- 第三章中，我们假设了零条件均值假设（MLR. 4）：

$\hat{\sigma}^2 = \frac{1}{n}\sum \hat{u}_i^2$

$$\boxed{E(u_i|x_{i1}, x_{i2}, \ldots, x_{ik}) = 0}$$

u

- 解释变量与误差项不相关被叫做外生的（exogenous）；外生性服从MLR. 4

$cov(x, u) = 0$

- 解释变量与误差项的相关被叫做内生的（endogenous）；内生性违背了MLR. 4
- 内生性会导致OLS不具有无偏性与一致性
- 在这一章中，我们讨论内生性下模型的估计问题

Chapter 15

工具变量估计与两阶段最小二乘法

章节框架

- 在这一章中，我们将探讨内生性下模型的估计问题
- 首先，我们介绍内生性可能的原因以及工具变量的定义
- 之后，我们讨论如何利用工具变量构造估计值
- 最后，我们提出检验内生性的方法 前提条件.

动机：简单回归模型中的遗漏变量

- **内生性问题在社会科学和经济学中很普遍**
  - (1)遗漏变量（omitted variables）
  - 例子：工资方程中的教育

$$\log(wage_i) = \beta_0 + \beta_1 educ_i + u_i$$

误差项包含包含一些因素，例如与教育相关的智力，工作经验等

  - (2)测量误差（measurement error）
  - 例子：NBA球员工资与实力（不能被准确观测）
  - (3)~~瞬时~~性（simultaneity）
  - 例子：警察密度与犯罪率

Friedman 1957.

同时.

x是y的原因，
y也是x的原因.

(1997).

Levitt.

Freakonomics

1. Simultaneity bias

(1928) (1950) 因果

联立方程 1980~1990 Angrist.

$$\begin{cases} q_t^d = \alpha_0 + \alpha_1 p_t + u_t & \text{(需求)} \\ q_t^s = \beta_0 + \beta_1 p_t + v_t & \text{(供给)} \\ q_t^d = q_t^s & \text{(均衡)} \end{cases}$$

Instrumental Variable (IV)

天气

生产技术问题

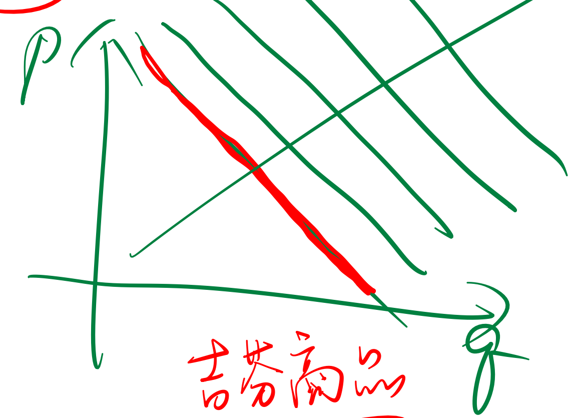
需求、需求量

令 $q_t \equiv q_t^d = q_t^s$，可得

$\alpha_1 < 0$

$$\begin{cases} q_t = \alpha_0 + \alpha_1 p_t + u_t \\ q_t = \beta_0 + \beta_1 p_t + v_t \end{cases}$$

$\beta_1 > 0$



两个方程中的被解释变量与解释变量完全一样。

如直接作回归 $q_t \xrightarrow{\text{OLS}} p_t$，估计的是需求函数还是供给函数？

Working (1927)

把线性方程组中的(p_t, q_t)看成是未知数(内生变量)，把(u_t, v_t)看作已知，可求解(p_t, q_t)为(u_t, v_t)的函数：

$$\operatorname{Cov}(p_t, u_t) = \frac{-\operatorname{Var}(u_t)}{\alpha_1 - \beta_1} = \frac{\operatorname{Var}(u_t)}{\beta_1 - \alpha_1} > 0$$

$$\begin{cases} p_t = p_t(u_t, v_t) = \dfrac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \dfrac{v_t - u_t}{\alpha_1 - \beta_1} \\ q_t = q_t(u_t, v_t) = \dfrac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} + \dfrac{\alpha_1 v_t - \beta_1 u_t}{\alpha_1 - \beta_1} \end{cases}$$

$$\operatorname{Cov}(p_t, v_t) = \frac{\operatorname{Var}(v_t)}{-(\beta_1 - \alpha_1)} < 0$$

由于p_t为(u_t, v_t)的函数，故$\operatorname{Cov}(p_t, u_t) \neq 0$，$\operatorname{Cov}(p_t, v_t) \neq 0$。

OLS 估计值$\hat{\alpha}_1, \hat{\beta}_1$不是$\alpha_1, \beta_1$的一致估计量。

$q_t \xrightarrow{OLS} p_t$， 回归系数(需求) $= \dfrac{\operatorname{Cov}(q, p)}{\operatorname{Var}(p)} = \dfrac{\operatorname{Cov}(p_t, \alpha_0 + \alpha_1 p_t + u_t)}{\operatorname{Var}(p_t)}$

$= \alpha_1 + \boxed{\dfrac{\operatorname{Cov}(p_t, u_t)}{\operatorname{Var}(p_t)}} > \alpha_1$

(供给) $\beta_1 + \dfrac{\operatorname{Cov}(p_t, v_t)}{\operatorname{Var}(p_t)} < \beta_1$

既不是供给，也不是需求

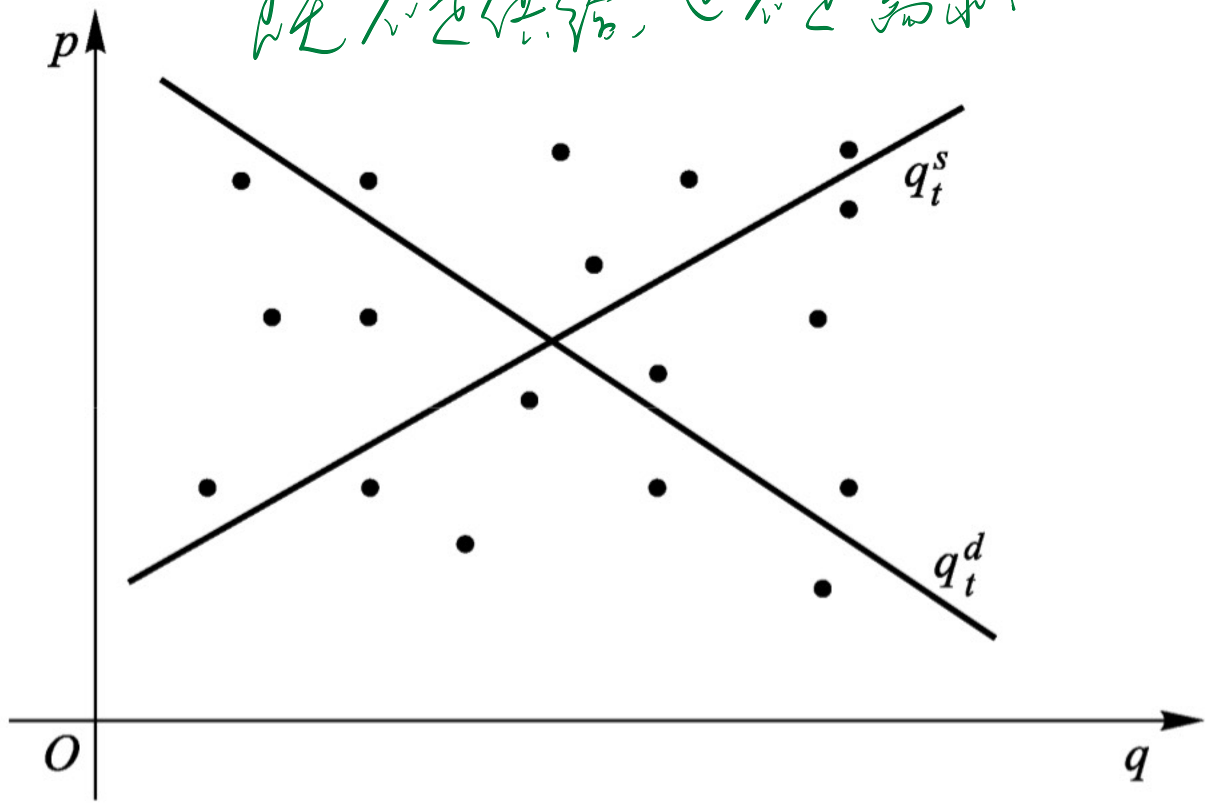


图 10.1　需求与供给决定市场均衡

2. Omitted variable problem

OLS估计量的期望值

真实：$y=\beta_0+\beta_1x_1+\beta_2x_2+v$. $E[v|x_1,x_2]=0$

- **回归中包含无关变量**

$E[y|x_1,x_2,x_3]=E[y|x_1,x_2]$

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+u$$

$=\beta_0+\beta_1x_1+\beta_2x_2$.

= 0 在总体中

没问题，因为 $E(\hat{\beta}_3)=\beta_3=0$

$E[u|x_1,x_2,x_3]\overset{?}{=}0$

但是.

不用管→然而，包含无关变量可能会增加抽样的方差（3.4小节）

$=E[y-\beta_0-\beta_1x_1-\beta_2x_2|x_1,x_2]$

- **遗漏变量：简单情形**

因此 $E(\hat{\beta}_3)=\beta_3=0$.

$$y=\beta_0+\beta_1x_1+\beta_2x_2+u$$ ← 真实样本(包含x_1和x_2)

$$\tilde{y}=\tilde{\beta}_0+\tilde{\beta}_1x_1$$ ← 估计模型(x_2是遗漏的)

① $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_{k-1} x_{ik-1} + \hat{\beta}_k x_{ik} + \hat{u}_i$

② $y_i = \tilde{\beta}_0 + \tilde{\beta}_1 x_{i1} + \cdots + \tilde{\beta}_{k-1} x_{ik-1} + \tilde{u}_i$

辅助回归： $x_{ik} = \tilde{\delta}_0 + \tilde{\delta}_1 x_{i1} + \tilde{\delta}_2 x_{i2} + \cdots + \tilde{\delta}_{k-1} x_{ik-1} + \tilde{v}_i$

结论： $\tilde{\beta}_j = \hat{\beta}_j + \hat{\beta}_k \tilde{\delta}_j$ $j=0,\cdots,k-1$

对 $j=1$ 证明. $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_k \tilde{\delta}_1$

证明：辅助回归. $x_{i1} \sim 1, x_{i2}, \cdots, x_{ik-1}$，得残差 $\tilde{r}_{i1}$

$\check{\gamma}_0, \check{\gamma}_2, \cdots \check{\gamma}_{k-1}$

$$\tilde{\beta}_1 = \left(\sum_{i=1}^{n} \tilde{r}_{i1}^2\right)^{-1} \left(\sum_{i=1}^{n} \tilde{r}_{i1} y_i\right)$$

$\sum \tilde{r}_{i1} = 0$, (i) $\sum x_{i2} \tilde{r}_{i1} = \cdots =$

$\sum x_{ik-1} \tilde{r}_{i1} = 0$ (ii)

$\tilde{r}_{i1} = x_{i1} - \check{\gamma}_0 - \check{\gamma}_2 x_{i2} - \cdots - \check{\gamma}_{k-1} x_{ik-1}$

(iii)

$x_{i1} = \tilde{r}_{i1} + \check{\gamma}_0 + \check{\gamma}_2 x_{i2} + \cdots + \check{\gamma}_{k-1} x_{ik-1}$

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik} + \hat{u}_i$$

$$\sum_{i=1}^{n} \hat{\tilde{r}}_{i1} y_i = \sum_{i=1}^{n} \hat{\tilde{r}}_{i1} (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik} + \hat{u}_i)$$

