第 10 章　工具变量法

OLS 能够成立的最重要条件是解释变量与扰动项不相关(即前定变量或同期外生的假设)。

但解释变量与扰动项相关(内生性)的例子比比皆是。

解决内生性的主要方法之一为工具变量法。

内生性的主要来源包括遗漏变量偏差、联立方程偏差(双向因果关系)，以及测量误差偏差(measurement error bias)。

10.1 联立方程偏差

例 考察如下农产品市场均衡模型：

$$\begin{cases} q_t^d = \alpha + \beta p_t + u_t & \text{(需求)} \\ q_t^s = \gamma + \delta p_t + v_t & \text{(供给)} \\ q_t^d = q_t^s & \text{(均衡)} \end{cases} \qquad (10.1)$$

q_t^d为农产品需求，q_t^s为农产品供给，p_t为农产品价格。市场出清的均衡条件要求$q_t^d = q_t^s$。令$q_t \equiv q_t^d = q_t^s$，可得

$$\begin{cases} q_t = \alpha + \beta p_t + u_t \\ q_t = \gamma + \delta p_t + v_t \end{cases} \qquad (10.2)$$

这两个方程的被解释变量与解释变量完全一样。

如果直接作回归$q_t \xrightarrow{\text{OLS}} p_t$，估计的究竟是需求函数还是供给函数呢？参见图 10.1。

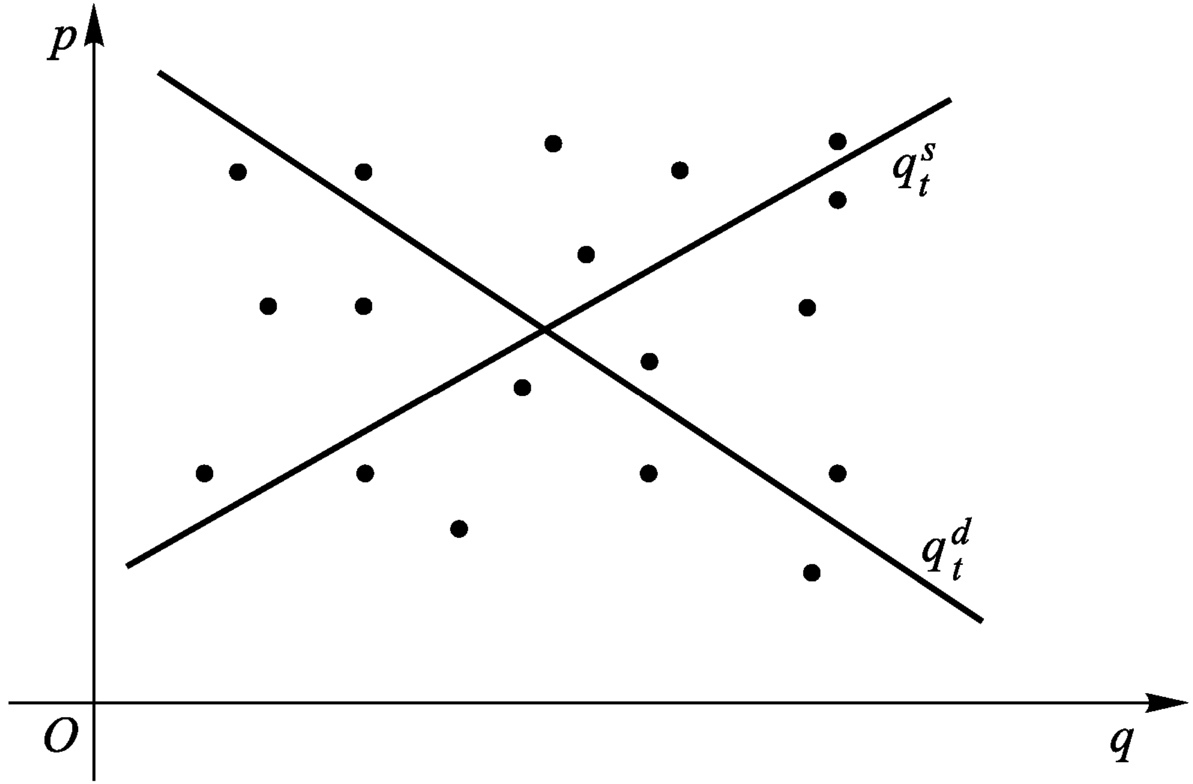

图 10.1　需求与供给决定市场均衡

从数据生成过程的视角，可把线性方程组(10.2)中的(p_t, q_t)看成是未知数(内生变量)，而把(u_t, v_t)看作已知，然后求解(p_t, q_t)为(u_t, v_t)的函数。

由此可知，解释变量p_t与两个方程的扰动项(u_t, v_t)都相关，即$\text{Cov}(p_t, u_t) \neq 0$，$\text{Cov}(p_t, v_t) \neq 0$。

对于需求函数的正冲击$(u_t > 0)$，将使得均衡价格p_t上升，故二者正相关。

对于供给函数的正冲击$(v_t > 0)$，将使得均衡价格p_t下降，故二者负相关。

因此，OLS 估计值$(\hat{\beta}, \hat{\delta})$不是$(\beta, \delta)$的一致估计量。

称这种偏差为**联立方程偏差**(simultaneity bias)或**内生性偏差**(endogeneity bias)。

例 考察宏观经济模型中的消费函数：

$$\begin{cases} C_t = \alpha + \beta Y_t + \varepsilon_t \\ Y_t = C_t + I_t + G_t + X_t \end{cases} \qquad (10.3)$$

Y_t, C_t, I_t, G_t, X_t分别为国民收入、总消费、总投资、政府净支出与净出口。

如果单独对消费方程进行 OLS 回归，将存在联立方程偏差，得不到一致估计。

10.2 测量误差偏差

导致内生性的另一来源是**解释变量的测量误差**(measurement error 或 errors-in-variables)。

例 假设真实模型为

$$y = \alpha + \beta x^* + \varepsilon \tag{10.4}$$

其中，$\beta \neq 0$，$\mathrm{Cov}(x^*, \varepsilon) = 0$。但$x^*$无法精确观测，而只能观测到$x$，二者满足如下关系：

$$x = x^* + u \tag{10.5}$$

其中，$\mathrm{Cov}(x^*, u) = 0$(测量误差u与被测量变量x^*不相关)，$\mathrm{Cov}(u, \varepsilon) = 0$(测量误差与扰动项$\varepsilon$不相关)。

