

工具变量

李德山

¹ 西南科技大学 经济管理学院

²School of Economics and Management
Southwest University of Science and Technology

2021 年 3 月 19 日

① 工具变量估计法的直观理解

主要内容

- ① 工具变量估计法的直观理解
- ② 两阶段最小二乘法

主要内容

- ① 工具变量估计法的直观理解
- ② 两阶段最小二乘法
- ③ 工具变量法的局限性

主要内容

- ① 工具变量估计法的直观理解
- ② 两阶段最小二乘法
- ③ 工具变量法的局限性
- ④ 工具变量运用的检验

主要内容

- ① 工具变量估计法的直观理解
- ② 两阶段最小二乘法
- ③ 工具变量法的局限性
- ④ 工具变量运用的检验
- ⑤ 工具变量操作步骤

主要内容

- ① 工具变量估计法的直观理解
- ② 两阶段最小二乘法
- ③ 工具变量法的局限性
- ④ 工具变量运用的检验
- ⑤ 工具变量操作步骤
- ⑥ 工具变量运用常见问题

工具变量估计法的直观理解

- 回顾
- 为了得到无偏和一致估计量，我们所需的条件是干扰项 e 的条件均值独立于处置变量 D ，即 $E(e_i|D_i, X_i) = 0$ 。
 - 添加控制变量的方法可以“清理”干扰项，直到剩余的干扰因素与处置变量无关。
 - 匹配模型和一般回归模型只是控制了可观测变量。
 - 面板数据也只能控住不可观测且不随时间变化的变量。
 - 但现实是，有很多解释变量是不可观测并随时间变化的。

工具变量估计法的直观理解

- **工具变量**提供了一种解决办法。它不是“清理”干扰项，而是“清理”解释变量。
 - 如果认为处置变量变化中与干扰项相关的部分是“坏的变化”、与干扰项不相关的部分是“好的变化”。工具变量法就是将处置变量中“坏的变化”清理掉，将“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。
 - 工具变量可用来估计线性联立方程的系数。

工具变量估计法的直观理解

- 一些例子
 - Angrist and Krueger (1991) 利用美国普查数据估计教育收益率。
 - 能力影响个体的教育选择和收入水平，但无法利用匹配或回归方法得到教育的正确因果效应。
- 如何寻找工具变量？
 - 与教育密切相关
 - 与能力或其他未观测因素无关

工具变量估计法的直观理解

- 一些例子
 - Angrist and Krueger (1991) 利用美国普查数据估计教育收益率。
 - 能力影响个体的教育选择和收入水平，但无法利用匹配或回归方法得到教育的正确因果效应。
- 如何寻找工具变量？
 - 与教育密切相关
 - 与能力或其他未观测因素无关
 - Angrist and Krueger 用出生季度作为教育的工具变量。

工具变量估计法的直观理解

- 一些例子
 - Angrist and Krueger (1991) 利用美国普查数据估计教育收益率。
 - 能力影响个体的教育选择和收入水平，但无法利用匹配或回归方法得到教育的正确因果效应。
- 如何寻找工具变量？
 - 与教育密切相关
 - 与能力或其他未观测因素无关
 - Angrist and Krueger 用出生季度作为教育的工具变量。
 - Angrist (1990) 利用参军资格作为越战服务经历的工具变量来研究其对退伍老兵收入的影响。

工具变量估计法的直观理解

- 一些例子
 - Angrist and Krueger (1991) 利用美国普查数据估计教育收益率。
 - 能力影响个体的教育选择和收入水平，但无法利用匹配或回归方法得到教育的正确因果效应。
- 如何寻找工具变量？
 - 与教育密切相关
 - 与能力或其他未观测因素无关
 - Angrist and Krueger 用出生季度作为教育的工具变量。
 - Angrist (1990) 利用参军资格作为越战服务经历的工具变量来研究其对退伍老兵收入的影响。
 - Angrist (1998) 利用双胞胎作为家庭规模的工具变量来研究其对母亲劳动力供给的影响。

工具变量估计法的直观理解

- 工具变量估计法的图解

- 假设回归模型为:

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

这里假设, D 和 e 是相关的, 即 $Cov(D_i, e_i) \neq 0$ 。

- OLS 回归系数 $\hat{\beta}_1^{ols}$ 为:

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

其中, $\tilde{Cov}(Y_i, D_i)$ 和 $\tilde{Var}(D_i)$ 是样本协方差和方差

工具变量估计法的直观理解

- $\hat{\beta}_1^{ols}$ 的样本概率极限值为:

$$\begin{aligned} p \lim \hat{\beta}_1^{ols} &= p \lim \frac{\widetilde{Cov}(Y_i, D_i)}{\widetilde{Var}(D_i)} \\ &= p \lim \frac{\widetilde{Cov}(\alpha + \beta_1 D_i + e_i, D_i)}{\widetilde{Var}(D_i)} \\ &= \beta_1 + p \lim \frac{\widetilde{Cov}(e_i, D_i)}{\widetilde{Var}(D_i)} \\ &= \beta_1 + \underbrace{\frac{Cov(e_i, D_i)}{Var(D_i)}}_{bias \neq 0} \end{aligned}$$