(i) $\sum \hat{\tilde{r}}_{i1} \hat{\beta}_0 = 0$, (ii) $\sum \hat{\tilde{r}}_{i1} x_{i2} = \sum \hat{\tilde{r}}_{i1} x_{i3} = \cdots = \sum \hat{\tilde{r}}_{i1} x_{ik-1} = 0$

$\sum \hat{\tilde{r}}_{i1} \hat{u}_i = 0$，因为 $\hat{u}_i$ 和 $1, x_{i1}, \cdots, x_{ik}$ 有 Normal Equation

$$= \sum \hat{\tilde{r}}_{i1} \hat{\beta}_1 x_{i1} + \sum \hat{\tilde{r}}_{i1} \hat{\beta}_k x_{ik}$$

$$= \hat{\beta}_1 \left(\sum \hat{\tilde{r}}_{i1} x_{i1}\right) + \hat{\beta}_k \left(\sum \hat{\tilde{r}}_{i1} x_{ik}\right)$$

(iii) $\sum \hat{\tilde{r}}_{i1} x_{i1} = \sum \hat{\tilde{r}}_{i1}^2$

故 $$\tilde{\beta}_1 = \left(\sum \hat{\tilde{r}}_{i1}^2\right)^{-1} \cdot \left(\hat{\beta}_1 \sum \hat{\tilde{r}}_{i1}^2 + \hat{\beta}_k \left(\sum \hat{\tilde{r}}_{i1} x_{ik}\right)\right)$$

$$= \hat{\beta}_1 + \hat{\beta}_k \underbrace{\left(\sum \hat{\tilde{r}}_{i1}^2\right)^{-1} \left(\sum \hat{\tilde{r}}_{i1} x_{ik}\right)}_{} \triangleq \tilde{\delta}_1 = \hat{\beta}_1 + \hat{\beta}_k \tilde{\delta}_1$$

$\tilde{\delta}_1$理解：x_j在辅助回归中去掉其它$x_{(j)}$影响的残差，和x_k回归。若x_j和$x_{(j)}$无关，则残差是x_j，此时$\tilde{\delta}_j$成为x_j和x_k之间的回归系数，也即是x_j和x_k的相关。

仅有$cov(x_{ik}, x_{ij})=0$，不足以让$\tilde{\delta}_j=0$。

OLS估计量的期望值

- **遗漏变量偏误**

$y = \beta_0 + \beta_1 x_1 + \eta_t$　　$X_1 = [1 \ x_1]$

如果x_1和x_2是相关的，假设一个线性回归关系

$$x_2 = \delta_0 + \delta_1 x_1 + v$$

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$

$$\Rightarrow \quad y = \beta_0 + \beta_1 x_1 + \beta_2(\delta_0 + \delta_1 x_1 + v) + u$$

$= [X_1 \ x_2]\begin{bmatrix}\beta \\ \beta_2\end{bmatrix} + u$

$$= \boxed{(\beta_0 + \beta_2\delta_0)} + \boxed{(\beta_1 + \beta_2\delta_1)}x_1 + \boxed{(\beta_2 v + u)}$$

$y = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \tilde{u}$

$\tilde{\beta} = (X_1'X_1)^{-1}(X_1'y)$

$= (X_1'X_1)^{-1}(X_1' \cdot (X_1\beta + x_2\beta_2 + u))$

如果y只对x_1做回归，这就是截距项

如果y只对x_1做回归，这就是斜率项

误差项

MR4?　MR4'?

- **结论：所有估计系数都是偏误的（All estimated coefficients will be biased）**

$\mathrm{cov}(x, u) \neq 0$

$= \beta + (X_1'X_1)^{-1}(X_1'x_2\beta_2)$

$\rightarrow \beta + \beta_2 \boxed{E(x_1x_1')^{-1} E(x_1x_2)}$

只是简单讨论
实际做项目时，因通常有多个变量，这讨论无用

- 当遗漏变量x_2时β_1的估计偏误情况汇总：

	Corr$(x_1, x_2) > 0$ $(\delta_1 > 0)$	Corr$(x_1, x_2) < 0$ $(\delta_1 < 0)$
$\beta_2 > 0$	Positive bias $\hat{\beta}_K \tilde{\delta}_1 > 0$	Negative bias $\hat{\beta}_K \tilde{\delta}_1 < 0$
$\beta_2 < 0$	Negative bias $\hat{\beta}_K \tilde{\delta}_1 < 0$	Positive bias $\hat{\beta}_K \tilde{\delta}_1 > 0$

此时 $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_K \tilde{\delta}_1$

$x_K = \tilde{\delta}_0 + \tilde{\delta}_1 x_1 + \tilde{v}$.

- 思考：

（1）~~如果x_1和x_2是相关的，哪条基本假设不成立~~？ 没什么不成立.

（2）如果x_1和x_2是不相关的呢？ 见上页$\tilde{\delta}_j$理解.

OLS估计量的期望值

- **例子：工资等式中遗漏能力**

$$wage = \beta_0 + \beta_1 educ + \beta_2 abil + u$$

$$abil = \delta_0 + \delta_1 educ + v$$

都是正的

$$wage = (\beta_0 + \beta_2\delta_0) + \boxed{(\beta_1 + \beta_2\delta_1)} educ + (\beta_2 v + u)$$

β_1 将被过度估计（overestimated）因为 $\beta_2\delta_1 > 0$ 。这看起来好像受到多年教育的人会挣很高的工资，但这部分是由于平均来说受教育长的人能力也很强。

- **什么么时候没有遗漏变量偏误？**
  - 如果遗漏变量与解释变量是不相关的

OLS估计量的期望值

$x_3 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + r$

- **遗漏变量偏误：更一般的情形**

代入：$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + r) + u$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

真实模型(包含x_1, x_2, 和x_3)

$= (\beta_0 + \beta_3\delta_0) + (\beta_1 + \beta_3\delta_1)x_1$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + w$$

估计模型(x_3是遗漏的)

$+ (\beta_2 + \beta_3\delta_2)x_2$

$+ (\beta_3 r + u)$

  - 很难得到偏误的方向，这是因为x间会两两相关
  - 如果x间不相关，那分析就和之前一样简单

- **例子：工资等式中遗漏能力**

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 abil + u$$

如果exper与（educ，abil）都近似不相关，则遗漏变量的偏误方向就像之前两变量的简单情形一致：β_1将被过度估计

但实际上， exper与educ通常多少会负相关

3. Measurement error problem

假设真实模型为

$$y = \alpha + \beta \boxed{x^*} + \varepsilon, \quad \mathrm{Cov}(x^*, \varepsilon) = 0, \;\; \beta \neq 0$$

但x^*无法精确观测，而只能观测到x，二者满足如下关系：

$$x = x^* + u, \quad \mathrm{Cov}(x^*, u) = 0, \;\; \mathrm{Cov}(u, \varepsilon) = 0$$

其中，测量误差u与被测量变量x^*不相关，也与扰动项ε不相关。

代入可得：

若多元回归 $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K^* + \varepsilon$

$= \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + (\varepsilon - \beta_K u)$

$$\hat{\beta}_K \to \beta_K + \frac{\mathrm{cov}(\tilde{x}_K, \varepsilon - \beta_K u)}{\mathrm{Var}(\tilde{x}_K)}$$

$\beta(x-u)+\varepsilon$

x_k 和 $\tilde{x}_k$ 是一样的.

$$y = \alpha + \beta x + (\varepsilon - \beta u)$$

新扰动项$(\varepsilon - \beta u)$与解释变量x存在相关性：

$$\begin{aligned}
\mathrm{Cov}(x, \varepsilon - \beta u) &= \mathrm{Cov}(x^* + u, \varepsilon - \beta u) \\
&= \underbrace{\mathrm{Cov}(x^*, \varepsilon)}_{=0} - \beta \underbrace{\mathrm{Cov}(x^*, u)}_{=0} + \underbrace{\mathrm{Cov}(u, \varepsilon)}_{=0} - \beta \mathrm{Cov}(u, u) \\
&= -\beta \mathrm{Var}(u) \neq 0
\end{aligned}$$

故 OLS 不一致，称为“测量误差偏差”(measurement error bias)。

可确定此偏差的方向：

$$\hat{\beta}=\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2}$$

$$\xrightarrow{p}\frac{\mathrm{Cov}(x_i,y_i)}{\mathrm{Var}(x_i)}=\frac{\mathrm{Cov}(x_i^*+u,\alpha+\beta x^*+\varepsilon)}{\mathrm{Var}(x_i^*+u)}$$

$$=\frac{\beta\,\mathrm{Var}(x_i^*)}{\mathrm{Var}(x_i^*)+\mathrm{Var}(u)}=\beta\cdot\frac{1}{1+\underbrace{\left(\sigma_u^2/\sigma_{x^*}^2\right)}_{>0}}$$

由于σ_u^2与$\sigma_{x^*}^2$一定为正，故$0<\frac{1}{1+\left(\sigma_u^2/\sigma_{x^*}^2\right)}<1$。

无论真实参数β大于或小于 0，此偏差总是使得$\hat{\beta}$的绝对值变小而趋向于 0。

故也称为“衰减偏差”(attenuation bias)或“向 0 衰减”(attenuation toward zero)。

相对于x_i^*的方差$\sigma_{x^*}^2$，如果测量误差u_i的方差σ_u^2越大，则$\left(\sigma_u^2/\sigma_{x^*}^2\right)$越大，$\dfrac{1}{1+\left(\sigma_u^2/\sigma_{x^*}^2\right)}\to 0$，则向 0 衰减的偏差越严重。

Friedman.

如果被解释变量存在测量误差，后果却不严重。真实模型：

$$y^{*} = \beta x + \varepsilon, \quad \mathrm{Cov}(x, \varepsilon) = 0, \ \beta \neq 0$$

$y - v = \beta x + \varepsilon$　　$y = \beta x + (\varepsilon + v)$

y^{*}无法精确观测，只能观测到y，二者满足如下关系：

$$y = y^{*} + v$$

$y^{*} = y - v$

Ln(次品率)

Ln(收入)

其中，v为测量误差。代入可得：

$$y = \beta x + (\varepsilon + v)$$

只要$\mathrm{Cov}(x, v) = 0$，则 OLS 一致，但可能增大扰动项的方差。

Endogenous 内生

Exogenous 外生

工具变量法直觉：

$$y = \alpha_0 + \alpha_1 x + u$$

x变 → y变　　α_1无法清晰看到.

x不变 → u变

找到一个z. ① z变 → u不变　　外生性

② z变 → x变　　相关性

通过z观察α_1

Instrumental Variable method (IV)

如能将内生变量分成两部分，一部分与扰动项相关，另一部分与扰动项不相关，可用与扰动项不相关的那部分得到一致估计。

这种分离常借助另一“工具变量”来实现。

假设在图 10.1 中，存在某个因素(变量)使得供给曲线经常移动，而需求曲线基本不动，则可估计需求曲线，参见图 10.2。

这个使得供给曲线移动的变量就是工具变量。 Z

假设供给方程的扰动项可分解为两部分，即可观测的气温 x_t 与不可观测的其他因素：

外生性
不能检验.