将表达式(10.5)代入方程(10.4)可得

$$y = \alpha + \beta x + (\varepsilon - \beta u) \qquad (10.6)$$

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

$$\begin{aligned}\mathrm{Cov}(x, \varepsilon - \beta u) &= \mathrm{Cov}(x^* + u, \varepsilon - \beta u) \\ &= \underbrace{\mathrm{Cov}(x^*, \varepsilon)}_{=0} - \beta \underbrace{\mathrm{Cov}(x^*, u)}_{=0} + \underbrace{\mathrm{Cov}(u, \varepsilon)}_{=0} - \beta\,\mathrm{Cov}(u, u) \\ &= -\beta\,\mathrm{Var}(u) \neq 0\end{aligned} \qquad (10.7)$$

因此，OLS 估计不一致。

由于解释变量测量误差所造成的 OLS 估计偏差，称为**测量误差偏差**(measurement error bias)。

如果被解释变量存在测量误差，后果却不严重。

比如，只要被解释变量的测量误差与解释变量不相关，则 OLS 依然一致(参见习题)。

10.3 工具变量法

OLS 的不一致性是由于内生变量与扰动项相关而引起。

如果能够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，则可用与扰动项不相关的那部分得到一致估计。

对内生变量的这种分离可借助于对内生变量的深入认识来完成，而更常见的方法则借助另外一个“工具变量”来实现。

回到农产品市场均衡的例子。

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。

此时，就可以估计需求曲线，参见图 10.2。这个使得供给曲线移动的变量就是工具变量。

假设影响方程组(10.2)中供给方程扰动项的因素可分解为两部分，即可观测的气温z_t与不可观测的其他因素：

$$q_t^s = \gamma + \delta p_t + \eta z_t + v_t \qquad (10.8)$$

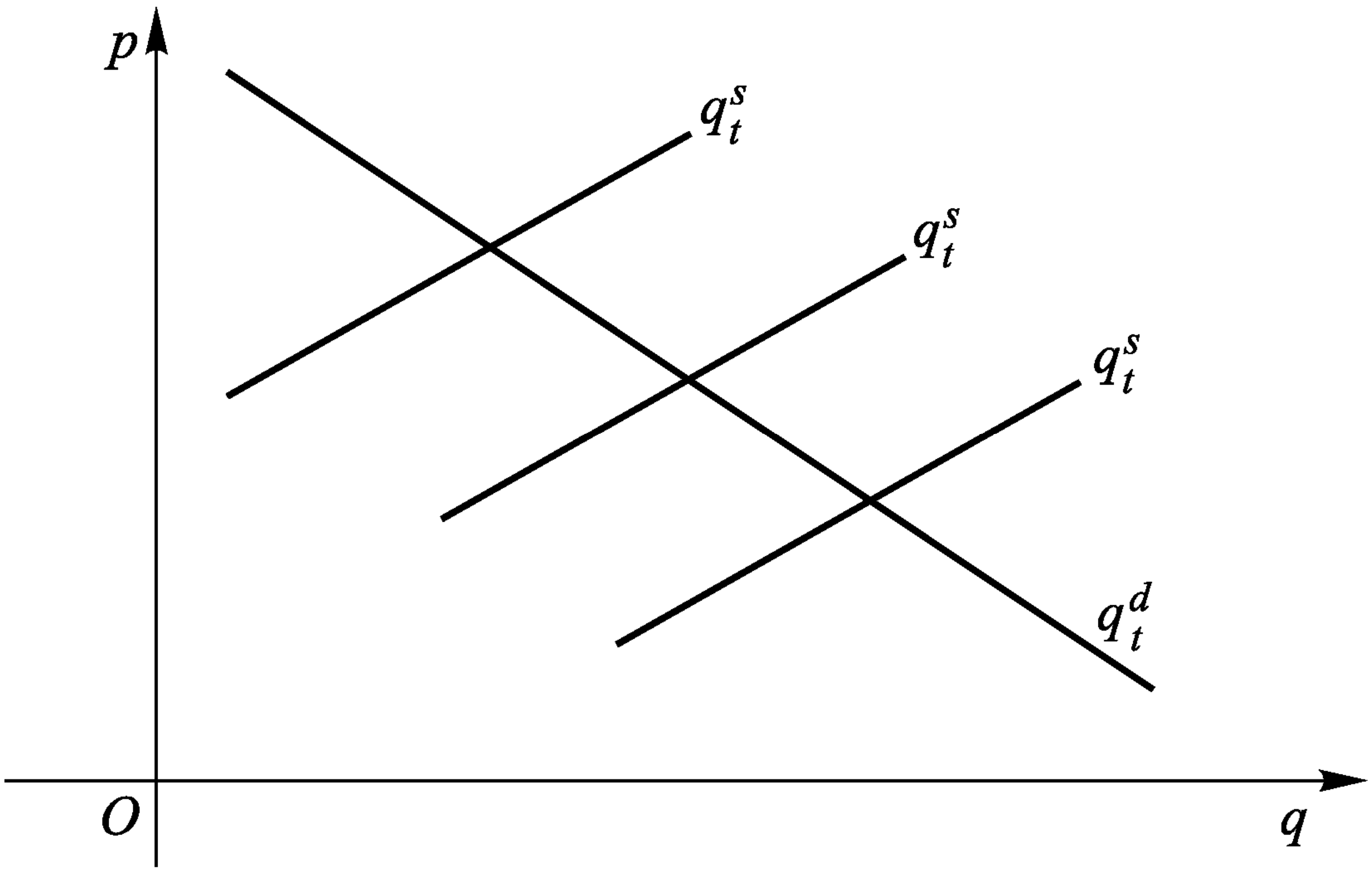

图 10.2　稳定的需求与变动的供给

假定气温 z_t 是前定变量，与需求方程的扰动项不相关，即 $\text{Cov}(z_t, u_t) = 0$。

由于气温 z_t 的变化使得供给函数 q_t^s 沿着需求函数 q_t^d 移动，故可以估计出需求函数 q_t^d。

在这种情况下，称 z_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**(relevance)：工具变量与内生解释变量相关，即 $\text{Cov}(z_t, p_t) \neq 0$。

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

在本例中，气温z_t满足这两个条件。

(i) 相关性：从联立方程组可以解出 $p_t = p_t(z_t, u_t, v_t)$，故 $\text{Cov}(z_t, p_t) \neq 0$。

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

利用工具变量的这两个性质，可得到对需求方程回归系数β的一致估计。

回顾需求方程为

$$q_t = \alpha + \beta p_t + u_t \tag{10.9}$$

在方程两边同时求与z_t的协方差：

$$\begin{aligned}\mathrm{Cov}(q_t, z_t) &= \mathrm{Cov}(\alpha + \beta p_t + u_t, z_t) \\ &= \beta \mathrm{Cov}(p_t, z_t) + \underbrace{\mathrm{Cov}(u_t, z_t)}_{=0} = \beta \mathrm{Cov}(p_t, z_t)\end{aligned} \tag{10.10}$$

