计量经济学(Econometrics)

胡华平

西北农林科技大学

经济管理学院数量经济教研室

huhuaping01@hotmail.com

2021-09-07

第11章：内生自变量问题

11.1 简单回归模型中的遗漏变量

一个简单示例

假定工资水平的“**真实模型**”为：

$$\log(wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 abil_i + u_i$$

然而，因为能力变量（$abil_i$）无法观测得到，所以我们经常用智商水平变量（IQ_i）来替换，并构建如下“**代理变量模型**”：

$$\log(wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 IQ_i + u_i^*$$

一个简单示例

进一步地，如果拿不到智力水平（IQ_i）数据，我们很可能构建出如下**遗漏重要变量的“偏误模型”**：

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

- 此时，我们可以认为重要变量 $abil_i$ 被遗漏，从而进入到随机干扰项 v_i 中。
- 又因为，一般情况下我们认为教育水平（$educ_i$）与能力（$abil_i$）是相关的，从而可以认为 自变量 $educ_i$ 与随机干扰项 v_i 是相关的。而这是违背CLRM假设的（违背了哪一条？）。

知识回顾

对于总体回归模型：

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad \text{(PRM)}$$

- 在CLRM假设下：**CLRM假设3**——X是固定的（给定的）或独立于误差项。也即自变量X**不是**随机变量。此时，我们可以使用OLS方法，并得到**BLUE**。

$$\begin{aligned} Cov(X_i, u_i) &= 0 \\ E(X_i u_i) &= 0 \end{aligned}$$

- 如果违背上述假设，也即自变量X与随机干扰项相关。此时使用OLS估计将不再能得到**BLUE**，而应该采用**工具变量法**（IV）进行估计。

$$\begin{aligned} Cov(X_i, u_i) &= 0 \\ E(X_i u_i) &= 0 \end{aligned}$$

事实上，无论 X_i与 u_i是否相关，我们都可以采用**IV法**得到**BLUE**。

代理变量

代理变量（proxy variable）：一般因为某些原因，某个变量 X_i 不能直接观测得到（数据不可得），那么常常会找一个能够观测到到的、并与 X_i 高度相关的变量 D_i 作为替代。

- 假定工资水平的"**真实模型**"为：

$$\log(wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 abil_i + u_i$$

- 然而，因为能力变量（$abil_i$）无法观测得到，所以我们经常用智商水平变量（IQ_i）来替换，并构建如下"**代理变量模型**"：

$$\log(wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 IQ_i + u_i^*$$

此时，智力水平 IQ_i 就可以认为是变量 能力 $abil_i$ 的**代理变量**。

工具变量

代理变量（proxy variable）：一般因为某些原因，某个变量 X_i不能直接观测得到（数据不可得），那么常常会找一个能够观测到到的、并与 X_i高度相关的变量 D_i作为替代。

工具变量（instrument variable）：一般记为 Z_i，是指一个具备如下**两个性质**的可观测的变量：

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad \text{(PRM)}$$

$$\begin{aligned} Cov(Z_i, u_i) &= 0 \qquad \text{(instrumental exogeneity)} \\ Cov(Z_i, X_i) &\neq 0 \qquad \text{(instrumental relevance)} \end{aligned}$$

- 工具外生性（instrumental exogeneity）：工具变量对于总体回归模型是外生的。
- 工具相关性（instrumental relevance）：工具变量对于解释 X_i的变异时有很重要的作用。

对示例的进一步扩展

对于“**真实模型**”：

$$\log(wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 abil_i + u_i$$

我们很可能构建出如下**遗漏重要变量**的“**偏误模型**”：

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

因此，对于教育 edu_i的工具变量 Z_i而言：

- 工具变量 Z_i必须与能力 $abil_i$不相关，但又必须与教育 edu_i相关。
- 选择方案1：家庭背景教育如母亲的教育 $MotherEdu_i$就可能是工具变量备选方案之一。因为我们通常可以认为**母亲教育水平**与孩子教育是正相关的。
- 选择方案2：另一个工具变量的备选方案是家庭中**兄弟姐妹的数量** $sibs$。因为我们通常可以认为家庭中兄弟姐妹数量会导致较低的平均教育水平（也即负相关）。