把 $q_t^s=\beta_0+\beta_1 p_t+v_t$ 中的

$$\underbrace{v_t=\delta_0+\delta_1 x_t+\eta_t}$$

$$q_t^s = \beta_0 + \beta_1 p_t + \overset{\delta_1}{\cancel{\beta_2}} x_t + \cancel{v_t}\ \eta_t$$

代入供给方程：$q_t^s = \underbrace{(\beta_0+\delta_0)} + \beta_1 p_t + \delta_1 x_t + \eta_t$

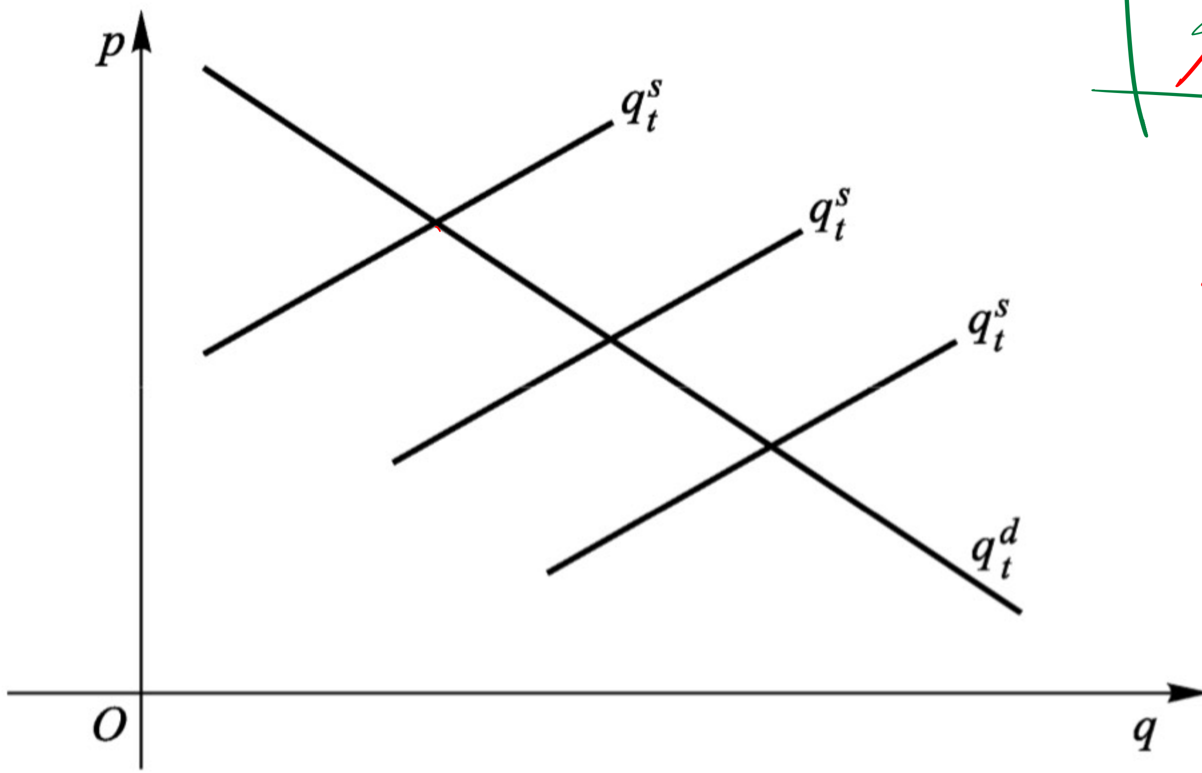


图 10.2　稳定的需求与变动的供给

假定气温 x_t 是前定变量，与两个扰动项都不相关，即 $\mathrm{Cov}(x_t, u_t) = 0$，$\mathrm{Cov}(x_t, v_t) = 0$。

由于气温 x_t 的变化使得供给函数 q_t^s 沿着需求函数 q_t^d 移动，故可估计需求函数 q_t^d。

此时，称 x_t 为“工具变量”(Instrumental Variable，简记 IV)。

在回归方程中(此处为需求方程)，一个有效(valid)的工具变量应满足以下两个条件。

(i) 相关性：工具变量与内生解释变量相关，即 $\mathrm{Cov}(x_t, p_t) \neq 0$。检验

(ii) 外生性：工具变量与扰动项不相关，即 $\mathrm{Cov}(x_t, u_t) = 0$。定性

$$\mathrm{cov}\left(x_t, \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{v - u}{\alpha_1 - \beta_1}\right) = \mathrm{cov}\left(x, \frac{\delta_0 + \delta_1 x_t + \eta - u}{\alpha_1 - \beta_1}\right)$$

$\delta_1 > 0 \qquad = \frac{\delta_1 \mathrm{Var}(x)}{-(\beta_1 - \alpha_1)} < 0$

气温↑ → 产量↑

→ 价格↓

工具变量的外生性也称“排他性约束”(exclusion restriction)，因为外生性意味着，工具变量影响被解释变量的唯一渠道是通过与其相关的内生解释变量，它排除了所有其他的可能影响渠道。

在本例中，气温 x_t 满足这两个条件。

(i) 相关性：从联立方程组可解出 $p_t = p_t(x_t, u_t, v_t)$，故 $\mathrm{Cov}(x_t, p_t) \neq 0$。

(ii) 外生性：因为气温 x_t 是前定变量，故 $\mathrm{Cov}(x_t, u_t) = 0$。

利用工具变量的这两个性质，可得到对 α_1 的一致估计。

x 工具

同时对需求方程$q_t=\alpha_0+\alpha_1 p_t+u_t$两边求与$x_t$的协方差：

$$\begin{aligned}\mathrm{Cov}(q_t,x_t)&=\mathrm{Cov}(\alpha_0+\alpha_1 p_t+u_t,x_t)\\&=\alpha_1\mathrm{Cov}(p_t,x_t)+\underbrace{\mathrm{Cov}(u_t,x_t)}_{=0}=\alpha_1\mathrm{Cov}(p_t,x_t)\end{aligned}$$

根据工具变量的相关性，$\mathrm{Cov}(p_t,x_t)\neq 0$，可把上式两边同除以$\mathrm{Cov}(p_t,x_t)$：

$\frac{\mathrm{cov}(q,p)}{\mathrm{Var}(p)}$

$$\boxed{\alpha_1=\frac{\mathrm{Cov}(q_t,x_t)}{\mathrm{Cov}(p_t,x_t)}}$$

← 相关性≠0 "方程有解"

使用对应的样本值，可得一致的“工具变量估计量”(Instrumental Variable Estimator)：

$\frac{\mathrm{cov}(q,x)}{\mathrm{cov}(p,x)}$

OLS： $X'X$ 可逆. $(X'X)^{-1}$

$$\hat{\alpha}_{1,\mathrm{IV}} = \frac{\widehat{\mathrm{Cov}(q_t, x_t)}}{\widehat{\mathrm{Cov}(p_t, x_t)}} \xrightarrow{p} \boxed{\frac{\mathrm{Cov}(q_t, x_t)}{\mathrm{Cov}(p_t, x_t)}} = \alpha_1$$

如果工具变量与内生变量无关，$\mathrm{Cov}(x_t, p_t) = 0$，则无法定义工具变量法。

如果工具变量与内生变量的相关性很弱，$\mathrm{Cov}(x_t, p_t) \approx 0$，会导致估计量$\hat{\alpha}_{1,\mathrm{IV}}$的方差变得很大，称为“弱工具变量问题”。

传统的工具变量法通过“二阶段最小二乘法”(Two Stage Least Square，简记 2SLS 或 TSLS)来实现。

$x_k \sim Z$.

Z.

第一阶段回归：用内生解释变量对工具变量回归，即 $p_t \xrightarrow{\text{OLS}} x_t$，得到拟合值 $\hat{p}_t$。

$p_t = \pi_0 + \pi_1 x_t + \gamma_t^{(1)}$

第二阶段回归：用被解释变量对第一阶段回归的拟合值进行回归，即 $q_t \xrightarrow{\text{OLS}} \hat{p}_t$。

$q_t = \alpha_0 + \alpha_1 \hat{p}_t + \gamma_t^{(2)}$

为什么这样做能得到好结果？把需求方程 $q_t = \alpha_0 + \alpha_1 p_t + u_t$ 分解

$$q_t = \alpha_0 + \alpha_1 \hat{p}_t + \underbrace{\left[u_t + \alpha_1 (p_t - \hat{p}_t)\right]}_{\equiv \varepsilon_t}$$

$\frac{1}{n}\sum x_i \hat{u}_i = 0$.

命题　在第二阶段回归中，$\hat{p}_t$ 与新扰动项 $\varepsilon_t \equiv u_t + \alpha_1 (p_t - \hat{p}_t)$ 不相关。

$$\hat{\alpha}_{1,2SLS} = \frac{\operatorname{cov}(\hat{p}_t, q_t)}{\operatorname{Var}(\hat{p}_t)} = \frac{\operatorname{cov}(\hat{\pi}_0 + \hat{\pi}_1 x_t, q_t)}{\underbrace{\operatorname{cov}(\hat{p}_t, \hat{p}_t + \hat{\gamma}_t^{(1)})}_{= \operatorname{cov}(\hat{\pi}_0 + \hat{\pi}_1 x_t,\ p_t)}}$$

$$= \frac{\cancel{\pi_1}\operatorname{cov}(x_t, q_t)}{\cancel{\pi_1}\operatorname{cov}(x_t, p_t)} = \hat{\alpha}_{IV}$$

证明： 由于$\varepsilon_t \equiv u_t + \alpha_1(p_t - \hat{p}_t)$，故

$$\operatorname{Cov}(\hat{p}_t, \varepsilon_t) = \underbrace{\operatorname{Cov}(\hat{p}_t, u_t)}_{=0} + \alpha_1 \operatorname{Cov}(\hat{p}_t, p_t - \hat{p}_t)$$

首先，由于$\hat{p}_t$是x_t的线性函数($\hat{p}_t$为第一阶段回归的拟合值)，而$\operatorname{Cov}(x_t, u_t) = 0$(工具变量的外生性)，故上式右边的第一项$\operatorname{Cov}(\hat{p}_t, u_t) = 0$。

其次，由于在第一阶段回归中，拟合值$\hat{p}_t$与残差$p_t - \hat{p}_t$正交(OLS的正交性)，故上式右边的第二项$\operatorname{Cov}(\hat{p}_t, p_t - \hat{p}_t) = 0$。

由于$\hat{p}_t$与ε_t不相关，故 2SLS 一致。

动机：简单回归模型中的遗漏变量

- **内生性的影响**
  - 考虑一个简单回归模型

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

  - OLS估计：

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \overline{x})(u_i - \overline{u})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

  - 如果$Cov(x_i, u_i) \neq 0$,

$$E(\hat{\beta}_1) \neq \beta_1,$$

$$plim\hat{\beta}_1 = \beta_1 + \frac{Cov(x_i,u_i)}{Var(x_i)} \neq \beta_1$$

  - OLS不具有无偏性与一致性

动机：简单回归模型中的遗漏变量

- **工具变量法（Instrument Variable，IV）**

y，x

- **IV（z_i）的定义：**
  - ~~1）不能出现在回归中~~ z
  - 2）与内生变量高度相关$Cov(z_i, x_i) \neq 0$
  - 3）与误差项不相关$Cov(z_i, u_i) = 0$

$y = \alpha_0 + \alpha_1 x + u$

$Cov(x, u) = 0$

x是自己的IV.

- **由性质3）可得，**

$$Cov(z_i, u_i) = 0$$

$$\Leftrightarrow \ 0 = Cov(z_i, y_i - \beta_0 - \beta_1 x_i) = Cov(z_i, y_i) - \beta_1 Cov(z_i, x_i)$$

- **又因为性质2），**

$$\Leftrightarrow \ \beta_1 = Cov(z_i, y_i)/Cov(z_i, x_i)$$

动机：简单回归模型中的遗漏变量

- 从而，我们可以定义如下IV估计量：

$$\hat{\beta}_{IV} = \frac{\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})}$$

- IV估计量满足一致性：

$$\Rightarrow \hat{\beta}_{IV} = \widehat{Cov}(z_i, y_i)/\widehat{Cov}(z_i, x_i) \rightarrow Cov(z_i, y_i)/Cov(z_i, x_i) = \beta_1$$

"无偏性" "非常不重要。"

- 注意：一般情况下IV估计量不具有无偏性

- 如果$E(u_i^2|z_1, \ldots, z_n) = \sigma^2$，可以计算$\hat{\beta}_{IV}$的渐近方差为

$$\frac{\sigma^2}{n\sigma_x^2\rho_{x,z}^2} \leftarrow \sigma^2(X^{*\prime}X^*)^{-1}$$

Jensen.