其中，由于工具变量的外生性，故$\mathrm{Cov}(u_t, z_t) = 0$。

根据工具变量的相关性，$\mathrm{Cov}(p_t, z_t) \neq 0$。

把上式两边同除以$\mathrm{Cov}(p_t, z_t)$可得

$$\beta = \frac{\mathrm{Cov}(q_t, z_t)}{\mathrm{Cov}(p_t, z_t)} \qquad (10.11)$$

上式相当于总体矩条件。

以相应的样本矩取代上式的总体矩(即以样本协方差替代总体协方差)，可得一致的**工具变量估计量**(Instrumental Variable Estimator)：

$$\hat{\beta}_{\mathrm{IV}} = \frac{\widehat{\mathrm{Cov}(q_t, z_t)}}{\widehat{\mathrm{Cov}(p_t, z_t)}} = \frac{\sum_{t=1}^{n}(q_t - \overline{q})(z_t - \overline{z})}{\sum_{t=1}^{n}(p_t - \overline{p})(z_t - \overline{z})} \xrightarrow{p} \frac{\mathrm{Cov}(q_t, z_t)}{\mathrm{Cov}(p_t, z_t)} = \beta \qquad (10.12)$$

其中，$\overline{q}, \overline{p}, \overline{z}$ 分别为q, p, z的相应样本均值。

由于样本矩为总体矩的一致估计，故工具变量估计量$\hat{\beta}_{\text{IV}}$是真实参数β的一致估计。

如果工具变量与内生变量无关，即$\text{Cov}(z_t, p_t) = 0$，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，即$\text{Cov}(z_t, p_t) \approx 0$，会导致估计量$\hat{\beta}_{\text{IV}}$的方差变得很大，称为“弱工具变量问题”。

10.4　二阶段最小二乘法

工具变量法一般通过**二阶段最小二乘法**(Two Stage Least Square，简记 2SLS 或 TSLS)来实现(Theil, 1953; Basmann, 1957)，顾名思义，即通过作两个回归来完成。

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

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 $q_t \xrightarrow{\text{OLS}} \hat{p}_t$。

为什么这样做能得到好结果呢？

把需求方程 $q_t = \alpha + \beta p_t + u_t$ 分解为

$$q_t = \alpha_0 + \beta \hat{p}_t + \underbrace{\left[u_t + \beta(p_t - \hat{p}_t)\right]}_{\equiv \varepsilon_t} \qquad (10.13)$$

上式就是在第二阶段所作的回归，其扰动项为 $\varepsilon_t \equiv u_t + \beta(p_t - \hat{p}_t)$。

命题　在第二阶段回归中，$\hat{p}_t$与扰动项ε_t不相关。

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

$$\mathrm{Cov}(\hat{p}_t, \varepsilon_t) = \mathrm{Cov}(\hat{p}_t, u_t) + \beta\,\mathrm{Cov}(\hat{p}_t,\, p_t - \hat{p}_t) \qquad (10.14)$$

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

其次，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估计量的正交性)，故上式右边的第二项$\mathrm{Cov}(\hat{p}_t,\, p_t - \hat{p}_t) = 0$。

因此，第二阶段回归的解释变量$\hat{p}_t$与扰动项ε_t不相关，故 2SLS 能得到一致估计。

2SLS 的实质是把内生解释变量p_t分成两部分，即由工具变量z_t所造成的外生部分($\hat{p}_t$)以及与扰动项相关的其余部分($p_t - \hat{p}_t$)；

然后，把被解释变量q_t对p_t中的外生部分($\hat{p}_t$)进行回归，从而满足 OLS 对前定变量的要求而得到一致估计。

如果存在多个工具变量，也不难应用 2SLS 法。

假设z_1与z_2为两个有效工具变量(都满足相关性与外生性)，则第一阶段回归变为

$$p = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + u \qquad (10.15)$$

由此可得拟合值$\hat{p} = \hat{\alpha}_0 + \hat{\alpha}_1 z_1 + \hat{\alpha}_2 z_2$，而第二阶段回归不变。

考虑存在多个内生变量的情形，比如

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \qquad (10.16)$$

其中，x_1与x_2均为内生解释变量，都与ε相关。

由于有两个内生变量，则至少需要两个工具变量，才能进行2SLS 估计。

如果只有一个工具变量 z，则由第一阶段回归可得，$\hat{x}_1 = \hat{\alpha}_0 + \hat{\alpha}_1 z$，而$\hat{x}_2 = \hat{\gamma}_0 + \hat{\gamma}_1 z$。将$\hat{x}_1$与$\hat{x}_2$代入原方程：

$$y = \beta_0 + \beta_1 \hat{x}_1 + \beta_2 \hat{x}_2 + v_t \qquad (10.17)$$

由于$\hat{x}_1$与$\hat{x}_2$都是z的线性函数，故此方程存在严格多重共线性，即可以找到$\hat{x}_1$与$\hat{x}_2$的线性组合为常数(参见习题)。

如果存在两个内生变量，至少需要两个工具变量，才能进行工具变量法的估计。

阶条件：进行 2SLS 估计的必要条件是工具变量个数不少于内生解释变量的个数，称为**阶条件**(order condition)。

根据阶条件是否满足可分为以下三种情况：

(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。

更一般地，考虑多个内生变量，且包含外生解释变量的情形。比如，

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 w + \varepsilon \quad (10.18)$$

其中，x_1与x_2为内生变量，而w为外生解释变量(与扰动项ε不相关)。

假设有三个有效工具变量z_1, z_2, z_3。

在 2SLS 的第一阶段回归中，应分别将两个内生解释变量(x_1, x_2)对所有外生变量(包括工具变量z_1, z_2, z_3及外生解释变量w)回归：

$$x_1 = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + \alpha_3 z_3 + \alpha_4 w + u \quad (10.19)$$

$$x_2 = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \gamma_3 z_3 + \gamma_4 w + v \quad (10.20)$$

外生解释变量w可视为自己的工具变量。

首先，w显然与w高度相关，故满足相关性。

其次，w与扰动项ε不相关(因为w为外生解释变量)。

若第一阶段回归不包括外生解释变量w，则会导致偏差。

若w不出现于第一阶段回归，则无法保证w与第一阶段回归的残差正交；

而w仍出现于第二阶段回归，也就无法保证w与第二阶段回归的残差不相关。

有时也称z_1, z_2, z_3为方程外的工具变量。

将方程(10.19)与(10.20)的拟合值分别记为$\hat{x}_1$与$\hat{x}_2$，并代入原方程(10.18)，进行第二阶段回归：

$$y = \beta_0 + \beta_1 \hat{x}_1 + \beta_2 \hat{x}_2 + \beta_3 w + \xi \qquad (10.21)$$