提问：哪一个方案更好呢？为什么？

学习成绩与逃课次数的例子

下面我们以学习成绩与逃课次数的例子进行分析讨论。

模型设定

假设“真实模型”是：

$$score_i = \alpha_1 + \alpha_2 skipped_i + \alpha_3 abil_i + \alpha_4 mot_i + \alpha_5 income_i + u_i$$

一个遗漏了重要变量的“偏误模型”是：

$$score_i = \beta_1 + \beta_2 skipped_i + v_i$$

- 学习成绩受到逃课次数的影响，但是我们也很担心以上模型中 $skipped_i$ 与 v_i 中的某些因素相关，例如越有能力 $abil_i$、越积极 mot_i 的学生，逃课也越少。
- 因为自变量 $skipped_i$ 可能与随机干扰项 v_i 相关。此时，对于以上简单的回归，可能得不出可靠的估计。

模型设定

$$score_i = \beta_1 + \beta_2 skipped_i + v_i$$

逃课次数 $skipped_i$的工具变量 Z_i有哪些可供备选的呢？

- 宿舍跟上课地点的距离 $distance$。我们一般认为，它与逃课次数相关 $skipped_i$，但是它与 v_i中的某些因素也会相关么？
- 如果收入水平 $income$确实影响了学习成绩，但是模型却没有引入收入水平 $income$变量，也就意味着 v_i中包含了遗漏的重要变量——收入水平 $income$。此时，距离 $distance$就会与收入水平 $income$相关，进而与 v_i相关。——因为收入少的学生，更倾向于在外租房（合租）；收入多的学生，更倾向于住校。

11.2 内生变量法下的估计问题

工具变量法下系数的估计过程

把上述"偏误模型"记为：

$$
\begin{aligned}
score_i &= \beta_1 + \beta_2 skipped_i + u_i \\
Y_i &= \beta_1 + \beta_2 X_i + u_i
\end{aligned}
$$

假设我们找到了理想的工具变量 Z_i，并构建如下的**工具变量模型**：

$$Y_i = \alpha_1 + \alpha_2 Z_i + v_i$$

$$
\begin{aligned}
cov(Z_i, Y_i) &= \alpha_2 cov(Z_i, X_i) + cov(Z_i, u_i) && \leftarrow [cov(Z_i, u_i) = 0] \\
\alpha_2|_{IV}^{plim} &= \frac{cov(Z_i, Y_i)}{cov(Z_i, X_i)} = \frac{\sum z_i y_i}{\sum z_i x_i} && \leftarrow [if \quad X_i = Z_i] \\
&= \frac{\sum x_i y_i}{\sum x_i^2} = \beta_2
\end{aligned}
$$

这将意味着工具变量法IV会得到最小二乘法OLS下的估计结果。

工具变量法下系数的真实方差

对于“偏误模型”和工具变量模型：

$$
\begin{aligned}
Y_i &= \beta_1 + \beta_2 X_i + u_i \qquad \text{(PRM)} \\
Y_i &= \alpha_1 + \alpha_2 Z_i + v_i \qquad \text{(IV)}
\end{aligned}
$$

如果如下三个条件成立：

$$
\begin{aligned}
Cov(Z_i, u_i) &= 0 \\
Cov(Z_i, X_i) &\neq 0 \\
E(v_i^2|Z_i) &\equiv \sigma^2 \equiv var(u_i)
\end{aligned}
$$

可证明斜率系数 α_2的渐近方差：

$$
var(\alpha_2) \simeq \frac{\sigma^2}{n\sigma^2_{X_i}\rho^2_{(X_i,Z_i)}}
$$

其中：

- σ^2是 v_i的总体方差，也即 $var(v_i) \equiv \sigma^2$。
- $\sigma^2_{X_i}$是 X_i的总体方差，也即 $var(X_i) \equiv \sigma^2_{X_i}$。
- $\rho^2_{(X_i,Z_i)}$是 X_i和 Z_i的总体相关系数的平方，也即 $\rho^2_{(X_i,Z_i)} \equiv \frac{[cov(X_i,Z_i)]^2}{var(X_i)var(Z_i)}$；