其中σ_x^2为x_i的总体方差，$\rho_{x,z}^2$为x_i与z_i总体相关系数的平方 同方差时，IV 差的 OLS

"有效性"　方差大.

"一致性"很重要.

10.2　工具变量法作为一种矩估计

1．矩估计 (Method of Moments，MM)

假设随机变量$x \sim N(\mu, \sigma^2)$，其中μ, σ^2为待估参数。

因为有两个待估参数，故使用两个总体矩条件(population moment conditions)：

一阶矩：$\mathrm{E}(x) = \mu$

二阶矩：$\mathrm{E}(x^2) = \mathrm{Var}(x) + \left[\mathrm{E}(x)\right]^2 = \sigma^2 + \mu^2$

用对应的样本矩(sample moments)来替代总体矩条件可得：

E(x).

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n} x_i = \hat{\mu} \\ \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \hat{\mu}^2 + \hat{\sigma}^2 \end{cases} \Rightarrow \begin{cases} \hat{\mu} = \overline{x} \\ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \end{cases}$$

其中，$\overline{x} \equiv \frac{1}{n}\sum_{i=1}^{n} x_i$ 为样本均值。推导中用到，$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$。

任何随机向量 $\boldsymbol{x}$ 的函数 $f(\boldsymbol{x})$ 的期望 $\mathrm{E}[f(\boldsymbol{x})]$ 都称为“总体矩”。

OLS 也是矩估计。利用解释变量与扰动项的正交性：

假设

MLR 4.

$$\mathrm{E}\left[\boldsymbol{x}_i \varepsilon_i\right]=\mathbf{0} \quad \Rightarrow \quad \mathrm{E}\left[\boldsymbol{x}_i (y_i - \boldsymbol{x}_i' \boldsymbol{\beta})\right]=\mathbf{0} \qquad (\text{代入}\ \varepsilon_i = y_i - \boldsymbol{x}_i'\boldsymbol{\beta})$$

$$\Rightarrow \quad \mathrm{E}(\boldsymbol{x}_i y_i) = \mathrm{E}(\boldsymbol{x}_i \boldsymbol{x}_i')\boldsymbol{\beta} \qquad (\text{展开、移项})$$

Normal Equation

$$\frac{1}{n}\sum x_i \varepsilon_i = 0$$

不同于.

$$\Rightarrow \quad \boldsymbol{\beta} = \left[\mathrm{E}(\boldsymbol{x}_i \boldsymbol{x}_i')\right]^{-1} \mathrm{E}(\boldsymbol{x}_i y_i) \qquad (\text{假设}\left[\mathrm{E}(\boldsymbol{x}_i \boldsymbol{x}_i')\right]^{-1}\text{存在，求解}\ \boldsymbol{\beta})$$

以样本矩替代总体矩，可得矩估计：

$$\hat{\boldsymbol{\beta}}_{\mathrm{MM}} = \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i y_i\right) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$$

2．工具变量法作为一种矩估计

假设回归模型为

外生　　内生,

$$y_i = \beta_1 x_{i1} + \cdots + \beta_{K-1} x_{i,K-1} + \beta_K x_{iK} + \varepsilon_i$$

只有最后一个解释变量x_{iK}为内生变量，即$\mathrm{Cov}(x_{iK}, \varepsilon_i) \neq 0$，故 OLS 不一致。

假设有一个有效工具变量w满足$\mathrm{Cov}(x_{iK}, w_i) \neq 0$(相关性)，以及$\mathrm{Cov}(w_i, \varepsilon_i) = 0$(外生性)。

由于$x_1, \cdots, x_{K-1}$不是内生变量，可把自身作为自己的工具变量(满足工具变量法的两个条件)。

记解释向量$\boldsymbol{x}_i \equiv (x_{i1} \cdots x_{i,K-1}\ x_{iK})'$，则原模型为$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$。

记工具向量$\boldsymbol{z}_i \equiv (z_{i1} \cdots z_{i,K-1}\ z_{iK})' \equiv (x_{i1} \cdots x_{i,K-1}\ w_i)'$。

定义$\boldsymbol{g}_i \equiv \boldsymbol{z}_i\varepsilon_i$。由于工具向量与扰动项正交，故$\mathrm{E}(\boldsymbol{g}_i) = \mathrm{E}(\boldsymbol{z}_i\varepsilon_i) = \boldsymbol{0}$为“总体矩条件”或“正交条件”(orthogonality condition)。

$$\boxed{\mathrm{E}(\boldsymbol{z}_i\varepsilon_i) = \boldsymbol{0}} \Rightarrow \mathrm{E}\left[\boldsymbol{z}_i(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})\right] = \boldsymbol{0} \qquad (\text{代入}\varepsilon_i = y_i - \boldsymbol{x}_i'\boldsymbol{\beta})$$

(1) 外生性条件

$$\Rightarrow \mathrm{E}(\boldsymbol{z}_i y_i) = \left[\mathrm{E}(\boldsymbol{z}_i\boldsymbol{x}_i')\right]\boldsymbol{\beta} \qquad (\text{展开、移项})$$

$x\beta+\varepsilon$

$$\Rightarrow \boldsymbol{\beta} = \left[\mathrm{E}(\boldsymbol{z}_i\boldsymbol{x}_i')\right]^{-1}\mathrm{E}(\boldsymbol{z}_i y_i) \qquad (\text{假设}\left[\mathrm{E}(\boldsymbol{z}_i\boldsymbol{x}_i')\right]^{-1}\text{存在})$$

存在. (2) 相关性条件. → β.

$$Z:\ \begin{bmatrix} x_{11} & \cdots & x_{1,k-1} & w_1 \\ \vdots & \ddots & \vdots & \vdots \\ x_{n1} & \cdots & x_{n,k-1} & w_n \end{bmatrix}_{n\times k}$$

以样本矩代替上式中的总体矩，可得 IV 估计量，

$$\hat{\boldsymbol{\beta}}_{\text{IV}} = \left(\frac{1}{n}\sum\nolimits_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{x}_i'\right)^{-1}\left(\frac{1}{n}\sum\nolimits_{i=1}^{n} \boldsymbol{z}_i y_i\right) = \boxed{(\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{y}}$$

其中，$\boldsymbol{Z} \equiv (\boldsymbol{z}_1 \cdots \boldsymbol{z}_{n-1}\ \boldsymbol{z}_n)'$。

$$(\underset{Z}{\cancel{X}}'X)^{-1}\underset{Z}{\cancel{X}}'y$$

OLS 也是一种工具变量法。

exogenous ('predetermined')

如果 $\boldsymbol{x}_i$ 全部是前定变量，可将自己作为工具变量，即 $\boldsymbol{z}_i = \boldsymbol{x}_i$，$\boldsymbol{Z} = \boldsymbol{X}$。因此，

$$\hat{\boldsymbol{\beta}}_{\text{IV}} = (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{y} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \hat{\boldsymbol{\beta}}_{\text{OLS}}$$

$n > K$.

命题 如果秩条件“$\operatorname{rank}\left[\mathrm{E}(z_i x_i')\right]=K$”成立(即方阵$\mathrm{E}(z_i x_i')$满秩)，则在一定的正则条件下，$\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$是$\boldsymbol{\beta}$的一致估计，且$\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$服从渐近正态分布。

OLS: $g_i = x_i \varepsilon_i$

IV: $g_i = z_i \varepsilon_i$

证明： 因为$\operatorname{rank}\left[\mathrm{E}(z_i x_i')\right]=K$，故$\left[\mathrm{E}(z_i x_i')\right]^{-1}$存在。

$$\hat{\boldsymbol{\beta}}_{\mathrm{IV}}-\boldsymbol{\beta}=(\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'(\underbrace{\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}}_{y})-\boldsymbol{\beta}=(\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{\varepsilon}$$

$$=\left(\frac{1}{n}\sum_{i=1}^{n} z_i x'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} z_i\varepsilon_i\right)=\boldsymbol{S}_{ZX}^{-1}\bar{\boldsymbol{g}}\xrightarrow{p}\left[\mathrm{E}(z_i x_i')\right]^{-1}\underbrace{\mathrm{E}(\boldsymbol{g}_i)}_{=\mathbf{0}}=\mathbf{0}$$

其中，$\boldsymbol{S}_{ZX}\equiv\frac{1}{n}\sum_{i=1}^{n} z_i x_i'$，$\bar{\boldsymbol{g}}\equiv\frac{1}{n}\sum_{i=1}^{n} z_i\varepsilon_i$。

根据与第 5 章类似的推导可知：

$$\sqrt{n}\,\bar{\boldsymbol{g}} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{S})$$

其中，$\boldsymbol{S} \equiv \mathrm{E}(\boldsymbol{g}_i \boldsymbol{g}_i') = \mathrm{E}(\varepsilon_i^2 \boldsymbol{z}_i \boldsymbol{z}_i')$。工具变量估计量 $\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$ 渐近地服从正态分布：

条件异方差.　　$\Omega = \begin{bmatrix} \sigma & & \\ & \ddots & \\ & & \sigma \end{bmatrix}$

$$\sqrt{n}\,(\hat{\boldsymbol{\beta}}_{\mathrm{IV}} - \boldsymbol{\beta}) \xrightarrow{d} N\left(\mathbf{0}, \mathrm{Avar}(\hat{\boldsymbol{\beta}}_{\mathrm{IV}})\right)$$

其中，渐近方差矩阵 $\mathrm{Avar}(\hat{\boldsymbol{\beta}}_{\mathrm{IV}}) = \left[\mathrm{E}(\boldsymbol{z}_i \boldsymbol{x}_i')\right]^{-1} \boldsymbol{S} \left[\mathrm{E}(\boldsymbol{x}_i \boldsymbol{z}_i')\right]^{-1}$。

秩条件 $\mathrm{rank}\left[\mathrm{E}(\boldsymbol{z}_i \boldsymbol{x}_i')\right] = K$ 意味着，工具变量 w_i 与解释变量 x_i 相关。

如果不相关，则秩条件无法满足。

以一元回归为例，此时，$K=2$，$\boldsymbol{x}_i=(1\ \ x_i)'$，$\boldsymbol{z}_i=(1\ \ w_i)'$，则

$$\mathrm{E}(\boldsymbol{z}_i\boldsymbol{x}_i')=\mathrm{E}\left[\begin{pmatrix}1\\ w_i\end{pmatrix}(1\quad x_i)\right]=\mathrm{E}\begin{bmatrix}1 & x_i\\ w_i & w_ix_i\end{bmatrix}=\begin{bmatrix}1 & \mathrm{E}(x_i)\\ \mathrm{E}(w_i) & \mathrm{E}(w_ix_i)\end{bmatrix}$$

$$\mathrm{rank}\left[\mathrm{E}(\boldsymbol{z}_i\boldsymbol{x}_i')\right]=K=2\qquad\Leftrightarrow\ \text{行列式}\begin{vmatrix}1 & \mathrm{E}(x_i)\\ \mathrm{E}(w_i) & \mathrm{E}(w_ix_i)\end{vmatrix}\neq 0$$

有n个内生x，

$$\Leftrightarrow\ \mathrm{E}(w_ix_i)-\mathrm{E}(w_i)\mathrm{E}(x_i)\neq 0$$

就有n个IV.