其中，ξ 为第二阶段回归的扰动项。记此估计量为 $\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$。

$\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$ 为真实参数 $\boldsymbol{\beta}$ 的一致估计，且服从渐近正态分布，可以照常进行大样本统计推断。

由于 2SLS 的第二阶段回归就是 OLS，故 $\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$ 的协方差矩阵 $\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\mathrm{IV}})$ 在形式上与 OLS 估计量相似。

考虑到可能存在异方差，建议使用异方差稳健的标准误。

需要注意的是，第二阶段回归所得残差为

$$\hat{\xi} \equiv y - \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_1 + \hat{\beta}_2 \hat{x}_2 + \hat{\beta}_3 w \qquad (10.22)$$

而原方程真正的残差却是

$$e_{IV} \equiv y - \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 w \qquad (10.23)$$

一般来说，二者并不相等，即$e_{IV} \neq \hat{\xi}$。

进行 2SLS 估计，最好不要自己手工进行两次回归，而直接使用 Stata 命令。

2SLS 的 Stata 命令格式为

```
. ivregress 2sls y x1 x2 (x3 = z1 z2),robust first
```

其中，“y”为被解释变量，“x1 x2”为外生解释变量，

“x3”为内生解释变量，而“z1 z2”为方程外的工具变量。

选择项“robust”表示使用异方差稳健的标准误(默认为普通标准误)。

选择项“first”表示显示第一阶段的回归结果。

在球形扰动项的情况下，2SLS 是最有效率的工具变量法。

在异方差的情况下，则存在更有效率的工具变量法，即**广义矩估计**(Generalized Method of Moments，简记 GMM)，是数理统计中**矩估计**(Method of Moments，简记 MM)的推广。

GMM 之于 2SLS，正如 GLS 与 OLS 的关系。

在恰好识别的情况，GMM 等价于 2SLS。

10.5 弱工具变量

如果工具变量与内生解释变量仅微弱地相关，则工具变量法估计量 $\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$ 的方差将变得很大。

由于工具变量仅包含极少与内生解释变量有关的信息，利用这部分信息进行的工具变量法估计就不准确，即使样本容量很大也很难收敛到真实的参数值。

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，会导致$\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$的小样本性质变得很差。

$\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$的小样本真实分布离大样本的渐近正态分布相去甚远，致使基于大样本理论的统计推断失效。

为了检验是否存在弱工具变量，可在第一阶段回归中，检验所有方程外的工具变量(不含外生解释变量)的系数是否联合为零。

假设原模型为

$$y = \beta_0 + \beta_1 x + \beta_2 w + \varepsilon \qquad (10.24)$$

其中，x为内生变量，而w为外生解释变量。

假设有两个有效工具变量z_1, z_2，则第一阶段回归为

$$x = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + \alpha_3 w + u \qquad (10.25)$$

然后检验$H_0 : \alpha_1 = \alpha_2 = 0$，即工具变量$z_1, z_2$的系数联合为 0。

由于工具变量的强弱程度可连续变化，很难确定明确的标准。

经验规则(rule of thumb)：

如果此检验的F统计量大于 10[①]，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。

① 由于技术性原因，此处应使用普通标准误，而非异方差稳健的标准误(Stock and Watson, 2012, p. 507)。

在 Stata 中作完 2SLS 回归后，可使用以下命令检验弱工具变量：

```
. estat firststage
```

此命令将根据第一阶段回归计算一些统计量，包括上文的F统计量。

如果发现存在弱工具变量，可能的解决方法包括：

(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** (Limited Information Maximum Likelihood Estimation，简记 LIML)。

在大样本下，LIML 与 2SLS 渐近等价。

在弱工具变量的情况下，LIML 的小样本性质可能优于 2SLS。

LIML 的 Stata 命令为

```
. ivregress liml y x1 x2(x3 = z1 z2)
```

此命令在格式上与“`ivregress 2sls`”(二段最小二乘法)完全相同。

10.6 对工具变量外生性的过度识别检验

工具变量的外生性是保证 2SLS 一致性的重要条件。

如果所使用的“工具变量”与扰动项相关，则可能导致严重的偏差。

在恰好识别的情况下，无法检验工具变量的外生性。只能进行定性讨论或依赖于专家的意见。

定性讨论通常基于以下逻辑：如果工具变量是外生的，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)已被包括在回归方程中，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。

此条件称为**排他性约束**(exclusion restriction)，因为它排除了工具变量除了通过内生变量而影响被解释变量的所有其他渠道。

在实际操作中，需要找出工具变量影响被解释变量的所有其他可能渠道，然后一一排除。

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。

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”。

如果拒绝该原假设，则认为至少某个变量不是外生的，与扰动项相关。

假设共有 K 个解释变量$\{x_1, \cdots, x_K\}$，其中前$(K-r)$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：

$$y = \underbrace{\beta_1 x_1 + \cdots + \beta_{K-r} x_{K-r}}_{\text{外生}} + \underbrace{\beta_{K-r+1} x_{K-r+1} + \cdots \beta_K x_K}_{\text{内生}} + \varepsilon \qquad (10.26)$$

假设共有m个方程外的工具变量$\{z_1, \cdots, z_m\}$，其中$m > r$；则过度识别的原假设为

$$H_0: \mathrm{Cov}(z_1, \varepsilon) = 0, \ \cdots, \ \mathrm{Cov}(z_m, \varepsilon) = 0 \qquad (10.27)$$

由于扰动项ε无法观测，故只能通过 2SLS 的残差e_{IV}来考察工具变量与扰动项的相关性。

把 2SLS 的残差e_{IV}对所有外生变量(即所有外生解释变量与工具变量)进行以下辅助回归：

$$e_{IV} = \gamma_1 x_1 + \cdots + \gamma_{K-r} x_{K-r} + \delta_1 z_1 + \cdots + \delta_m z_m + error \qquad (10.28)$$

原假设(10.27)可写为

$$H_0: \delta_1 = \cdots = \delta_m = 0 \qquad (10.29)$$

记辅助回归(10.28)的可决系数为R^2，则 Sargan 统计量(Sargan, 1958)为

$$nR^2 \xrightarrow{d} \chi^2(m-r) \qquad (10.30)$$

其中，Sargan 统计量的渐近分布为$\chi^2(m-r)$，其自由度$(m-r)$是过度识别约束的个数，即方程外工具变量个数(m)，减去内生解释变量个数(r)。

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

此检验的直观思想：在过度识别的情况下，可使用不同的工具变量组合来进行工具变量法估计；而如果所有工具变量都有效，则这些工具变量估计量$\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$都将收敛到相同的真实参数$\boldsymbol{\beta}$。