工具变量法下系数的样本方差

对于给定的样本数据，我们可以计算出

$$var(\alpha_2) \simeq \frac{\sigma^2}{n\sigma^2_{X_i}\rho^2_{(X_i,Z_i)}} \simeq \frac{\hat{\sigma}^2}{nS^2_{X_i}R^2_{(X_i,Z_i)}}$$

其中：

- $\sigma^2_{X_i} \simeq S^2_{X_i} = \frac{\sum(X_i-\bar{X})^2}{n-1}$。
- $\rho^2_{(X_i,Z_i)} \simeq R^2$，其中 R^2 为通过做 X_i 对 Z_i 的回归来获得的**判定系数**。

$$X_i = \hat{\pi}_1 + \hat{\pi}_2 Z_i + \epsilon_i$$

- $\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2}$，是来自对**工具变量回归**的残差计算。

$$Y_i = \hat{\alpha}_1 + \hat{\alpha}_2 Z_i + e_i$$

已婚女性的教育回报案例

下面我们对已婚女性的教育回报案例进行分析讨论。

研究者关注428名已婚女性**时均工资** *wage*与其**受教育年数** *educ*之间的关系，并考虑如下变量：

变量说明

vars	mark
wage	时均工资
educ	受教育年数
exper	就业次数
fatheduc	父亲的受教育年数
motheduc	母亲的受教育年数
inlf	是否是劳动力
hours	工作时长

Showing 1 to 7 of 22 entries

Previous 1 2 3 4 Next

原始数据

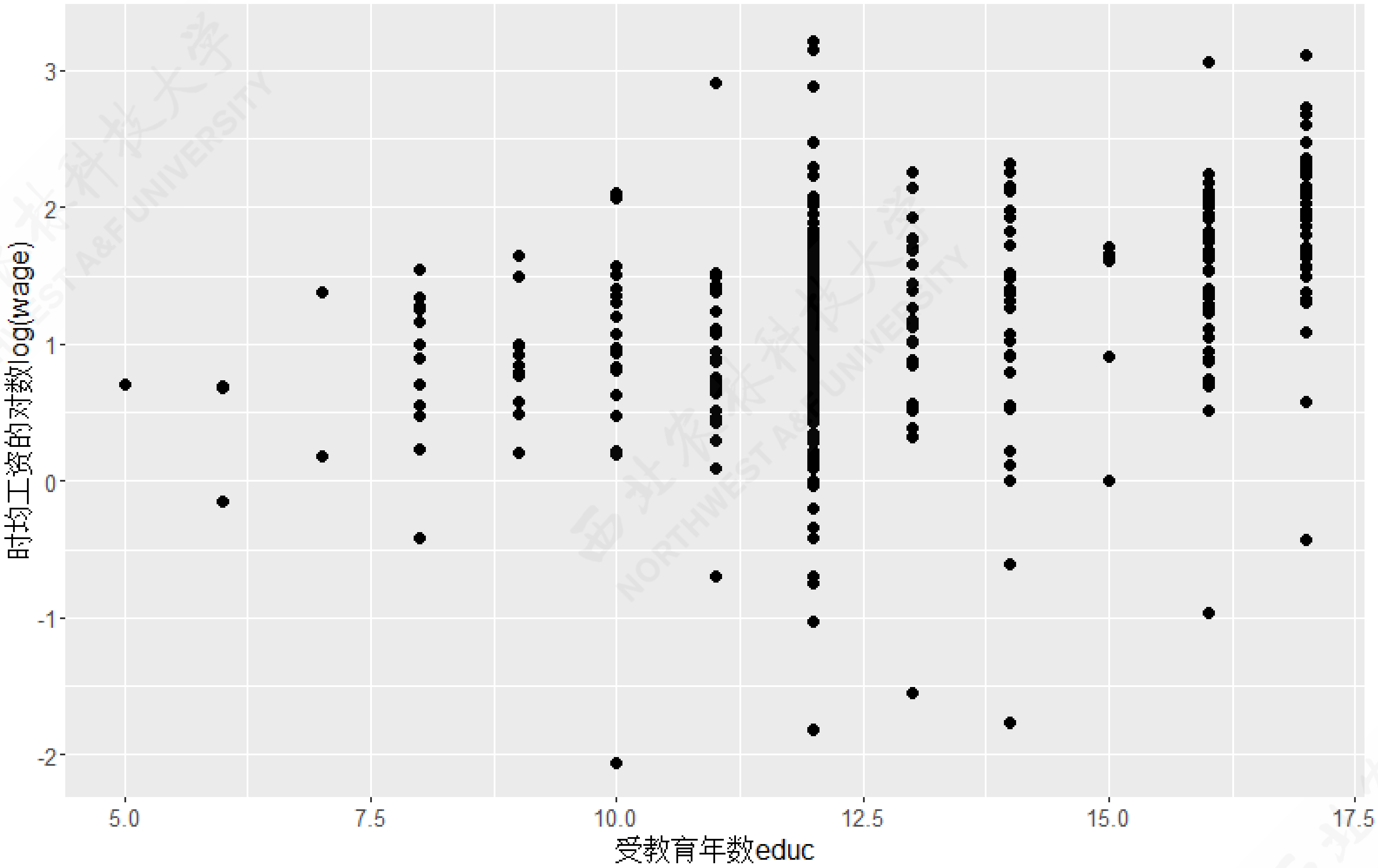
已婚女性的教育回报数据n=(428)