- 因为 D_i 是内生的, $Cov(D_i, e_i) \neq 0$, 所以 $\hat{\beta}_1^{ols}$ 不依概率收敛于 β_1 , 即不是一致估计量。

工具变量估计法的直观理解

- 左图中, D 到 Y 的路径有两条: $D \rightarrow Y$ (因果路径) 和 $D \leftarrow e \rightarrow Y$ (混淆路径)
- 通过增加控制变量将干扰项 e 与 D 相关的部分分离出来。如果可通过观测得到 X , 把干扰项分解为 $e_i = \beta_2 X_i + \nu_i$, 方程 (1) 可以写成:

$$Y_i = \alpha + \beta_1 D_i + \beta_2 X_i + \nu_i$$

其中, D 和 X 相关, D 和 ν 不相关。

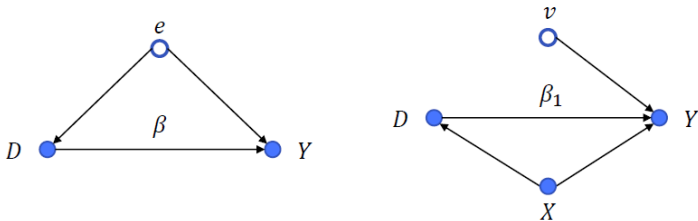


Figure: 通过控制变量清理干扰项以解决内生性问题

工具变量估计法的直观理解

- 假设 X 是观测不到的，那么增加控制变量就不可行了。
- 工具变量先“清理”掉内生变量中与干扰项相关的变化，再用与干扰项不相关的变化（“好”的变化）去估计对 Y 的作用。
- 工具变量 Z 需要满足两个条件：
 - 外生性： Z 本身是“干净”的， Z 和 e 不相关。
 - 相关性： Z 能够清理内生变量 D ， Z 和 D 必须相关。（下图中， D 是对撞变量）

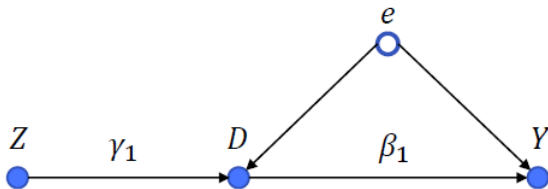


Figure: 工具变量的相关性和外生性

工具变量估计法的直观理解

- 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：

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

- 将上式代入到方程 (1)，工具变量 Z 和被解释变量 Y 的关系为：

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

- 前提是满足相关性和外生性。即 $Cov(D_i, Z_i) \neq 0$, $Cov(Z_i, e_i) = 0$ 。

工具变量估计法的直观理解

- $\pi_1 = \beta_1 \gamma_1$ 的直观理解

$$\underbrace{Z \rightarrow Y}_{\pi_1} = \underbrace{(Z \rightarrow D)}_{\gamma_1} \times \underbrace{(D \rightarrow Y)}_{\beta_1}$$

- 根据上式可得：

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

- 从而可以得到间接最小二乘法的系数估计量：

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

- 间接估计法避免了使用 D 信息中“坏”的部分。这个方法也被称为工具变量法或系数比例法。

工具变量估计法的直观理解

- 两阶段最小二乘法

- 通过直接分解出 D 与干扰项 e 不相关的变化部分 \tilde{D} 来进行估计的方法。

- 第一阶段：通过工具变量 Z 将 D 分解为两个不相关的变量，

$D_i = \tilde{D}_i + \nu_i$ ，其中 $\tilde{D}_i = \gamma_0 + \gamma_1 Z_i$ 。因为 Z 与 e 不相关，所以 \tilde{D}_i 与 e 也不相关，即 \tilde{D}_i 是 D_i 中“好”的部分。 ν_i 是 D_i 中“坏”的部分。

$$D_i = \tilde{D}_i + \nu_i = \gamma_0 + \gamma_1 Z_i + \nu_i$$

工具变量估计法的直观理解

- 由图可以看出，将 D 分解为两个变量 \tilde{D}_i 和 ν_i 。分解后的 \tilde{D}_i 与 e 之间不存在路径； \tilde{D}_i 到 Y 的开放路径只有一条。