可以检验不同的工具变量估计量之间的差是否收敛于$\mathbf{0}$；如果不是，则说明这些工具变量不全是有效的。

在恰好识别的情况下，只有唯一的工具变量估计量，无法进行这种比较，故过度识别检验失效。

如果拒绝原假设，过度识别检验并不能告诉我们，哪些工具变量是无效的。

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。

过度识别检验成立的大前提是，该模型至少是恰好识别的。此大前提无法检验，只能假定其成立。

如果只有一个内生变量，则隐含地假定至少有一个工具变量是外生的，然后检验所有其他工具变量的外生性。

即使不同的工具变量估计量$\hat{\boldsymbol{\beta}}_{\text{IV}}$的概率极限都相同，并不能保证它们都收敛到真实的参数$\boldsymbol{\beta}$；也可能都收敛到其他值，比如$\boldsymbol{\beta}^* \neq \boldsymbol{\beta}$。

恰好识别的大前提则保证了，在这些工具变量估计量中至少有一个估计量收敛到真实参数。

此时，如果所有工具变量都外生，则所有工具变量估计量都会收敛到真实参数。

在 Stata 中作完 2SLS 估计后，可用以下命令进行过度识别检验

```
. estat overid
```

10.7　对解释变量内生性的豪斯曼检验：究竟该用 OLS 还是 IV

使用工具变量法的前提是存在内生解释变量，这也需要检验。

如何从统计上检验解释变量是否为内生呢？

扰动项不可观测，无法直接检验解释变量与扰动项的相关性。

如果找到有效的工具变量，则可借助工具变量来检验解释变量的内生性。

假设存在方程外的工具变量。

如果所有解释变量都是外生变量，则 OLS 比工具变量法更有效。

在这种情况下使用工具变量法，虽然估计量一致，但相当于“无病用药”，反而会增大估计量的方差。

反之，如果存在内生解释变量，则 OLS 不一致，而工具变量法是一致的。

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 与工具变量法都一致。

在大样本下$\hat{\boldsymbol{\beta}}_{\text{IV}}$与$\hat{\boldsymbol{\beta}}_{\text{OLS}}$都收敛于真实的参数值$\boldsymbol{\beta}$，因此$(\hat{\boldsymbol{\beta}}_{\text{IV}} - \hat{\boldsymbol{\beta}}_{\text{OLS}})$(称为“对比向量”，vector of contrast)依概率收敛于 $\mathbf{0}$。

反之，如果H_0不成立，则工具变量法一致而 OLS 不一致，故$(\hat{\boldsymbol{\beta}}_{\text{IV}}-\hat{\boldsymbol{\beta}}_{\text{OLS}})$不会收敛于 $\mathbf{0}$。

如果$(\hat{\boldsymbol{\beta}}_{\text{IV}}-\hat{\boldsymbol{\beta}}_{\text{OLS}})$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可得

$$(\hat{\boldsymbol{\beta}}_{\text{IV}}-\hat{\boldsymbol{\beta}}_{\text{OLS}})'\left[\widehat{\text{Var}(\hat{\boldsymbol{\beta}}_{\text{IV}}-\hat{\boldsymbol{\beta}}_{\text{OLS}})}\right]^{-1}(\hat{\boldsymbol{\beta}}_{\text{IV}}-\hat{\boldsymbol{\beta}}_{\text{OLS}})\xrightarrow{d}\chi^2(r) \tag{10.31}$$

其中，r为内生解释变量的个数，$\left[\widehat{\text{Var}(\hat{\boldsymbol{\beta}}_{\text{IV}}-\hat{\boldsymbol{\beta}}_{\text{OLS}})}\right]$为对比向量$(\hat{\boldsymbol{\beta}}_{\text{IV}}-\hat{\boldsymbol{\beta}}_{\text{OLS}})$的协方差矩阵之样本估计值。

如果此豪斯曼统计量很大，超过了其渐近分布$\chi^2(r)$的临界值，则可拒绝“所有解释变量均外生”的原假设，认为存在内生解释变量，故应使用 IV。

豪斯曼检验的 Stata 命令为

. reg y x1 x2

. estimates store ols (存储 OLS 的结果，记为 ols)

. ivregress 2sls y x1 (x2=z1 z2) (假设 x2 为内生变量，z1,z2 为 IV)

. estimates store iv (存储 2SLS 的结果，记为 iv)

. hausman iv ols,constant sigmamore (根据存储结果进行豪斯曼检验)

选择项“`sigmamore`”表示统一使用更有效率的估计量(即 OLS)所对应的残差来计算扰动项方差$\hat{\sigma}^2$。这样有助于保证根据样本数据计算的$\left[\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\mathrm{IV}}-\hat{\boldsymbol{\beta}}_{\mathrm{OLS}})}\right]$为正定矩阵，便于求其逆矩阵。

选择项“`constant`”表示$\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$与$\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$中都包括常数项(默认不含常数项)。

传统豪斯曼检验的缺点：为简化矩阵$\left[\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\mathrm{IV}}-\hat{\boldsymbol{\beta}}_{\mathrm{OLS}})}\right]$的计算，假设在$H_0$成立的情况下，OLS 最有效率，故不适用异方差的情形(OLS 只在球形扰动项的情况下才最有效率)。