$\Leftrightarrow$ $\underline{\mathrm{Cov}(w_i,x_i)\neq 0}$，即 w_i 与 x_i 相关。

阶条件：满足秩条件的必要条件是z_i中至少包含 K 个变量，即不在方程中出现的工具变量个数不少于方程中内生解释变量的个数。称此条件为“阶条件”(order condition)。

根据是否满足阶条件可分为三种情况：

(1) 不可识别(unidentified)：工具变量个数小于内生解释变量个数；

(2) 恰好识别(just or exactly identified)：工具变量个数等于内生解释变量个数；

(3) 过度识别(overidentified)：工具变量个数大于内生解释变量个数。

① 外生性，定性，

② 相关性，要检验。　　秩条件。　　弱工具变量问题，

(所有工具变量的维度。$1 \cdot x_1 + 0$，

命题：$y = \beta_1 x_1 + \cdots + \beta_{K-1} x_{K-1} + \beta_K x_K + \varepsilon$　　$\swarrow$ 内生

$x_1 \sim Z$

$\{z_1, \cdots, z_M\}$，$Z = \begin{bmatrix} x_1 \\ \vdots \\ x_{K-1} \\ z_1 \\ \vdots \\ z_M \end{bmatrix}$.

$\mathrm{rank}(E[Zx']) = K \iff x_K = \delta_1 x_1 + \cdots + \delta_{K-1} x_{K-1} + \theta_1 z_1 + \cdots + \theta_M z_M + r_K$

$\theta_1, \cdots, \theta_M$ 至少有一个不为 0.

证：$x_K = x_K^* + r_K$，$x^* = \begin{bmatrix} x_1 \\ \vdots \\ x_{K-1} \\ x_K^* \end{bmatrix}$，$x = x^* + r$

$E[Zx']) = E[Z(x^{*\prime} + r')] = E[Zx^{*\prime}]$

$$X^* = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & & & & 0 \\ & & \vdots & & & & \\ 0 & \cdots & \underset{(k-1)}{1} & 0 & \cdots & & 0 \\ \delta_1 & \cdots & \delta_{k-1} & \theta_1 & \cdots & & \theta_M \end{bmatrix}}_{\substack{K\times L \\ \Pi}} \underset{L\times 1}{Z} = \Pi Z.$$

$$(X^*)' = Z'\Pi'$$

$$E[ZX'] = E[ZZ'\Pi'] = E[ZZ']\Pi'$$

$$\Pi' = \begin{array}{c} \begin{bmatrix} 1 & & 0 & \delta_1 \\ 0 & & \vdots & \vdots \\ \vdots & \cdots & 1\,(k-1) & \delta_{k-1} \\ & & \vdots & \theta_1 \\ \boxed{0 \quad \cdots \quad 0} & & & \vdots \\ & & & \theta_M \end{bmatrix} \end{array} \begin{array}{l} \left.\vphantom{\begin{matrix}1\\0\\1\end{matrix}}\right\} K-1\text{列} \\ \\ \\ \end{array}$$

$\underbrace{\boxed{0 \cdots 0}}_{K-1\text{列}}$ 全为0

若所有$\theta_1 = \theta_2 = \cdots = \theta_M = 0$

以上介绍的工具变量法仅适用于“恰好识别”的情形。

在“过度识别”的情况下，$\boldsymbol{Z'X}$ 不是方阵，$(\boldsymbol{Z'X})^{-1}$ 不存在，也就无法定义工具变量估计量 $\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$。

解决方法之一是扔掉“多余”的工具变量，但被扔掉的工具变量包含有用的信息。

Efficient.

有效率的做法是 2SLS。

2SLS 在所有工具变量回归中方差最小。

10.3　二阶段最小二乘法

多个工具变量的线性组合仍然是工具变量，因为仍满足工具变量的两个条件(相关性与外生性)。

如果生成工具变量的 K 个线性组合，则又回到恰好识别的情形。

在球型扰动项的假定下，由 2SLS 所提供的工具变量线性组合是所有线性组合中最渐近有效的。

第一阶段(分离出内生变量的外生部分)

外生的东西参与第一阶段、回归、

将解释变量$\boldsymbol{x}_1, \cdots, \boldsymbol{x}_K$分别对所有 L 个工具变量$\{\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_L\}$作 OLS 回归，其中第 k 个解释变量$\boldsymbol{x}_k \equiv (x_{1k} \cdots x_{nk})'_{n\times 1}$。

得到拟合值：

$$\hat{\boldsymbol{x}}_1 = \boldsymbol{P}\boldsymbol{x}_1,\quad \hat{\boldsymbol{x}}_2 = \boldsymbol{P}\boldsymbol{x}_2,\quad \cdots,\quad \hat{\boldsymbol{x}}_K = \boldsymbol{P}\boldsymbol{x}_K$$

其中，$\boldsymbol{P} \equiv \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$为对应于$\boldsymbol{Z}$的投影矩阵。定义

$$\hat{\boldsymbol{X}} \equiv (\hat{\boldsymbol{x}}_1\ \hat{\boldsymbol{x}}_2 \cdots \hat{\boldsymbol{x}}_K) = \boldsymbol{P}(\boldsymbol{x}_1\ \boldsymbol{x}_2 \cdots \boldsymbol{x}_K) = \boldsymbol{P}\boldsymbol{X} = \boldsymbol{Z}\left[(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}\right]$$

$y \sim \hat{X}$　　P 投影　　$P = P^2$

idempotent　　$P' = P$

第二阶段(使用此外生部分进行回归)

由于 $\hat{\boldsymbol{X}}$ 是 $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_L\}$ 的线性组合(参见第一阶段回归)，故 $\hat{\boldsymbol{X}}$ 恰好包含 K 个工具变量。

$(Z'X)^{-1} Z'y$

使用 $\hat{\boldsymbol{X}}$ 为工具变量对原模型 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ 进行工具变量法估计：

$$\hat{\boldsymbol{\beta}}_{\mathrm{IV}} = (\hat{\boldsymbol{X}}'\boldsymbol{X})^{-1}\hat{\boldsymbol{X}}'\boldsymbol{y} = (\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}})^{-1}\hat{\boldsymbol{X}}'\boldsymbol{y}$$

$(PX)'X$　　$X'P'X$　　$X'P'PX$

因为 $\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}} = (\boldsymbol{P}\boldsymbol{X})'(\boldsymbol{P}\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{X} = \boldsymbol{X}'\boldsymbol{P}'\boldsymbol{X} = \hat{\boldsymbol{X}}'\boldsymbol{X}$，投影矩阵 $\boldsymbol{P}$ 为对称幂等矩阵，$\boldsymbol{P}' = \boldsymbol{P}$，$\boldsymbol{P}^2 = \boldsymbol{P}$。

可将 $\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$ 视为把 $\boldsymbol{y}$ 对 $\hat{\boldsymbol{X}}$ 进行 OLS 回归而得到，故名“二阶段最小二乘法”。

第二阶段回归所得到的残差为$\boldsymbol{e}_2 \equiv \boldsymbol{y} - \hat{\boldsymbol{X}}\hat{\boldsymbol{\beta}}_{2SLS}$，而原方程残差却是$\boldsymbol{e} \equiv \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{2SLS}$。

执行 2SLS 最好直接使用 Stata 的命令。

在同方差的假设下，$\hat{\boldsymbol{\beta}}_{IV}$ 的协方差矩阵估计量为 $\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}}_{IV})} = s^2(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}})^{-1}$，其中 $s^2 \equiv \dfrac{\boldsymbol{e}'\boldsymbol{e}}{n-K}$。 $\sigma^2(X'X)^{-1}$

在异方差的情况下，应使用稳健的协方差矩阵估计量，即 $\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}}_{IV})} = (\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}})^{-1}\left(\sum_{i=1}^{n} e_i^2 \hat{\boldsymbol{x}}_i \hat{\boldsymbol{x}}_i'\right)(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}})^{-1}$。

Projection

将 $\hat{\boldsymbol{X}} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}$ 代入方程，可得 2SLS 的最终表达式：

$$\hat{\boldsymbol{\beta}}_{2SLS} = (\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P}\boldsymbol{y} = \left[\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}\right]^{-1}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y}$$

IV 是有偏的。

以恰好识别为例。

$E[\hat{\beta}_{IV}] = \beta + E[(ZX')^{-1} Z u]$ $\qquad E[u|Z]=0$ 外生性假设。

$\qquad\qquad\qquad\quad x \qquad x$ $\qquad\qquad E[u|x]=0.$

$E[(ZX')^{-1} Z E[u|Z]]$ ×

$E[u|Z, x]=0$ ×

可以 $E[(ZX')^{-1} Z E[u|Z, x]]$

x 外生。

Consistency?

$\hat{\beta}_{IV} = \beta + \left(\frac{Z'X}{n}\right)^{-1}\left(\frac{Z'u}{n}\right)$

Slutsky Theorem

$E(ZX')^{-1} \underbrace{E(Zu)}_{=0}$

OLS 和 2SLS 的方差（同方差时，u）

$$Avar(\hat{\beta}_{OLS}-\beta) = \sigma^2 E[xx']^{-1}$$

$$Avar(\hat{\beta}_{2SLS}-\beta) = \sigma^2 E[x^* x^{*'}]^{-1},$$

$x = x^* + r$

$E[x^* r] = 0.$

$$E[xx'] = E[(x^*+r)(x^*+r')]$$

$$= E[x^* x^{*'}] + E[rr']$$

$$\sigma^2 E[x^* x^{*'}]^{-1} - \sigma^2 E[xx']^{-1}$$

半正定.

p.s.d.

因为 $E[xx'] - E[x^* x^{*'}] = E[rr'].$

异方差：

$$Var(\hat{\beta}_{OLS}) = (X'X)^{-1}(X'\Omega X)(X'X)^{-1}$$

$$Var(\hat{\beta}_{2SLS}) = (X'PX)^{-1}X'P\Omega PX(X'PX)^{-1}$$

同方差，命题：2SLS 的渐近方差在所有工具变量估计量中最小。

证：$\hat{\beta}$ 是 2SLS，$\tilde{\beta}$ 是 IV 估计量，IV 的线性组合的估计

IV 维度 L 维，K 个 X，$\tilde{\beta}$ 须从 L 维 → K 维，

2SLS：$\underset{K\times 1}{X^*} = \underset{K\times L}{\Pi'}\underset{L\times 1}{Z}$，$\Pi = E[ZZ']^{-1}E[ZX']$

"LIML"

$$\tilde{x} = \underset{K\times L}{\Gamma'} z,$$

$$\tilde{\beta} \to E[\tilde{x}x']^{-1} E[\tilde{x}'y]$$

$$= \beta + E[\tilde{x}x']^{-1} E[\tilde{x}\varepsilon]$$

2SLS: $\hat{\beta} \longrightarrow E[x^* x']^{-1} E[x^* y]$

$$= \beta + E[x^* x']^{-1} E[x^* \varepsilon]$$

同方差.

$$E[x^* x'] = E[x^*(x^* + r)']$$

$$= E[x^* x^{*\prime}]$$

$$\mathrm{Avar}(\hat{\beta} - \beta) = \sigma^2 E[x^* x^{*\prime}]^{-1}$$

$$\updownarrow$$

$$\mathrm{Avar}(\tilde{\beta} - \beta) = \sigma^2 E[\tilde{x}x']^{-1} E[\tilde{x}\tilde{x}'] E[x\tilde{x}']^{-1}$$

$$E[x^* x^{*\prime}] - E[\tilde{x} x'] E[\tilde{x}\tilde{x}']^{-1} E[x \tilde{x}'] \quad \text{p.s.d.}$$

$$E[x\tilde{x}'] = E[(x^*+r)\tilde{x}'] = E[x^*\tilde{x}'] + \cancel{E[r\tilde{x}']}$$

$$= E[x^*\tilde{x}']$$

$$E[x^* x^{*\prime}] - E[x^*\tilde{x}'] E[\tilde{x}\tilde{x}']^{-1} E[\tilde{x} x^{*\prime}] \quad ①$$

x^* 对 $\tilde{x}$ 回归, 拟合值

$$E[(\tilde{x}' E[\tilde{x}\tilde{x}']^{-1} E[\tilde{x} x^{*\prime}])' (\tilde{x}' E[\tilde{x}\tilde{x}']^{-1} E[\tilde{x} x^{*\prime}])]$$

$$= E[x^*\tilde{x}'] E[\tilde{x}\tilde{x}']^{-1} \cancel{E[\tilde{x}\tilde{x}']} \cancel{E[\tilde{x}\tilde{x}']^{-1}} E[\tilde{x} x^{*\prime}]$$

Linear Projection

$$① = E[x^* x^{*\prime}] - E[L(x^*|\tilde{x}) L(x^*|\tilde{x})']$$

$$= E[s^* s^{*\prime}] \quad \text{p.s.d.}$$

OLS，IV，2SLS.