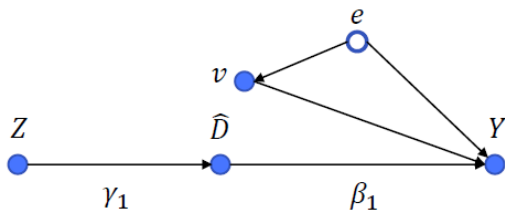


Figure: 通过工具变量清理内生变量以解决内生性问题

工具变量估计法的直观理解

- 第二阶段：用 D 中“好”的部分 \tilde{D}_i 估计 D 对 Y 的影响，将分解的 D 代入方程 (1) 中，可得：

$$\begin{aligned} Y_i &= \alpha + \beta_1 D_i + e_i \\ &= \alpha + \beta_1 (\tilde{D}_i + \nu_i) + e_i \\ &= \alpha + \beta_1 \tilde{D}_i + \underbrace{\delta_i}_{\beta_1 \nu_i + e_i} \end{aligned}$$

- \tilde{D}_i 与 ν_i 和 e_i 都不相关，因此 $Cov(\tilde{D}_i, \delta_i) = 0$ 。
- 对上式回归，可得：

$$\begin{aligned} \beta_1^{2sls} &= \frac{Cov(Y_i, \tilde{D}_i)}{Var(\tilde{D}_i)} = \frac{Cov(Y_i, \gamma_0 + \gamma_1 Z_i)}{Var(\gamma_0 + \gamma_1 Z_i)} = \frac{\gamma_1 Cov(Y_i, Z_i)}{\gamma_1^2 Var(Z_i)} \\ &= \frac{Cov(Y_i, Z_i)}{\gamma_1 Var(Z_i)} = \frac{Cov(Y_i, Z_i)}{\frac{Cov(D_i, Z_i)}{Var(Z_i)} Var(Z_i)} = \frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)} \end{aligned} \quad (2)$$

- 与间接最小二乘法得到的系数相同，只是处理方式稍有不同，本质上是一样的。

工具变量估计法的直观理解

- 工具变量数量问题

- 1 工具变量数小于内生变量数

- “识别不足”：两种方法都无法估计出内生变量的系数。

- 2 工具变量数等于内生变量数

- “恰好识别”：模型中内生变量系数可得到唯一估计值。

- 3 工具变量数大于内生变量数

- 如果分别使用不同的工具变量，会得到不同的系数估计值。
- 例子，假设有 2 个工具变量，无论使用哪个方法，可以得到两个一致的估计系数。由于每个工具变量都包含了有效的信息，如果将这两个工具变量组合成一个工具变量，在这种情况下，两阶段最小二乘法比间接最小二乘法更优越。此时两阶段最小二乘法是用两个工具变量的线性组合最大限度地分离出 D 中的外生部分 \tilde{D}_i ，再使用“最佳”的 \tilde{D}_i 与 Y 进行回归。因此，2SLS 提供了一个多工具变量的“最佳组合”方法。

两阶段最小二乘法

- 假设模型是：

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

其中， D 是内生变量， X 是外生变量

- D 有一个工具变量 Z , 满足以下两个条件：

1 外生性

- Z 与 e 不相关，即控制了所有解释变量（内生变量和外生变量）后， Z 对 Y 没有作用。外生性是一个“干净”的工具变量要求，即当控制了外生变量后， Z 对 Y 的影响只能通过内生变量来实现，当内生变量也被控制后， Z 对 Y 就没有了作用。因此外生性也被称为“排他性”。

两阶段最小二乘法

2 相关性

- 当控制了模型里所有外生变量与内生变量后, Z 与 D 仍然存在相关性。

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

其中, $\gamma_1 \neq 0$

- 图 (a) 显示控制外生变量后, 内生变量 D 与 e 仍然相关。图 (b) 显示控制外生变量后, $Z_1 \rightarrow D_1$ (相关性); $Z_1 \rightarrow D_1 \rightarrow Y$ (外生性)。
- 注意: $Z_1 \rightarrow D_1 \leftarrow e \rightarrow Y$ 是对撞路径; $Z_1 \rightarrow [X_1, \dots, X_k] \rightarrow Y$ 是截断路径。

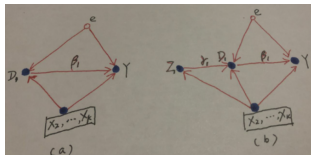


Figure: 控制其他外生变量情况下的工具变量路径图

两阶段最小二乘法

- 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:

$$\underbrace{D_{1i}}_{\text{内生变量}} = \gamma_0 + \underbrace{\gamma_1 Z_{1i}}_{IV} + \underbrace{\gamma_2 X_{2i} + \cdots + \gamma_k X_{ki}}_{\text{所有外生变量}} + \nu_i$$

- 利用所得的估计系数 $(\hat{\gamma}_0, \hat{\gamma}_1, \cdots + \hat{\gamma}_k)$ 计算内生变量的预测值:

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

- 第二阶段:
 - 将预测值 \hat{D}_{1i} 替代内生变量 D_i , 并进行回归:

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

- 得到的系数即为样本两阶段最小二乘法系数。

两阶段最小二乘法

- 多个内生变量和多个工具变量，假设模型为：

$$Y_i = \alpha + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + e_i \quad (4)$$