改进的**杜宾-吴-豪斯曼检验**(Durbin–Wu–Hausman Test，简记 DWH)即使在异方差的情况下也适用。

在 Stata 中作完 2SLS 估计后，可输入以下命令进行异方差稳健的 DWH 检验：

```
. estat endogenous
```

10.8 如何获得工具变量

使用工具变量法的前提是存在有效的工具变量。

如何寻找工具变量十分重要。

工具变量的两个要求(相关性与外生性)常常自相矛盾。

寻找合适的工具变量比较困难，需要一定的创造性与想象力。

寻找工具变量的步骤大致可以分为两步：

(i) 列出与内生解释变量(x)相关的尽可能多的变量的清单(这一步较容易)；

(ii) 从这一清单中剔除与扰动项相关的变量(这一步较难)。

(ii)的操作有一定难度，因为扰动项不可观测。

如何判断某候选变量(z)是否与不可观测的扰动项(ε)相关呢？

由于扰动项是被解释变量(y)的扰动项，故可从该候选变量与被解释变量的相关性着手。

z与y相关，因为z与内生解释变量x相关。

重要的是，z对y的影响仅仅通过x来起作用，因为如果z与ε相关，则z对y的影响必然还有除x以外的渠道，参见图 10.3。

至于是否“z对y的影响仅仅通过x来起作用”，有时可以通过定性的讨论来确定。

这就是上文提到的**排他性约束**(exclusion restriction)。

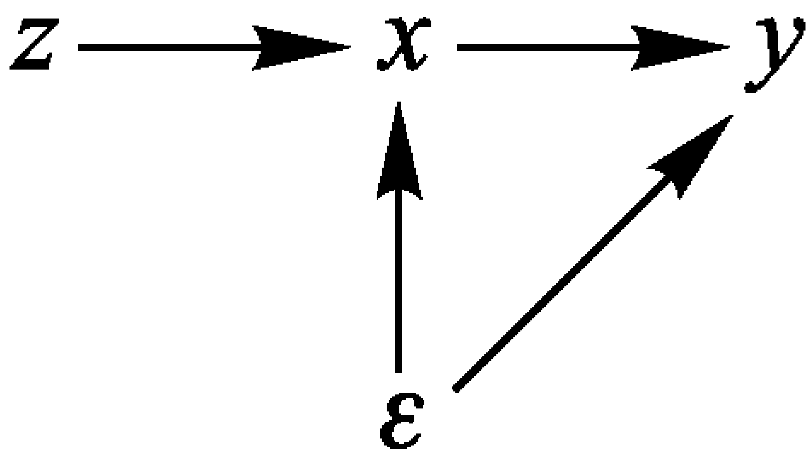

图 10.3　工具变量示意图

例　滞后变量。对于时间序列或面板数据，常使用内生解释变量的滞后作为工具变量。显然，内生解释变量与其滞后变量相关。而由于滞后变量已经发生，故为“前定”(从当期的角度看，其取值已经固定)，可能与当期的扰动项不相关。

比如，Groves, Hong, McMillan and Naughton (1994)考察国企改革(员工奖金激励制度)对企业生产率的作用。

奖金占员工报酬比重越高，则越能促进生产率的提高。但生产率越高的企业越有能力给员工发奖金，故存在双向因果关系。

该文使用奖金比重的滞后值作为当期奖金比重的工具变量。

二者的相关性是显然的。

另一方面，当期的生产率不可能影响过去的奖金比重，故奖金比重的滞后值可能具有外生性。

例　警察人数与犯罪率。

警察人数越多，则犯罪率应越低。但警察人数是内生变量，因为存在逆向因果。

Levitt (1997)使用“市长选举的政治周期”作为警察人数的工具变量。

在任市长在竞选连任时，为了拉选票，会增加警察人数以保证治安，故满足相关性。

选举周期一般以机械的方式确定，除了对警察人数有影响外，不会单独地对犯罪率起作用，故满足外生性。

例 制度对经济增长的影响。好的制度能促进经济增长，但制度变迁常常也依赖于经济增长，故制度是内生变量。

Acemoglu, Johnson and Robinson (2001)使用“移民死亡率”(settler mortality)作为制度的工具变量。

在移民死亡率高的地方(比如非洲)，移民难以定居，故在当地建立掠夺性制度(extractive institutions)。在移民死亡率低的地方(比如北美)，则建立有利于经济增长的制度(比如较好的产权保护)。

这种初始制度上的差异一直延续到今天，故移民死亡率与今天的制度相关，满足相关性。

另一方面，移民死亡率除了对制度有影响外，不再对当前的经济增长有任何直接影响，故满足外生性。

例 看电视过多引发小儿自闭症？Waldman, Nicholson and Adilov (2006)与 Waldman, Nicholson, Adilov and Williams (2008)研究过多观看电视是否引发小儿自闭症。

有自闭倾向的儿童可能更经常看电视，而不喜户外活动或与人交往；故存在双向因果关系。

作者使用降雨量作为电视观看时间的工具变量。二者存在相关性，即降雨越多的地区，人们呆在室内的时间越长，故看电视时间也越长。

降雨量很可能是外生的(只通过看电视时间而影响被解释变量)。研究结果支持过多观看电视为小儿自闭症的诱因。

例 科举制对人力资本积累的长期影响。Chen, Kung and Ma (2020)以 278 个“府”(prefecture)为横截面的观测单位，考察明清时期各府的进士密度(进士数量除以府人口)对 2010 年教育年限的影响。

尽管已控制不少协变量，但仍可能有遗漏变量偏差。由于科举成功与教辅资料密切相关，而 278 个府仅有 19 个印刷中心，其分布依赖于木版印刷的两种主要材料，即松(用于制墨)与竹(用于造纸)。

该文使用“各府离最近的松与竹产地的水运距离”(a prefecture's shortest river distance to its nearest sites of pine and bamboo)作为工具变量进行 2SLS 估计。

结果发现，如果明清时期的进士数量翻倍，则 2010 年教育年限可增加 8.5%。考虑到各府的进士密度差异较大，此效应较大。

10.9 工具变量法的 Stata 实例

以数据集 grilic.dta 为例演示工具变量法，继续对教育投资回报率的探讨。

此数据集的主要变量包括：*lnw*(工资对数)，*s*(教育年限)，*expr*(工龄)，*tenure*(在现单位的工作年数)，*iq*(智商)，*med*(母亲的教育年限)，*kww*(在“knowledge of the World of Work”测试中的成绩)，*rns*(美国南方虚拟变量，住在南方=1)，*smsa*(大城市虚拟变量，住在大城市=1)。

(1) 作为参照系，首先进行 OLS 回归，并使用稳健标准误。

```
. use grilic.dta,clear
. reg lnw s expr tenure rns smsa,r
```

我们感兴趣的关键解释变量为 *s*(教育年限)，而 *expr*, *tenure*, *rns*, *smsa* 为控制变量。

```
Linear regression                               Number of obs     =        758
                                                F(5, 752)         =      84.05
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3521
                                                Root MSE          =     .34641

------------------------------------------------------------------------------
             |               Robust
         lnw | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           s |    .102643   .0062099    16.53   0.000     .0904523    .1148338
        expr |   .0381189   .0066144     5.76   0.000      .025134    .0511038
      tenure |   .0356146   .0079988     4.45   0.000     .0199118    .0513173
         rns |  -.0840797    .029533    -2.85   0.005    -.1420566   -.0261029
        smsa |   .1396666    .028056     4.98   0.000     .0845893     .194744
       _cons |   4.103675   .0876665    46.81   0.000     3.931575    4.275775
------------------------------------------------------------------------------
```

教育投资的年回报率高达 10.26%，而且在 1%的水平上显著不为 0。

这个教育投资回报率似乎太高了。

可能的原因是，由于遗漏变量“能力”与教育年限正相关，故“能力”对工资的贡献也被纳入教育的贡献，高估了教育的回报率。

(2) 引入智商(*iq*)作为“能力”的代理变量(proxy)，再进行 OLS 回归。

```
. reg lnw s iq expr tenure rns smsa,r
```