OLS	IV	2SLS
有偏	有偏	有偏
不一致	一致	一致
方差小.（同方差时）	比OLS大	方差比OLS大 比IV小.

① 外生性，定性

定量检验必须在过度识别时才可进行.

② 相关性.

$\mathrm{rank}(E[zx'])=K$.

OLS：$E[xx']$ 可逆（多重共线性）

Z.

$$\hat{\beta}_{IV}=(Z'X)^{-1}Z'y$$

动机：简单回归模型中的遗漏变量

- **例子：父亲的教育程度作为IV**

OLS:
$$\widehat{\log}(wage) = -\underset{(.185)}{.185} + \underset{(.014)}{.109}\, educ$$

教育回报可能被高估了

$$n = 428, R^2 = .118$$

$$\widehat{educ} = \underset{(.28)}{10.24} + \underset{(.029)}{.269}\, \boxed{fatheduc}$$

父亲的教育程度是一个好IV吗？

1） 没有作为解释变量
2） 与教育程度显著相关
3） 与误差项不相关（？）

$$n = 428, R^2 = .173$$

教育回报率低了（符合预期）

IV:
$$\widehat{\log}(wage) = \underset{(.446)}{.441} + \underset{(.035)}{.059}\, educ$$

标准误也大了，不精确了

$$n = 428, R^2 = 1 - RSS_{IV}/TSS = .093$$

动机：简单回归模型中的遗漏变量

- **文献中教育程度的其他IV：**
- **兄弟姐妹的个数**
  - 1）不是工资的决定因素，2）与教育程度相关，因为预算约束，3）与先天因素无关
- **当16岁时与大学的地理远近**
  - 1）不是工资的决定因素，2）与教育相关，因为离大学近将会受到更多教育，3）与误差项不相关
- **出生月份** Angrist & Krueger (1991)
  - 1）不是工资的决定因素，2）与教育相关，因为义务教育法，3）与误差项不相关

动机：简单回归模型中的遗漏变量

- **低劣工具变量条件下IV的性质**

$\beta + (Z'X)^{-1}(Z'u)$

  - 如果工具变量不是完全外生的，而且与 x 的相关性较弱，则IV较OLS来说可能非常不一致

$\frac{cov(x,u)}{var(x)}$ $\frac{Corr(x,u)\,\sigma_x\,\sigma_u}{\sigma_x^2}$

$$plim\ \hat{\beta}_{1,OLS} = \beta_1 + Corr(x,u)\cdot\frac{\sigma_u}{\sigma_x}$$

如果IV是外生的，则没问题。如果不是，则IV与x的相关性越弱，渐近偏差就越大。

$$plim\ \hat{\beta}_{1,IV} = \beta_1 + \frac{\boxed{Corr(z,u)}}{Corr(z,x)}\cdot\frac{\sigma_u}{\sigma_x}$$

$\frac{cov(z,u)}{cov(z,x)} = \frac{Corr(z,u)\,\sigma_z\,\sigma_u}{Corr(z,x)\,\sigma_z\,\sigma_x}$

<u>IV比OLS更差，如果：</u> $\frac{Corr(z,u)}{Corr(z,x)} > Corr(x,u)$ 例如 $\frac{0.03}{0.2} > 0.1$

- IV与内生变量相关性过于弱导致的IV估计不准确问题被称为弱工具变量问题

多元回归模型的IV估计

- **考虑如下模型：**

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$$

  - y_2为内生变量（endogeneous variable）：$Cov(y_2, u_1) \neq 0$
  - z_1为外生变量（exogeneous variable）：$Cov(z_1, u_1) = 0$

- **工具变量z_2的条件:**
  - 1）不出现在回归中
  - 2）与误差项不相关:$Cov(z_2, u_1) = 0$
  - 3）与内生解释变量相关:$Cov(y_2, z_2) \neq 0$

- u_1 满足:

$$E(u_1) = 0, \qquad E(z_1 u_1) = 0, \qquad E(z_2 u_1) = 0$$

多元回归模型的IV估计

- 因为$u_1 = y_1 - \beta_0 - \beta_1 y_2 - \beta_2 z_1$，

$$E\left(y_1 - \beta_0 - \beta_1 y_2 - \beta_2 z_1\right) = 0, \qquad E\left(z_1(y_1 - \beta_0 - \beta_1 y_2 - \beta_2 z_1)\right) = 0,$$

$$E\left(z_2\left(y_1 - \beta_0 - \beta_1 y_2 - \beta_2 z_1\right)\right) = 0.$$

- 样本版本：

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1}\right) = 0, \qquad \frac{1}{n}\sum_{i=1}^{n}\left(z_{i1}\left(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1}\right)\right) = 0,$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(z_{i2}\left(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1}\right)\right) = 0.$$

- 定义$X_i = (1, y_{i2}, z_{i1})'$, $\boldsymbol{Z}_i = (1, z_{i1}, z_{i2})'$, $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)'$,

$$\hat{\beta} = \left(\sum_{i=1}^{n} Z_i X_i{}'\right)^{-1} \sum_{i=1}^{n} Z_i y_{i1}$$

多元回归模型的IV估计

- **更一般的情况：**

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 \boxed{z_1} + \ldots + \beta_k \boxed{z_{k-1}} + u_1$$

内生的　　　外生变量

- **工具变量 z_k 的条件**
  - 1）不出现在回归中
  - 2）与误差项不相关
  - 3）与内生解释变量相关

- u_1 满足：

$$E(u_1) = 0, \qquad E(z_j u_1) = 0, \qquad j = 1, \ldots, k$$

多元回归模型的IV估计

- **样本版本：**

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_{i1}-\hat{\beta}_0-\hat{\beta}_1 y_{i2}-\hat{\beta}_2 z_{i1}-\cdots-\hat{\beta}_k z_{ik-1}\right)=0,$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(z_{ij}\left(y_{i1}-\hat{\beta}_0-\hat{\beta}_1 y_{i2}-\hat{\beta}_2 z_{i1}-\cdots-\hat{\beta}_k z_{ik-1}\right)\right)=0,$$

- 用 k+1 个方程来解 k+1 个估计量 $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$
- **定义**$X_i=(1, y_{i2}, z_{i1}, \ldots, z_{ik-1})'$, $\boldsymbol{Z}_i=(1, z_{i1}, z_{i2}, \ldots, z_{ik})'$, $\hat{\beta}=(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_{\boldsymbol{k}})'$,

$$\hat{\beta}=\left(\sum_{i=1}^{n} Z_i X_i'\right)^{-1}\sum_{i=1}^{n} Z_i y_{i1}$$

多元回归模型的IV估计

- 可以拓展到有多个内生变量与多个工具变量的情况
- 该方法要求工具变量的个数等于内生解释变量的个数
- 如果工具变量的个数小于内生解释变量的个数，则没有方法估计
- 如果工具变量的个数大于(或等于)内生解释变量的个数，则可以采用两阶段最小二乘法（2SLS）

两阶段最小二乘

- **两阶段最小二乘（2SLS）估计**

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \ldots + \beta_k z_{k-1} + u_1$$

$$y_2 = \pi_0 + \pi_1 z_1 + \ldots + \pi_k z_{k-1} + \pi_k z_k + v_2$$

第一阶段（= 简约型回归）:

用外生信息来解释内生变量y_2

$$\widehat{y}_2 = \widehat{\pi}_0 + \widehat{\pi}_1 z_1 + \ldots + \widehat{\pi}_k z_{k-1} + \widehat{\pi}_k z_k$$

额外的外生变量（即工具变量）

第二阶段（= 用第一阶段 y_2 的预测值来做一个OLS回归）:

$$y_1 = \beta_0 + \beta_1 \widehat{y}_2 + \beta_2 z_1 + \ldots + \beta_k z_{k-1} + error$$

两阶段最小二乘

- **为什么2SLS有效？**
  - 因为y_2由其预测值所替代（是由外生变量预测的），所以在第二阶段的所有变量都是外生的
  - 因为使用由外生信息产生的预测值，所以y_2清除了其内生部分（即与误差项相关的部分）
- **2SLS的特征**
  - 如果仅有一个内生变量且仅有一个IV，那么2SLS=IV
  - 如果仅有一个内生变量且工具变量有多个，则2SLS估计也可以使用
  - 如果内生变量有多个，第一阶段中对每个内生变量都进行OLS回归
  - 第二阶段的OLS回归的标准误是不准确的
  - 2SLS/IV一般来说精确性不如OLS，因为在第二阶段回归中的多重共线性以及更少的解释变量的变动

两阶段最小二乘

- **例子：在工资等式中使用两个IV的2SLS**

第一步回归（educ对所有外生变量做回归）：

$$\widehat{educ} = \underset{(.27)}{8.37} + \underset{(.026)}{.085}\ exper - \underset{(.001)}{.002}\ exper^2$$

$$+ \underset{(.024)}{.185}\ \boxed{fatheduc} + \underset{(.026)}{.186}\ \boxed{motheduc}$$

教育程度显著与父母的教育程度相关

2SLS估计结果：

$$\widehat{\log}(wage) = \underset{(.400)}{.048} + \underset{(.031)}{\boxed{.061}}\ educ + \underset{(.013)}{.044}\ exper - \underset{(.0004)}{.0009}\ exper^2$$

$$n = 428, R^2 = 0.136$$

较OLS来说，教育回报低了很多，标准误也大了（不精确了）

两阶段最小二乘

- **检验弱工具变量**

第一阶段回归.

$$y_2 = \pi_0 + \pi_1 z_1 + \ldots + \pi_k z_{k-1} + \pi_k z_k + v_2$$

- **通过t检验可以验证z_k与y_2的相关性：**

$$H_0\colon \pi_k = 0$$

- **斯托克和约吉（Stock and Yogo， 2005）提出如果t 统计量大于3.2（多个工具变量F 统计量大于10）时，工具变量可以继续使用。**

rule of thumb

经验法则.

内生性检验

- **检验解释变量的内生性**

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \ldots + \beta_k z_{k-1} + u_1$$

怀疑这个变量是内生的

$$y_2 = \pi_0 + \pi_1 z_1 + \ldots + \pi_k z_{k-1} + \pi_k z_k + v_2$$

变量y_2是外生的，当且仅当v_2与u_1不相关，即在下面回归中 δ_1 为0:

$$u_1 = \delta_1 v_2 + e_1$$

- **总结：**变量y_2是外生的，即在下面回归中 δ_1 为0:

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \cdots + \beta_k z_{k-1} + \delta_1 v_2 + e_1$$

内生性检验

- **回归模型：**

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \ldots + \beta_k z_{k-1} + \delta_1 \hat{v}_2 + e_1$$

- **检验原假设**

$$H_0\colon \delta_1 = 0$$

- **如果** δ_1 **显著异于0， y_2是外生的原假设被拒绝**
- **注意：检验内生性也需要工具变量**

$$\hat{\boldsymbol{\beta}}_{2\mathrm{SLS}} = (\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P}\boldsymbol{y} = \left[\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}\right]^{-1}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y}$$

$cov(z,x) \to 0$?