其中， D_{1i}, D_{2i} 是内生变量。 D_{1i} 有一个工具变量 Z_{1i} ； D_{2i} 有两个工具变量 Z_{2i} 、 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进行回归：

$$\underbrace{D_{1i}}_{\text{内生变量}} = \gamma_0 + \underbrace{\gamma_1 Z_{1i} + \gamma_2 Z_{2i} + \gamma_3 Z_{3i}}_{\text{所有 IV}} + \underbrace{\gamma_4 X_{3i} \cdots + \gamma_k X_{ki}}_{\text{所有其他外生变量}} + \nu_{1i}$$

$$\underbrace{D_{2i}}_{\text{内生变量}} = \theta_0 + \underbrace{\theta_1 Z_{1i} + \theta_2 Z_{2i} + \theta_3 Z_{3i}}_{\text{所有 IV}} + \underbrace{\theta_4 X_{3i} \cdots + \theta_k X_{ki}}_{\text{所有其他外生变量}} + \nu_{2i}$$

- 用得到的系数计算每个内生变量的预测值： \hat{D}_{1i} 、 \hat{D}_{2i} 。

两阶段最小二乘法

- 第二阶段:
 - 用内生变量的预测值替代模型中的内生变量并进行回归:

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

- 得到的 $\hat{\beta}_1^{2sls}$ 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

- 大样本下的局限性：偏差性

- $\hat{\beta}_1^{2sls}$ 的大样本概率极限值为：

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

- 当工具变量完全外生时，2SLS 大样本偏差项 = 0。
- 当工具变量不完全是外生的，即使偏差项的分子很小，如果分母也很小（IV 与内生变量的相关性很小），这个偏差也会被放的很大。
- 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:

$$\frac{p \lim \hat{\beta}_1^{2sls} - \beta_1}{p \lim \hat{\beta}_1^{ols} - \beta_1} = \frac{Cov(Z_i, e_i)}{Cov(D_i, Z_i)} \times \underbrace{\frac{Var(D_i)}{Cov(D_i, e_i)}}_{OLS}$$

- 如果工具变量不完全外生, 这个比率大于 1, 则 $\hat{\beta}_1^{2sls}$ 的大样本偏差大于 $\hat{\beta}_1^{ols}$ 的大样本偏差。

工具变量估计法的局限性

- 大样本下的局限性：有效性

- 工具变量两阶段估计量 $\hat{\beta}_1^{2sls}$ 是渐进正态分布：

$$\hat{\beta}_1^{2sls} \xrightarrow{d} N(\beta_1, Avar(\hat{\beta}_1^{2sls}))$$

- 在同方差情况下，其渐进方差为：

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

其中， ρ_{DZ} 是工具变量与内生变量的相关系数。

- 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：

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

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：

$$\frac{\text{Avar}(\hat{\beta}_1^{2sls})}{\text{Avar}(\hat{\beta}_1^{ols})} = \frac{1}{\rho_{DZ}^2}$$

- ρ_{DZ}^2 小于 1，估计供便利估计值的方差总是大于 OLS 估计值的方差。因为工具变量只是用了内生变量与工具变量相关的一部分信息，而 OLS 使用了全部信息。
- 在弱工具变量的情况下， ρ_{DZ}^2 很小，因此能分解出的内生变量“好”的信息很少，造成方差很大，此时显著性并不可靠。

工具变量估计法的局限性

- 有限样本下的局限性：偏差性

- 在有限样本中， $\hat{\beta}_1^{2sls}$ 的偏差（Hahn and Hausman, 2002）为：

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

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， R^2 为工具变量的拟合优度，它是一阶段回归模型中工具变量对内生变量的解释力度。

- 在有限样本量（ N 较小），如果工具变量是弱工具变量（ R^2 较小），2SLS 估计量的偏差比 OLS 估计量的偏差还大。因此，在有限样本里，工具变量并不比内生性的 OLS 估计更准确。

$$\frac{E(\hat{\beta}_1^{2sls}) - \beta_1}{E(\hat{\beta}_1^{ols}) - \beta_1} = \frac{K\rho}{NR^2}$$

工具变量估计法的局限性

- 有限样本下的局限性：有效性
 - 在有限样本中，我们不知道 $\hat{\beta}_1^{2sls}$ 的分布，因此通常使用 t 检验在小样本里不适用。

工具变量运用的检验

- 内生性检验

- 1 Durbin-Wu-Hausman χ^2_J 检验

- 如果可能的内生变量有一个合理的工具变量，构造下面的统计量：

$$H = (\hat{\beta}_1^{2sls} - \hat{\beta}_1^{ols})' [Avar(\hat{\beta}_1^{2sls}) - Avar(\hat{\beta}_1^{ols})]^{-1} (\hat{\beta}_1^{2sls} - \hat{\beta}_1^{ols}) \sim \chi^2_J$$

