sim x_q^2$，n是样本数，q是自由度=工具变量数-内生变量数。

常用的过度识别检验

- 如果NR^2大于相关的x_q^2关键值，那么我们可以得出结论：不是所有的工具变量都是外生的。虽然我们知道存在一些工具变量不是外生的，但是我们并不知道哪些工具变量不是外生的。

- 如果nR小于相关的x_q^2关键值，虽然通过了过度识别检验，但我们仍然不能确定所有工具变量都是外生的，因为有可能有的工具变量是内生的，造成残差$\hat{e}_i$的估计是错误的，检验也是无效的。

工具变量使用步骤

工具变量使用步骤

1. 清楚地定义研究问题，描述经济机制，设置基本的模型，对基本模型进行OLS回归得到初步结果，理解并描述模型可能存在的内生性问题的原因（反向因果关系问题，遗漏变量问题，测量误差问题）。

2. 寻找有效的工具变量，我们需要根据经济机制和理论基础提出有效的工具，并解释为什么选择的工具变量是相关的和外生的。由于外生性本质上没法通过统计方法检验，因此使用描述性语言和经济原理说明工具变量的外生性是使用工具变量的关键之处。

工具变量使用步骤

3. 使用工具变量估计法对模型进行估计，同时进行必要的统计检验并谨慎地对结果进行解释，这包括：
   - 检验变量外生性：
   - 检验工具变量相关性：报告第一阶段的F统计值，检验是否有弱工具变量；
   - 检验工具变量外生性：在过度识别情况下，使用过度识别检验检验工具变量是否是外生的，如果检验没通过，则存在至少有一个工具变量是内生的可能。

4. 将工具变量估计结果和OLS结果进行对比，理解结果为何有差异。

常见问题

一、用计量软件估计工具变量模型，不要自己手动进行两步回归
二、第一阶段回归应包含所有的外生变量
三、避免用组（组可以是行业，学校，省等）均值作为工具变量
四、避免用内生变量的滞后项做工具变量
五、模型含有二次项的工具变量的用法
六、模型存在交叉项时工具变量的用法
七、工具变量是越多越好吗
八、工具变量是解决内生性的万灵药吗？
九、理解工具变量的结果只是局部平均处置效应