Linear regression						
				Number of obs =		758
				F(6, 751) =		71.89
				Prob > F =		0.0000
				R-squared =		0.3600
				Root MSE =		.34454

lnw	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
s	.0927874	.0069763	13.30	0.000	.0790921	.1064826
iq	.0032792	.0011321	2.90	0.004	.0010567	.0055016
expr	.0393443	.0066603	5.91	0.000	.0262692	.0524193
tenure	.034209	.0078957	4.33	0.000	.0187088	.0497092
rns	-.0745325	.0299772	-2.49	0.013	-.1333815	-.0156834
smsa	.1367369	.0277712	4.92	0.000	.0822186	.1912553
_cons	3.895172	.1159286	33.60	0.000	3.667589	4.122754

加入“能力”的代理变量 *iq* 后，教育投资的回报率下降为 9.28%，更为合理些，但仍然显得过高。

(3) 由于用 *iq* 来度量能力存在“测量误差”，故 *iq* 是内生变量。

考虑使用变量(*med*, *kww*)作为 *iq* 的工具变量。

母亲的教育年限(*med*)与 KWW 测试成绩(*kww*)都与 *iq* 正相关；并假设 *med* 与 *kww* 为外生。

进行 2SLS 回归，使用稳健标准误，并显示第一阶段的回归结果。

```
. ivregress 2sls lnw s expr tenure rns smsa
(iq=med kww),r first
```

First-stage regressions

Number of obs	= 758
F(7, 750)	= 47.74
Prob > F	= 0.0000
R-squared	= 0.3066
Adj R-squared	= 0.3001
Root MSE	= 11.3931

iq	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
s	2.467021	.2327755	10.60	0.000	2.010052	2.92399
expr	-.4501353	.2391647	-1.88	0.060	-.9196471	.0193766
tenure	.2059531	.269562	0.76	0.445	-.3232327	.7351388
rns	-2.689831	.8921335	-3.02	0.003	-4.441207	-.938455
smsa	.2627416	.9465309	0.28	0.781	-1.595424	2.120907
med	.3470133	.1681356	2.06	0.039	.0169409	.6770857
kww	.3081811	.0646794	4.76	0.000	.1812068	.4351553
_cons	56.67122	3.076955	18.42	0.000	50.63075	62.71169

```
Instrumental variables 2SLS regression            Number of obs   =        758
                                                  Wald chi2(6)    =     370.04
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.2775
                                                  Root MSE        =     .36436
```

lnw	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
iq	.0139284	.0060393	2.31	0.021	.0020916	.0257653
s	.0607803	.0189505	3.21	0.001	.023638	.0979227
expr	.0433237	.0074118	5.85	0.000	.0287968	.0578505
tenure	.0296442	.008317	3.56	0.000	.0133432	.0459452
rns	-.0435271	.0344779	-1.26	0.207	-.1111026	.0240483
smsa	.1272224	.0297414	4.28	0.000	.0689303	.1855146
_cons	3.218043	.3983683	8.08	0.000	2.437256	3.998831

```
Endogenous: iq
Exogenous:  s expr tenure rns smsa med kww
```

教育投资回报率降为 6.08%，且在 1%水平上显著；比较合理。

(4)下面进行过度识别检验：

```
. estat overid
```

```
  Test of overidentifying restrictions:

  Score chi2(1)          =  .151451  (p = 0.6972)
```

由于 p 值为 0.697，故接受原假设，认为(*med*, *kww*)外生，与扰动项不相关。

(5) 进一步考察工具变量与内生变量的相关性

从第一阶段的回归结果可知，工具变量(*med*, *kww*)对内生变量 *iq* 均有较好的解释力， p 值都小于 0.05。

正式检验须计算第一阶段回归的普通(非稳健)F统计量。

使用普通标准误重新进行 2SLS 估计。

```
. quietly ivregress 2sls lnw s expr tenure rns smsa (iq=med kww)

. estat firststage
```

```
First-stage regression summary statistics
--------------------------------------------------------------------------
             |            Adjusted      Partial
    Variable |   R-sq.       R-sq.        R-sq.       F(2,750)   Prob > F
-------------+------------------------------------------------------------
          iq |  0.3066      0.3001       0.0382        14.9058    0.0000
--------------------------------------------------------------------------


Minimum eigenvalue statistic = 14.9058

Critical Values                      # of endogenous regressors:    1
H0: Instruments are weak             # of excluded instruments:     2
---------------------------------------------------------------------
                                   |    5%     10%     20%     30%
2SLS relative bias                 |         (not available)
-----------------------------------+---------------------------------
                                   |   10%     15%     20%     25%
2SLS size of nominal 5% Wald test  |  19.93   11.59    8.75    7.25
LIML size of nominal 5% Wald test  |   8.68    5.33    4.42    3.92
---------------------------------------------------------------------
```

第一阶段回归的两个工具变量系数联合显著性的F统计量为14.91，超过10，故认为不存在弱工具变量。

(6) 为了稳健起见(也为了示范)，使用对弱工具变量更不敏感的有限信息最大似然法(LIML)：

```
. ivregress liml lnw s expr tenure rns smsa
(iq=med kww),r
```

```
Instrumental variables LIML regression                Number of obs   =        758
                                                      Wald chi2(6)    =     369.62
                                                      Prob > chi2     =     0.0000
                                                      R-squared       =     0.2768
                                                      Root MSE        =     .36454

------------------------------------------------------------------------------
             |               Robust
         lnw | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
          iq |   .0139764   .0060681     2.30   0.021     .0020831    .0258697
           s |   .0606362    .019034     3.19   0.001     .0233303    .0979421
        expr |   .0433416   .0074185     5.84   0.000     .0288016    .0578816
      tenure |   .0296237    .008323     3.56   0.000     .0133109    .0459364
         rns |  -.0433875    .034529    -1.26   0.209    -.1110631    .0242881
        smsa |   .1271796   .0297599     4.27   0.000     .0688512     .185508
       _cons |   3.214994   .4001492     8.03   0.000     2.430716    3.999272
------------------------------------------------------------------------------
Endogenous: iq
Exogenous:  s expr tenure rns smsa med kww
```

LIML 的系数估计值与 2SLS 非常接近，从侧面印证了“不存在弱工具变量”。

(7) 使用工具变量法的前提是存在内生解释变量。下面进行豪斯曼检验，其原假设为“所有解释变量均为外生”。

```
. qui reg lnw iq s expr tenure rns smsa

. estimates store ols

. qui ivregress 2sls lnw s expr tenure rns smsa (iq=med kww)

. estimates store iv

. hausman iv ols,constant sigmamore
```