- DWH 检验的原假设 H_0 : D 是外生的。在原假设下，OLS 和工具变量得到的参数估计量是一致的。在备择假设下，OLS 估计量是有偏的，此时 H 值远远不等于 0。

工具变量运用的检验

2 Wu-Hausman F 检验

- 该检验基于回归形式，使用较为方便。

$$Y_i = \alpha + \beta_1 D_i + e_i$$
$$D_i = \gamma_0 + \gamma_1 Z_i + \mu_i$$

- 检验 D 是外生性就是检验 $Cov(u_i, e_i)$ 是否为零。
- 两个干扰项的线性关系写为：

$$e_i = \rho \mu_i + \tau_i$$

- 如果二者不相关，意味着 $\rho = 0$ ，因此原假设 $H_0: \rho = 0$ 。

工具变量运用的检验

- 将上式代入原估计方程，可得：

$$Y_i = \alpha + \beta_1 D_i + \rho \mu_i + \tau_i$$

- 由于 μ_i 是不可观测的，可以将第一阶段回归得到的残差项 $\hat{\mu}_i = D_i - \hat{\gamma}_0 - \hat{\gamma}_1 Z_i$ 代入模型，从而可得最终用来检验的回归方程：

$$Y_i = \alpha + \beta_1 D_i + \rho \hat{\mu}_i + \varepsilon_i$$

- 如果 ρ 显著不等于 0，就拒绝变量 D 是外生性变量的假设。

工具变量运用的检验

- 弱工具变量检验

- 如果只有一个内生变量：观察 2SLS 中第一阶段关于所有工具变量的系数同时为 0 的 F 检验，如果第一阶段得到的 F 统计量非常低，则表示存在弱工具变量问题。当工具变量数量为 1,2,3,5,10 时，第一阶段 F 统计量分别设为 8.96,11.59,12.83,15.09,22.88(Stock,Wright and Yogo,2002)。如果第一阶段 F 值低于这些关键值，则可能存在弱工具变量问题。
- 如果存在多个内生变量和多个工具变量，Stock and Yogo (2005) 提供了一个检验方法（略）。这种方法的思路是先计算一个 Minimum Eigenvalue 的统计量，如果这个统计量高于 Stock and Yogo (2005) 提供的对应关键值，则不存在弱工具变量问题。

工具变量运用的检验

- 过度识别检验

- 工具变量的另一个条件是外生性。本质而言，我们是不可能检验这个条件的，因为干扰项无法被观测。但我们可以在一定程度上对外生条件进行检验。
- 恰好识别：无法检验
- 因为，当我们去估计 $\hat{e}_i = Y_i - \hat{\beta}_1 D_i$ 时，需要满足外生条件（已经假设了）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 $Cov(Z_i, \hat{e}_i) = 0$ 检验工具变量是否外生是没有意义的。
- 简言之，在恰好识别的情况下，工具变量的信息已经用在内生变量的系数估计上了，不能重复使用来检验其本身是否为外生。

工具变量运用的检验

- **过度识别：**假设有一个内生变量和两个工具变量。
 - 先假设第一个工具变量 Z_1 是外生的，并只用 Z_1 作为工具变量得到系数估计值 $\hat{\beta}_1^{Z_1}$ ，那么残差 $\hat{e}_i^{Z_1} = Y_i - \hat{\beta}_1^{Z_1} D_i$ 是 e 的一致估计量。接着使用值 $\hat{\beta}_1^{Z_1}$ 对第二个工具变量 Z_2 的外生性进行检验。
 - 这个检验方法的局限性是必须先假设一个工具变量满足外生性的假设成立的前提下，才能检验另一个变量的外生性。
 - 如果过度识别检验没有通过，虽然无法知道哪个变量是内生的，但是至少知道有一个工具变量是内生的结论。
 - 如果一个工具变量通过了外生检验，并不能说明 Z_1 一定是外生的，因为可能是另一个工具变量 Z_2 的外生性假设错误导致 Z_1 通过了外生性检验。

工具变量运用的检验

- 在实际操作中，我们的操作步骤是：
 - 1 原假设为所有工具变量都是外生的；
 - 2 先使用所有工具变量进行 2SLS 回归得到残差。如果所有工具变量都是外生的，残差 \hat{e}_i 就是干扰项 e 的一致估计量。
 - 3 将残差作为被解释变量，所有工具变量作为解释变量，用 OLS 进行回归得到 R^2 。如果所有工具变量都是外生的，那么它们与残差是无关的， R^2 会较小。
 - 4 进行假设检验。在原假设下 $NR^2 \sim \chi_q^2$ (q 是自由度 = 工具变量数 - 内生变量数)。如果 NR^2 大于相关的 χ_q^2 关键值，得到不是所有工具变量都是外生的。如果 NR^2 小于相关的 χ_q^2 关键值，通过了过度识别检验，但不能确定所有工具变量都是外生的。