10.4 有关工具变量的检验

$cov(z,x)=0$?

1. 不可识别检验

使用工具变量法的前提之一是秩条件成立，即 $\mathrm{rank}\left[\mathrm{E}(\boldsymbol{z}_i\boldsymbol{x}_i')\right] = K$ (满列秩)，其中 $\boldsymbol{z}_i \equiv (z_{i1} \cdots z_{iL})'$ (L 个工具变量)，$\boldsymbol{x}_i \equiv (x_{i1} \cdots x_{iK})'$ (K 个解释变量)，$\boldsymbol{z}_i$ 与 $\boldsymbol{x}_i$ 可有重叠元素，且 $L \geq K$ (阶条件)。

如果矩阵$\mathrm{E}(\boldsymbol{z}_i\boldsymbol{x}_i')$的列秩小于$K$，则不可识别。

对于秩条件是否成立，可进行“不可识别检验”(underidentification test)。

原假设为“H_0: $\mathrm{rank}\left[\mathrm{E}(\boldsymbol{z}_i\boldsymbol{x}_i')\right]=K-1$”（不可识别），而替代假设为“$H_1$: $\mathrm{rank}\left[\mathrm{E}(\boldsymbol{z}_i\boldsymbol{x}_i')\right]=K$”。

在扰动项为 iid 的假设下(不存在异方差)，可使用“Anderson LM 统计量” (Anderson, 1951)，其渐近分布为$\chi^2(L-K+1)$。

如果不作 iid 扰动项的假设(允许存在异方差)，应使用

频率学派 (Bayesian)

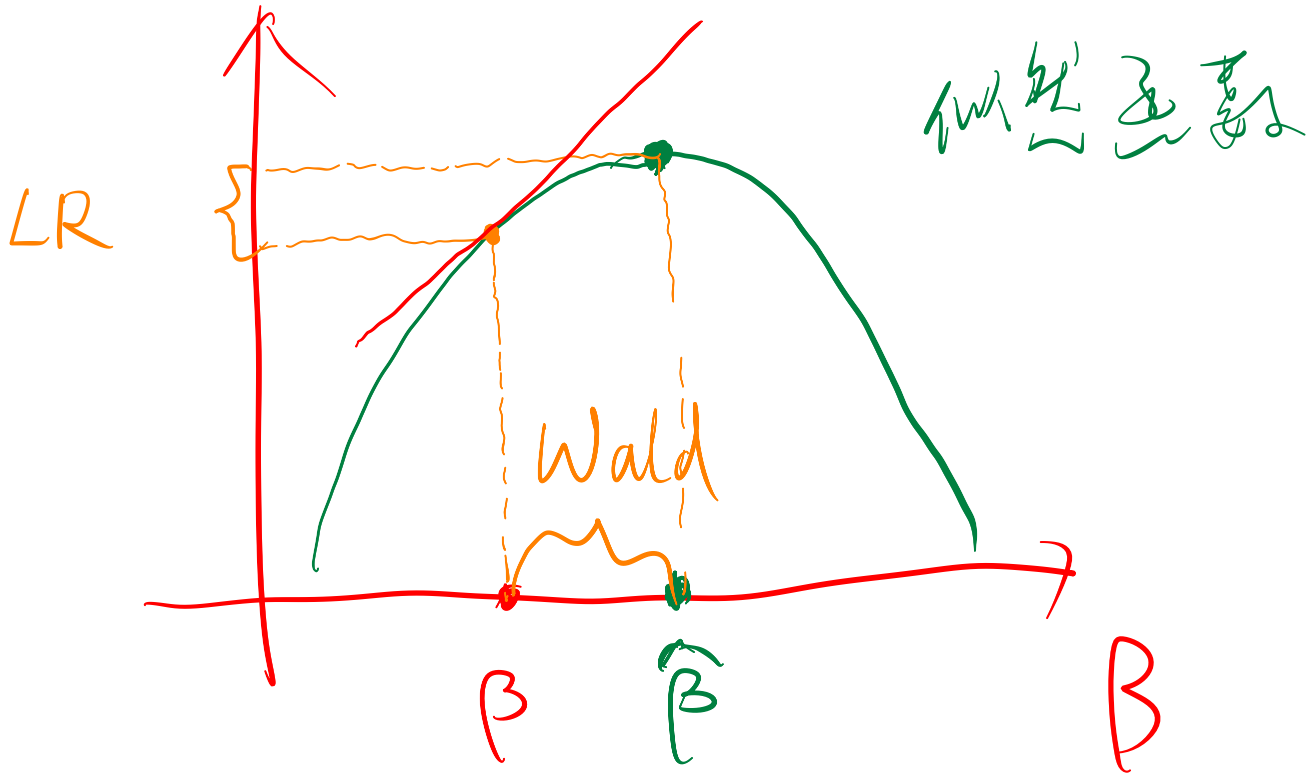
Wald, LM, LR

Likelihood Ratio

最大似然估计，

☆ 知道数据分布，

☆ 给定数据，建立出现这样数据概率 β。



$$y = x_1'\beta_1 + x_2'\beta_2 + u$$

x_1 外生, x_2 内生.

$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$, $z_1 = x_1$,

$z: L\times 1$, $x: K\times 1$, $z_1: K_1\times 1$

$z_2: K_2\times 1$

$\text{rank}(E[z'x])$.

$$E\left[\begin{bmatrix} x_1 \\ z_2 \end{bmatrix}\begin{bmatrix} x_1' & x_2' \end{bmatrix}\right] = E\begin{bmatrix} x_1x_1' & x_1x_2' \\ z_2x_1' & z_2x_2' \end{bmatrix}$$

满列秩 $\Longleftrightarrow$ $E[\tilde{z}_2\tilde{x}_2']$ 满秩.

$\tilde{z}_2 = z_2 - L(z_2|z_1)$ (残差), $\tilde{x}_2 = x_2 - L(x_2|z_1)$

$\Longleftrightarrow E[\tilde{x}_2\tilde{x}_2']^{-1}E[\tilde{x}_2\tilde{z}_2']E[\tilde{z}_2\tilde{z}_2']E[\tilde{z}_2\tilde{x}_2']$

特征值不为0.

Andersen LM.

(iid) $\text{rank}(E[\tilde{z}_2\tilde{x}_2']) = K_2 - 1$

$\Longleftrightarrow$ 最小特征值为 $r_{K_1} = 0$

$n r_{K_1}^2 \sim \chi^2(L-K+1)$

Cragg-Donald Wald

(Robust) ✓

Kleibergen & Paap (2006)

LM, Wald.

★ 只有1个 endogenous x_K 时，

Andersen，CD，KP 都 $\Longleftrightarrow$ 第一阶段回归的 F 统计量，

Robust.

“Kleibergen-Paap rk LM 统计量(Kleibergen-Paap, 2006)，其渐近分布也是$\chi^2(L-K+1)$。

2．弱工具变量检验

如果工具变量$\boldsymbol{z}$与内生解释变量$\boldsymbol{x}$完全不相关，则无法使用工具变量法，因为$\left[\mathrm{E}(\boldsymbol{z}_i\boldsymbol{x}_i')\right]^{-1}$不存在。

如果$\boldsymbol{z}$与$\boldsymbol{x}$仅仅微弱地相关，可认为$\left[\mathrm{E}(\boldsymbol{z}_i\boldsymbol{x}_i')\right]^{-1}$很大，导致 IV 估计量的渐近方差$\mathrm{Avar}(\hat{\boldsymbol{\beta}}_{\mathrm{IV}})=\left[\mathrm{E}(\boldsymbol{z}_i\boldsymbol{x}_i')\right]^{-1}\boldsymbol{S}\left[\mathrm{E}(\boldsymbol{z}_i\boldsymbol{x}_i')\right]^{-1}$变得很大。

由于$\boldsymbol{z}$中仅包含很少与$\boldsymbol{x}$有关的信息，利用这部分信息进行的 IV 估计不准确，即使样本容量很大也很难收敛到真实参数值。这种

工具变量称为“弱工具变量”(weak instruments)。

弱工具变量的后果类似于样本容量过小，导致 $\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$ 的小样本性质变得很差，而 $\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$ 的大样本分布也可能离正态分布相去甚远，致使基于大样本理论的统计推断失效。

判断弱工具变量的方法主要有以下四种。**方法之一为使用“偏 R^2”。**假设回归模型为

$$y = \boldsymbol{x}_1'\boldsymbol{\beta}_1 + x_2\beta_2 + \varepsilon$$

其中，只有 x_2 为内生解释变量。记工具变量为 $(\boldsymbol{x}_1\ \boldsymbol{z}_2)$，其中 $\boldsymbol{z}_2$ 为方程外的工具变量。

在 2SLS 的第一阶段回归中，$\underline{x_2 \xrightarrow{\text{OLS}} \boldsymbol{x}_1, \boldsymbol{z}_2}$，其 R^2 包含了内生

y x_1, x_2

变量x_2与工具变量$\boldsymbol{z}_2$相关性的信息，但也可能由于x_2与$\boldsymbol{x}_1$的相关性所造成。应使用滤去$\boldsymbol{x}_1$影响的“偏R^2”(partial R^2)，记为R_p^2。

具体步骤：首先把x_2对$\boldsymbol{x}_1$回归，$x_2 \xrightarrow{\text{OLS}} \boldsymbol{x}_1$，记其残差为$e_{x_2}$，代表$x_2$中不能由$\boldsymbol{x}_1$解释的部分；

系数实际是

$\theta_1, \cdots, \theta_M$.

其次，把$\boldsymbol{z}_2$对$\boldsymbol{x}_1$回归，$\boldsymbol{z}_2 \xrightarrow{\text{OLS}} \boldsymbol{x}_1$，记其残差为$\boldsymbol{e}_{z_2}$，代表$\boldsymbol{z}_2$中不能由$\boldsymbol{x}_1$解释的部分；

Frisch_Waugh Theorem.

最后，对两个残差进行回归，即$e_{x_2} \xrightarrow{\text{OLS}} \boldsymbol{e}_{z_2}$，所得可决系数就是$R_p^2$。具体$R_p^2$多低才构成弱工具变量，目前尚无共识。

Shea (1997)将R_p^2推广到多个内生解释变量的情形，Stata 称为

“Shea’s partial R^2”。

判断弱工具变量的方法之二为，在第一阶段回归中，$x_2 = \boldsymbol{x}_1'\boldsymbol{\gamma}_1 + \boldsymbol{z}_2\boldsymbol{\gamma}_2 + error$，检验原假设“$H_0 : \boldsymbol{\gamma}_2 = \mathbf{0}$”。

经验规则(rule of thumb)是，如果此检验的 F 统计量大于 10，则可拒绝“存在弱工具变量”的原假设，不必担心弱工具变量问题。

在多个内生解释变量的情况下，有多个第一阶段回归，故有多个 F 统计量。可使用 Stock and Yogo (2005)提出的“最小特征值统计量”(minimum eigenvalue statistic)。

如果只有一个内生解释变量，该统计量还原为 F 统计量。

判断弱工具变量的方法之三为，如果假设扰动项为 iid，可使用“Cragg-Donald Wald F 统计量”(Cragg and Donald, 1993)，其临界值由 Stock and Yogo (2005)提供。

判断弱工具变量的方法之四为，如果不作 iid 扰动项的假设，则应使用“Kleibergen-Paap Wald rk F 统计量”，其临界值也来自 Stock and Yogo (2005)。

解决弱工具变量问题的方法：

(i) 寻找更强的工具变量；

(ii) 使用对弱工具变量更不敏感的“有限信息最大似然估计法”

KP, CD, SY 给了关键值.