由于传统的豪斯曼检验建立在同方差的前提下，故未使用稳健标准误(没有用选择项“r”）。

```
Note: the rank of the differenced variance matrix (1) does not equal the number of
        coefficients being tested (7); be sure this is what you expect, or there may be
        problems computing the test.  Examine the output of your estimators for anything
        unexpected and possibly consider scaling your variables so that the coefficients
        are on a similar scale.

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       iv          ols         Difference          S.E.
-------------+----------------------------------------------------------------
          iq |    .0139284     .0032792        .0106493        .0054318
           s |    .0607803     .0927874        -.032007        .0163254
        expr |    .0433237     .0393443        .0039794        .0020297
      tenure |    .0296442      .034209       -.0045648        .0023283
         rns |   -.0435271    -.0745325        .0310054        .0158145
        smsa |    .1272224     .1367369       -.0095145        .0048529
       _cons |    3.218043     3.895172       -.6771285        .3453751
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from ivregress
            B = inconsistent under Ha, efficient under Ho; obtained from regress

    Test:  Ho:  difference in coefficients not systematic

                  chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        3.84
                Prob>chi2 =      0.0499
                (V_b-V_B is not positive definite)
```

p 值(`Prob>chi2`)为 0.0499，故可在 5%的显著性水平上拒绝“所有解释变量均为外生”的原假设，认为 iq 为内生变量。

由于传统的豪斯曼检验在异方差的情形下不成立，下面进行异方差稳健的 DWH 检验：

```
. estat endogenous
```

```
  Tests of endogeneity
  H0: Variables are exogenous

  Durbin (score) chi2(1)          =  3.87962  (p = 0.0489)
  Wu-Hausman F(1,750)             =  3.85842  (p = 0.0499)
```

上表提供了一个 F 统计量与一个 χ^2 统计量，二者在大样本下渐近等价。二者的 p 值都小于 0.05，故认为 iq 为内生解释变量。

(8) 汇报结果：将以上各种估计法的系数及标准误列在同一表格中，可使用以下命令：

```
. qui reg lnw s expr tenure rns smsa,r
. est sto ols_no_iq
. qui reg lnw iq s expr tenure rns smsa,r
. est sto ols_with_iq
. qui ivregress 2sls lnw s expr tenure rns smsa (iq=med kww),r
. est sto tsls
. qui ivregress liml lnw s expr tenure rns smsa (iq=med kww),r
. est sto liml
. estimates table ols_no_iq ols_with_iq tsls liml,b se
```

选择项“b”表示显示回归系数，选择项“se”表示标准误。

Variable	ols_no_iq	ols_with~q	tsls	liml
s	.10264304	.09278735	.06078035	.06063623
	.00620988	.00697626	.01895051	.01903397
expr	.0381189	.03934425	.04332367	.04334159
	.00661439	.00666033	.00741179	.0074185
tenure	.03561456	.03420896	.02964421	.02962365
	.00799884	.00789567	.00831697	.00832297
rns	-.08407974	-.07453249	-.04352713	-.04338751
	.02953295	.02997719	.03447789	.03452902
smsa	.13966664	.13673691	.12722244	.1271796
	.02805598	.02777116	.02974144	.02975994
iq		.00327916	.01392844	.01397639
		.00113212	.00603931	.00606812
_cons	4.103675	3.8951718	3.2180433	3.2149943
	.08766646	.11592863	.39836829	.40014925

Legend: b/se

如果希望用一颗星表示 10%的显著性水平，两颗星表示 5%的显著性水平，而三颗星表示 1%的显著性水平，可使用如下命令：

```
. estimates table ols_no_iq ols_with_iq tsls liml,star(0.1 0.05 0.01)
```

Variable	ols_no_iq	ols_with_iq	tsls	liml
s	.10264304***	.09278735***	.06078035***	.06063623***
expr	.0381189***	.03934425***	.04332367***	.04334159***
tenure	.03561456***	.03420896***	.02964421***	.02962365***
rns	-.08407974***	-.07453249**	-.04352713	-.04338751
smsa	.13966664***	.13673691***	.12722244***	.1271796***
iq		.00327916***	.01392844**	.01397639**
_cons	4.103675***	3.8951718***	3.2180433***	3.2149943***

Legend: * p<.1; ** p<.05; *** p<.01

但 Stata 官方命令“`estimates table`”无法同时显示回归系数、标准误与表示显著性的星号。

下载非官方命令“`estout`”。

```
. ssc install estout
. esttab ols_no_iq ols_with_iq tsls liml,se r2
mtitle star(* 0.1 ** 0.05 *** 0.01)
```

选择项“se”表示在括弧中显示标准误(默认显示t统计量，如果使用选择项“p”则显示p值)。

选择项“r2”表示显示R^2，选择项“mtitle”表示使用模型名称(model title)作为表中每列的标题(默认以被解释变量为标题)。

选择项“star(* 0.1 ** 0.05 *** 0.01)”表示以星号表示显著性水平。更多说明，参见“help estout”。

	(1) ols_no_iq	(2) ols_with_iq	(3) tsls	(4) liml
s	0.103***	0.0928***	0.0608***	0.0606***
	(0.00621)	(0.00698)	(0.0190)	(0.0190)
expr	0.0381***	0.0393***	0.0433***	0.0433***
	(0.00661)	(0.00666)	(0.00741)	(0.00742)
tenure	0.0356***	0.0342***	0.0296***	0.0296***
	(0.00800)	(0.00790)	(0.00832)	(0.00832)
rns	-0.0841***	-0.0745**	-0.0435	-0.0434
	(0.0295)	(0.0300)	(0.0345)	(0.0345)
smsa	0.140***	0.137***	0.127***	0.127***
	(0.0281)	(0.0278)	(0.0297)	(0.0298)
iq		0.00328***	0.0139**	0.0140**
		(0.00113)	(0.00604)	(0.00607)
_cons	4.104***	3.895***	3.218***	3.215***
	(0.0877)	(0.116)	(0.398)	(0.400)
N	758	758	758	758
R-sq	0.352	0.360	0.278	0.277

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

如果要将上表输出到 Microsoft Word 文档，并以文件名 iv 来命名此文档，则可运行如下命令：

```
. esttab ols_no_iq ols_with_iq tsls liml using
iv.rtf,se r2 mtitle star(* 0.1 ** 0.05 *** 0.01)
replace
(file iv.rtf not found)
(output written to iv.rtf)
```

其中，“`iv.rtf`”的扩展名“`rtf`”表示“rich text format”。选择项“`replace`”表示，若已有同名的文件，可覆盖之。

点击输出结果中的“iv.rtf”链接，即可打开此文件，然后可在 Word 中继续编辑此文件。