工具变量运用的检验

- 如果使用 2SLS 估计量，但使用同方差（没有使用选项 robust VCE），在 STATA 中会报告 χ^2_q 统计值。
- 如果使用 2SLS 估计量，但考虑异方差和自相关，在 STATA 中会报告 Wooldridge 的稳健得分过度识别检验。
- 如果使用 IV-GMM 估计量，则报告 Hansen(1982) 的 J 统计值。

工具变量操作步骤

- 工具变量的使用步骤

- 1 清晰定义研究问题，描述经济机制，设定基本模型，对基本模型进行 OLS 回归，理解并描述模型可能存在的内生性问题。
- 2 根据经济机制和理论基础选择有效的工具变量，并解释为什么选择工具变量是相关的和外生的（很关键）。
- 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，并谨慎对结果进行解释。
 - 检验变量内生性：Hausman 检验；
 - 检验工具变量相关性：报告第一阶段的 F 统计量。检验是否存在弱工具变量问题；
 - 检验工具变量的外生性：过度识别检验。
- 4 将工具变量估计结果与 OLS 估计结果进行比较，理解为何结果存在差异。

工具变量操作步骤

- 实例操作

- Acemoglu. The Colonial Origins of Comparative Development : An Empirical Investigation[J]. AER,2001,91(5):1369-1401.
- 研究问题：好的社会制度对经济发展是否有促进作用
- 经济机制：良好的社会制度意味着更好的产权保护和更少的扭曲资源配置的政策，会促进固定资产和人力资源的投入，并更有效率地产出
- 基本模型：

$$\log GDP_i = \alpha + \beta_1 avexpr + \beta_2 lat_abst + e_i$$

其中，*avexpr* 表示 1985-1995 年企业免受政府盘剥的指数平均值；
lat_abst：首都的纬度（地理位置）

工具变量操作步骤

变量的描述

variable name	storage type	display format	value label	variable label
lat_abst	float	%9.0g		Abs(latitude of capital)/90
euro1900	float	%9.0g		European settlers 1900, AJR
excolony	float	%9.0g		=1 if was colony FLOPS definiti
avexpr	float	%9.0g		average protection against expropriation risk
logpgp95	float	%9.0g		log PPP GDP pc in 1995, World Bank
cons1	float	%9.0g		cons on exec in 1st year indep
indtime	float	%9.0g		years independent: 1995 minus firstyr
democ00a	float	%9.0g		democracy in 1900
cons00a	float	%9.0g		constraint on executive in 1900
extmort4	float	%9.0g		corrected mort.
logem4	float	%9.0g		log settler mortality

OLS 回归结果

```
. regress logpgp95 avexpr lat_abst,robust
```

Linear regression

Number of obs = 111
F(2, 108) = 183.95
Prob > F = 0.0000
R-squared = 0.6225
Root MSE = .7108

logpgp95	Robust			P> t	[95% Conf. Interval]	
	Coef.	Std. Err.	t			
avexpr	.4634816	.0521728	8.88	0.000	.360066	.5668972
lat_abst	.8721613	.4993736	1.75	0.084	-.1176837	1.862006
_cons	4.872922	.2807513	17.36	0.000	4.316424	5.42942

- OLS 可能存在的问题
 - 发展较好的经济体更有可能建立良好的制度（反向因果）
 - 文化差异等其他变量同时影响国家经济状况和制度（混淆路径）
 - 社会制度不易准确衡量，测量可能存在较大偏差（测量误差）

工具变量操作步骤

- 寻找有效的工具变量

- 相关性：国家现在的社会制度一定程度上是过去制度的延续，早期社会制度又与欧洲殖民者的殖民政策有关。
- 外生性：殖民者死亡率的外生性最强。例如，早期定居人数可能反映定居者更喜欢迁移到富裕地区并允许现有经济发展；早期社会制度和现在经济发展水平也有可能同时受到某些地域特征一些而存在相关关系。

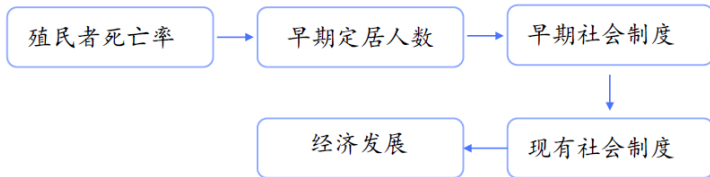


Figure: 国家现有社会制度的追根溯源

工具变量操作步骤

- 使用工具变量法对模型进行估计

- 第一阶段结果

```
. ivregress 2sls logpgp lat_abst (avexpr = logem4), first
```

First-stage regressions

第一阶段报告结果

被解释变量

内生变量

工具变量

其他外生解释变量

```
Number of obs   =      70
F(   2,    67)   =     19.53
Prob > F         =     0.0000
R-squared        =     0.3682
Adj R-squared    =     0.3494
Root MSE        =     1.2523
```

avexpr	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lat_abst	3.125466	1.203964	2.60	0.012	.7223438	5.528588
logem4	-.4537058	.1304823	-3.48	0.001	-.7141496	-.1932619
_cons	8.094943	.7590112	10.67	0.000	6.57995	9.609936