① maximal relative bias → $\dfrac{E[\hat{\beta}_{IV} - \beta]}{E[\hat{\beta}_{OLS} - \beta]}$ ← IV bias ← OLS bias

→ 1

→ 0. ✓

② maximal "size"

→ 第I类错误发生.

reject too often. 误认为 α 是5%, 实际 size 可能很大.

size	10.53.
20%	11.
25%	8

2SLS. 推断不那么就心.

细节：Stock & Yogo 给出的 critical value 是 iid 下的。

(Limited Information Maximum Likelihood Estimation，简记 LIML)；在大样本下，LIML 与 2SLS 是渐近等价的。在存在弱工具变量的情况下，LIML 的小样本性质可能优于 2SLS。

(iii) 如果有较多工具变量，可舍弃弱工具变量。在选择舍弃哪个工具变量时，可进行“冗余检验”(redundancy test)。

冗余工具变量(redundant instruments)：使用这些工具变量不会提高估计量的渐近效率(asymptotic efficiency)。

具体表现：在第一阶段回归中，这些工具变量的系数不显著；或者这些工具变量与内生解释变量的“偏相关系数”(partial correlations)为 0 或接近于 0(偏相关系数的平方即为偏 R^2)。

基于偏相关的思想，该检验考察内生变量与可能的冗余工具变量之间在过滤掉(partial out)其他工具变量影响之后的矩阵交叉乘积(matrix cross product)的秩是否为 0。

该冗余检验的原假设是，指定的工具变量为多余的。

该检验统计量的渐近分布为χ^2分布，自由度为“内生变量个数”乘以“冗余工具变量个数”。

E[xu]=0 ?

3．过度识别检验 外生性问题

大前提

工具变量个数=内生变量个数

在恰好识别的情况下，无法检验 IV 的外生性，只能定性讨论。

如果 IV 是外生的，则其对被解释变量发生影响的唯一渠道就是通过内生变量，除此以外别无其他渠道。

由于此唯一渠道(内生变量)已包括在回归方程中，故工具变量不会再出现在被解释变量的扰动项中，或对此扰动项有影响。此条件被称为“排他性约束”(exclusion restriction)。

需要找出 IV 影响被解释变量的所有其他可能渠道，一一排除，才能信服地说明工具变量的外生性。

在过度识别的情况下，可进行“过度识别检验”(overidentification test)。

此检验的大前提(maintained hypothesis)是该模型至少是恰好识别的，即有效工具变量至少与内生解释变量一样多。

“Hansen J”

在此大前提下，过度识别检验的原假设为“H_0:所有工具变量都是外生的”。

希望接受H_0

如果拒绝该原假设，则认为至少某个变量与扰动项相关。

假设前$(K-r)$个解释变量$\{x_1, \cdots, x_{K-r}\}$为外生解释变量，而后 r 个解释变量$\{x_{K-r+1}, \cdots, x_K\}$为内生解释变量。

假设共有 m 个方程外的工具变量$\{z_1, \cdots, z_m\}$，其中$m > r$。

把工具变量法的残差对所有外生变量(即所有外生解释变量与IV)进行以下辅助回归：

θ_1 θ_m

$$e_{i,\mathrm{IV}} = \gamma_1 x_{i1} + \cdots + \gamma_{K-r} x_{i,K-r} + \boxed{\delta_1} z_{i1} + \cdots + \boxed{\delta_m} z_{im} + error_i$$

原假设可写为“$H_0: \delta_1 = \cdots = \delta_m = 0$”。记此辅助回归的可决系数为$R^2$，则 Sargan 统计量为

$$nR^2 \xrightarrow{d} \chi^2(m-r)$$

其中，χ^2分布的自由度$(m-r)$为过度识别约束的个数(即“多余”的工具变量个数)。

如果恰好识别，则$m-r=0$(自由度为 0)，$\chi^2(0)$无定义，故无法使用这个“过度识别检验”。

即使接受了过度识别的原假设，也并不能证明这些工具变量的外生性。

过度识别检验的成立有一个大前提，即至少该模型是恰好识别的。此大前提无法检验，只能假定其成立。

异方差时，“Hansen J”，χ^2 J ←

↓

GMM. “Generalized Methods of Moments”

4．究竟该用 OLS 还是工具变量法：对解释变量内生性的检验

“豪斯曼检验”(Hausman specification test)(Hausman, 1978)的原假设为“H_0：所有解释变量均为外生变量”。

如果H_0成立，则 OLS 与 IV 都一致，在大样本下$\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$与$\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$都收敛于真实的参数值$\boldsymbol{\beta}$，故$(\hat{\boldsymbol{\beta}}_{\mathrm{IV}}-\hat{\boldsymbol{\beta}}_{\mathrm{OLS}})$依概率收敛于 $\mathbf{0}$。

反之，如果H_0不成立，则 IV 一致而 OLS 不一致，故$(\hat{\boldsymbol{\beta}}_{\mathrm{IV}}-\hat{\boldsymbol{\beta}}_{\mathrm{OLS}})$不会收敛于 $\mathbf{0}$。

如果$(\hat{\boldsymbol{\beta}}_{\mathrm{IV}}-\hat{\boldsymbol{\beta}}_{\mathrm{OLS}})$的距离很大，则倾向于拒绝原假设。

根据沃尔德检验原理，以二次型来度量此距离：

$$
(\hat{\boldsymbol{\beta}}_{\text{IV}} - \hat{\boldsymbol{\beta}}_{\text{OLS}})' \boldsymbol{D}^{-} (\hat{\boldsymbol{\beta}}_{\text{IV}} - \hat{\boldsymbol{\beta}}_{\text{OLS}}) \xrightarrow{d} \chi^2(r)
$$

其中，$\boldsymbol{D} \equiv \text{Var}(\hat{\boldsymbol{\beta}}_{\text{IV}} - \hat{\boldsymbol{\beta}}_{\text{OLS}})$，而 $\boldsymbol{D}^{-}$ 为 $\boldsymbol{D}$ 的广义逆矩阵(因为 $\boldsymbol{D}$ 不一定可逆；但如果 $\boldsymbol{D}$ 正定，则可逆)。当 $\boldsymbol{D}$ 可逆时，广义逆矩阵就是一般的逆矩阵，即 $\boldsymbol{D}^{-} = \boldsymbol{D}^{-1}$。

r 为内生解释变量的个数(不包括外生解释变量)。

$$
\text{Var}(\hat{\boldsymbol{\beta}}_{\text{IV}} - \hat{\boldsymbol{\beta}}_{\text{OLS}}) = \text{Var}(\hat{\boldsymbol{\beta}}_{\text{IV}}) + \text{Var}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) - 2\text{Cov}(\hat{\boldsymbol{\beta}}_{\text{IV}}, \hat{\boldsymbol{\beta}}_{\text{OLS}})
$$

同方差

如果在 H_0 成立的情况下，且 OLS 最有效率，则可证明 $\text{Cov}(\hat{\boldsymbol{\beta}}_{\text{IV}}, \hat{\boldsymbol{\beta}}_{\text{OLS}}) = \text{Var}(\hat{\boldsymbol{\beta}}_{\text{OLS}})$，故

假如

$$\boldsymbol{D} = \widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\mathrm{IV}})} - \widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\mathrm{OLS}})}$$

上述检验的缺点是，它假设在H_0成立的情况下，OLS 最有效率。

如果存在异方差，OLS 并不最有效率(不是 BLUE)。故传统的豪斯曼检验不适用于异方差的情形。

解决方法之一为，通过“自助法”(bootstrap)，即计算机模拟“再抽样”(resampling)的方法来计算$\boldsymbol{D} \equiv \widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\mathrm{IV}} - \hat{\boldsymbol{\beta}}_{\mathrm{OLS}})}$。

解决方法之二为，使用“杜宾－吴－豪斯曼检验”(Durbin−Wu−Hausman Test，简记 DWH)，该检验在异方差的情况下也适用，更为稳健。

GMM.

w

Hayashi (2000). 林文夫 Econometrics.

假设模型为，$y = \boldsymbol{x}_1'\boldsymbol{\beta}_1 + x_2\beta_2 + \varepsilon$，其中 x_2 为唯一的内生解释变量。

记工具变量为 $\boldsymbol{z} \equiv (\boldsymbol{x}_1\ \boldsymbol{z}_2)$，其中 $\boldsymbol{z}_2$ 为方程外的工具变量。

考虑 2SLS 的第一阶段回归，即 $x_2 = \boldsymbol{x}_1'\boldsymbol{\gamma} + \boldsymbol{z}'\boldsymbol{\delta} + v$。

由于工具变量 $\boldsymbol{z}$ 与 ε 不相关，故

$$\mathrm{E}(x_2\varepsilon) = \mathrm{E}\left[(\boldsymbol{x}_1'\boldsymbol{\gamma} + \boldsymbol{z}'\boldsymbol{\delta} + v)\varepsilon\right] = \underbrace{\mathrm{E}(\boldsymbol{x}_1'\boldsymbol{\gamma}\varepsilon)}_{=0} + \underbrace{\mathrm{E}(\boldsymbol{z}'\boldsymbol{\delta}\varepsilon)}_{=0} + \mathrm{E}(v\varepsilon) = \mathrm{E}(v\varepsilon)$$

$\neq 0$.　　　　$\neq 0$.

这意味着

$$“x_2 \text{为内生变量}” \Leftrightarrow “\mathrm{E}(x_2\varepsilon) \neq 0” \Leftrightarrow “\mathrm{E}(v\varepsilon) \neq 0”$$

故只需检验第一阶段回归的扰动项v是否与原模型的扰动项ε相关即可：

$E[v\varepsilon]$?

$$\varepsilon = \rho v + \xi$$

如果ε与v不相关，则$\rho = 0$。代入原模型可得

ε

$$y = \boldsymbol{x}_1'\boldsymbol{\beta}_1 + x_2\beta_2 + \rho v + \xi$$

使用第一阶段回归的残差$\hat{v}$来代替v，进行辅助回归：

$$y = \boldsymbol{x}_1'\boldsymbol{\beta}_1 + x_2\beta_2 + \hat{v}\rho + error$$

Robust.

对原假设"$H_0: \rho = 0$"进行 t 检验。如果拒绝"$H_0: \rho = 0$"，则

认为存在内生解释变量；否则，认为所有解释变量均为外生。

考虑到可能存在异方差，则在作 t 检验时使用稳健标准误。

如果存在多个内生解释变量，即 $y = \boldsymbol{x}_1'\boldsymbol{\beta}_1 + \boldsymbol{x}_2'\boldsymbol{\beta}_2 + \varepsilon$。

则在第一阶段回归中可得到与内生解释变量 $\boldsymbol{x}_2$ 相对应的多个残差 $\hat{\boldsymbol{v}}$，进行以下辅助回归，

$$y = \boldsymbol{x}_1'\boldsymbol{\beta}_1 + \boldsymbol{x}_2'\boldsymbol{\beta}_2 + \hat{\boldsymbol{v}}'\boldsymbol{\rho} + error$$

对原假设“$H_0 : \boldsymbol{\rho} = \boldsymbol{0}$”进行 F 检验即可。如果担心存在异方差，则可在作 F 检验时使用稳健标准误。

本章小节

- 在这一章中，我们介绍了内生性以及解决办法（工具变量方法)
- 我们首先讨论了内生性的成因以及影响
- 工具变量必须具备两个性质:(1)它必须是外生的（2）它必须与内生解释变量相关
- 当工具变量个数等于内生变量个数时，我们可以从矩条件出发获得一致估计值
- 当工具变量个数大于或等于内生变量个数时，我们可以使用2SLS方法
- 我们讨论了弱工具变量检验以及内生性检验问题