工具变量操作步骤

- 使用工具变量法对模型进行估计
 - 第二阶段结果

Instrumental variables (2SLS) regression

第二阶段结果

Number of obs = 70
Wald chi2(2) = 39.18
Prob > chi2 = 0.0000
R-squared = 0.0670
Root MSE = 1.0159

内生变量	logpgp95	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	avexpr	1.029084	.2332977	4.41	0.000	.5718288	1.486339
其他外生变量	lat_abst	-1.784366	1.494275	-1.19	0.232	-4.713092	1.144359
	_cons	1.65175	1.322986	1.25	0.212	-.9412546	4.244754

Instrumented: avexpr
Instruments: lat_abst logem4

工具变量操作步骤

- Hausman 检验解释变量是否外生
 - 两种方法都拒绝了内生变量是外生的假设。

```
. estat endogenous
```

Tests of endogeneity

Ho: variables are exogenous

Durbin (score) $\chi^2(1)$ = 16.4466 (p = 0.0001)

Wu-Hausman F(1,66) = 20.2691 (p = 0.0000)

卡方检验

F检验

工具变量操作步骤

- 检验工具变量是否为弱工具变量

- $F=12.09 > 10$;

Stock and Yogo 给的 Minimum Eigenvalue=8.96 < 16.38。

- 这说明这个工具变量在能否确定为弱工具变量的边际上。

```
. estat firststage
```

First-stage regression summary statistics					
Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	F(1,67)	Prob > F
avexpr	0.3682	0.3494	0.1529	12.0905	0.0009

$F=12 > 10$

Minimum eigenvalue statistic = 12.0905

Stock an Yogo提供的检验方法

Critical Values
Ho: Instruments are weak

不常用，需要工具变量至少比内生变量多2个

of endogenous regressors: 1
of excluded instruments: 1

2SLS relative bias	5%	10%	20%	30%
	(not available)			
2SLS Size of nominal 5% Wald test	10%	15%	20%	25%
LIML Size of nominal 5% Wald test	16.38	8.96	6.66	5.53

工具变量操作步骤

- 过度识别情况下，检验工具变量是否是外生的
 - 采用殖民者死亡率和定居人数同时作为工具变量进行估计
 - 第一阶段结果

```
. ivregress 2sls logpgp lat_abst (avexpr = logem4 euro1900),first
```

First-stage regressions

新增加的IV

```
Number of obs      =          69
F(   3,          65) =        15.07
Prob > F            =         0.0000
R-squared           =         0.4103
Adj R-squared       =         0.3830
Root MSE           =         1.2271
```

avexpr	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lat_abst	1.845078	1.355143	1.36	0.178	-.8613306	4.551486
logem4	-.3197937	.1453475	-2.20	0.031	-.6100726	-.0295148
euro1900	.0160626	.0077541	2.07	0.042	.0005765	.0315487
_cons	7.458333	.8117933	9.19	0.000	5.83707	9.079597

工具变量操作步骤

- 过度识别情况下，检验工具变量是否是外生的
 - 第二阶段结果
 - 可以看到，avexpr 的系数为 0.97，与只使用一个工具变量的结果相近。

Instrumental variables (2SLS) regression

第二阶段结果

Number of obs = 69
Wald chi2(2) = 47.23
Prob > chi2 = 0.0000
R-squared = 0.1599
Root MSE = .96035

logpgp95	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
avexpr	.9735225	.1886416	5.16	0.000	.6037918	1.343253
lat_abst	-1.629427	1.291443	-1.26	0.207	-4.160609	.9017557
_cons	1.975545	1.073346	1.84	0.066	-.1281749	4.079265

Instrumented: avexpr

Instruments: lat_abst logem4 euro1900

工具变量操作步骤

- 过度识别情况下，检验工具变量是否是外生的
 - 下面使用 STATA 的 `estat overid` 命令来检验过度识别检验。
 - 两种方法的卡方值都不显著，说明无法拒绝所有工具变量都是外生的假设。

```
. estat overid
```

Tests of overidentifying restrictions:

Sargan (score) $\chi^2(1) = .104916$

Basmann $\chi^2(1) = .098985$

(p = 0.7460)

(p = 0.7531)

都不显著

- 解释工具变量的经济显著性，并比较 OLS 结果
 - 2SLS 的回归系数为 0.97 比 OLS 的回归系数 0.46 高，说明 OLS 有向下偏差。很可能是测量误差造成的两个系数的差别。
 - 例如，尼日利亚的现有制度较差，智利和尼日利亚之家在制度指数上的差距会导致 GDP 产生大约 7 倍的差异。说明制度对经济发展水平有较大实质性的影响。

工具变量运用常见问题

- 1 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
- 2 第一阶段回归的解释变量应包含所有的外生变量
- 3 避免用组均值作为工具变量
- 4 避免用内生变量的滞后项作为工具变量
- 5 模型包含有二次项的工具变量的用法
- 6 模型存在交互项时工具变量用法
- 7 理解工具变量估计结果只是局部平均处置效应
- 8 工具变量越多越好吗？
- 9 工具变量是解决内生性的万灵药吗？

工具变量运用常见问题

- 1 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
 - 自己手动进行两步回归可能得到错误的标准误差。

$$\text{Avar}(\hat{\beta}_1^{2sls}) = \hat{\sigma}_e^2 (\hat{D}' \hat{D})^{-1}$$

$$\hat{\sigma}_e^2 = \frac{\hat{e}' \hat{e}}{N}$$

$$\hat{e} = Y - D' \hat{\beta}_1^{2sls}$$

- 在计算 \hat{e}_i 时用的是内生变量的原值 D' ，而不是预测值 \hat{D}' 。如果手动计算第二阶段回归，就容易犯直接使用第二阶段回归残差的错误 $\hat{e} = Y - \hat{D}' \hat{\beta}_1^{2sls}$ 。

工具变量运用常见问题

2 第一阶段回归的解释变量应包含所有的外生变量

- 在第一阶段回归不能只是用工具变量作为解释变量进行回归。
- 第一阶段回归没有加入其他外生变量作为控制变量，造成残差项中包含其他外生变量的信息，并且进入第二阶段的干扰项，进而造成第二阶段的其他外生变量与干扰项相关。

工具变量运用常见问题

3 避免用组均值作为工具变量

- 这里的组可以是行业、学校、省等。一些文献的理由是：组内个体特征与组内其他个体的平均或总特征有关，但是组内其他个体的平均或加总特征不直接影响个体的结果。
- 这个方法看起来是很有道理，但实际上通常是错误的，原因是组均值作为工具变量不满足外生性要求（Gormley and Matsa, 2014）。

$$Y_i = \alpha + \beta_1 X_i + \underbrace{e_i}_{Industry_i + v_{ij}}$$

- 通常一个行业内的企业特性受到行业固定效应的影响。如果行业内每个行业的特征都与行业固定效应相关，它们的特征均值必然也与行业固定效应相关。即使在模型里加入行业的固定效应，也不能解决这个问题（均值与行业固定效应几乎是共线的）。

4 避免用内生变量的滞后项作为工具变量

- 如果内生变量 X 是序列相关的, 那么 $Cov(X_{it}, X_{it-1}) \neq 0$ 。但是, 如果干扰项也存在时间序列相关,

$$Cov(X_{it-1}, e_t) = Cov(X_{it-1}, \rho e_{i,t-1} + \zeta_t) = \rho Cov(X_{it-1}, e_{i,t-1}) \neq 0,$$
那么外生性不满足。只有当 $\rho = 0$, 滞后项作为工具变量才满足外生性条件。

5 模型包含有二次项的工具变量的用法

- 错误的做法是，用手动法在第一步拿一个工具变量 Z 来计算拟合值 \hat{X} ，然后用 \hat{X} 来计算 \hat{X}^2 ，再把 \hat{X} 、 \hat{X}^2 代入第二步回归。
- 正确做法是，用 Z 和 Z^2 作为 X 和 X^2 的工具变量，使用计量软件估计模型。

工具变量运用常见问题

6 模型存在交互项时工具变量用法

- 正确做法是， Z 作为 X_1 的工具变量， ZX_2 作为 X_1X_2 的工具变量。

工具变量运用常见问题

7 理解工具变量估计结果只是局部平均处置效应

- OLS 参数估计值使用了解释变量所有的信息去估计被解释变量的变化，因此 OLS 估计得到的系数是平均处置效应。
- 工具变量估计系数是使用了与工具变量相关的信息去估计被解释变量的变化，得到的系数描述的是对于那些其特征 X 会受到工具变量 Z 影响的个体。局部指的是一部分个体。
- 工具变量的估计结果是局部平均处置效应也以意味着选择不同的工具变量会得到不同的估计结果。

8 工具变量越多越好吗？

- 找到更多的好的工具变量可以提高有效性，使用差的工具变量会放大估计偏差。
- 使用多工具变量要权衡有效性和偏差项。
- 避免添加弱工具变量或同质工具变量。
- 比较不同工具变量的估计结果来决定是否添加某个工具变量。

工具变量运用常见问题

9 工具变量是解决内生性的万灵药吗？

- 好的工具变量可以有效解决内生性问题。
- 外生性无法通过统计检验确认。
- 在实际运用中，外生性和相关性很难兼得。

参考文献:

Angrist and Pischke(2008). Mostly harmless econometrics: An empiricist' s companion. Princeton university press.

连玉君, Stata 讲义或者视频资料.