wage	educ	fatheduc
3.3540	12	7
1.3889	12	7
4.5455	12	7
1.0965	12	7
4.5918	14	14
4.7421	12	7
8.3333	16	7
7.8431	12	3

Showing 1 to 8 of 428 entries

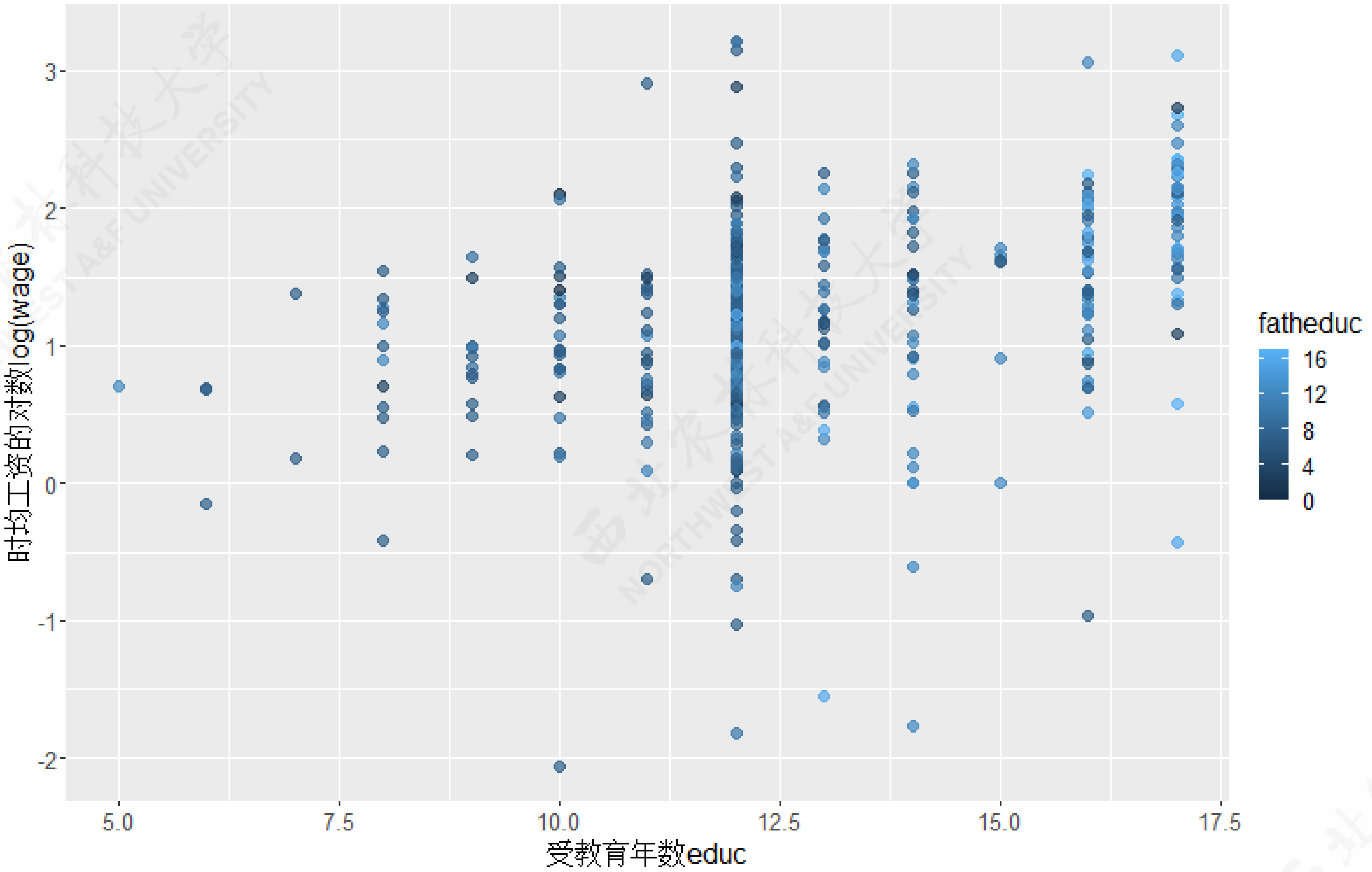
Previous 1 2 3 4 5 … 54 Next

散点图I



受教育年数与时均工资的散点图

散点图2



考虑父亲受教育年数的散点图

如果直接构建如下的“**偏误模型**”，并坚持采用OLS估计：

$$log(wage) = + \beta_1 + \beta_2 educ + u_i$$

$$\begin{array}{lll}
\widehat{log(wage)} = & -0.19 & +0.11educ \\
\text{(t)} & (-0.9998) & (7.5451) \\
\text{(se)} & (0.1852) & (0.0144) \\
\text{(fitness)} & R^2 = 0.1179; \bar{R}^2 = 0.1158 & \\
 & F^* = 56.93;\ p = 0.0000 &
\end{array}$$

工具变量法回归(IV):手工分步计算

采用工具变量法的第一阶段回归:

$$educ = +\beta_1 + \beta_2 fatheduc + u_i$$

$$\begin{array}{llll}
\widehat{educ} = & +10.24 & +0.27 fatheduc \\
\text{(t)} & (37.0993) & (9.4255) \\
\text{(se)} & (0.2759) & (0.0286) \\
\text{(fitness)} & R^2 = 0.1726; \bar{R}^2 = 0.1706 & \\
 & F^* = 88.84;\ p = 0.0000 &
\end{array}$$

采用工具变量法的第二阶段回归:

工具变量法回归（IV）：R软件自动计算

采用R包`AER`的工具变量回归函数`ivreg()`，可以得到如下回归结果：

```
Call:
ivreg(formula = log(wage) ~ educ | fath

Residuals:
    Min      1Q  Median      3Q     Max
-3.0870 -0.3393  0.0525  0.4042  2.0677

Coefficients:
            Estimate Std. Error t value
(Intercept)  0.44110    0.44610   0.989
educ         0.05917    0.03514   1.684
---
Signif. codes:  0 '***' 0.001 '**' 0.01

Residual standard error: 0.6894 on 426
Multiple R-Squared: 0.09344,    Adjuste
Wald test: 2.835 on 1 and 426 DF,  p-va
```

工具变量回归模型：

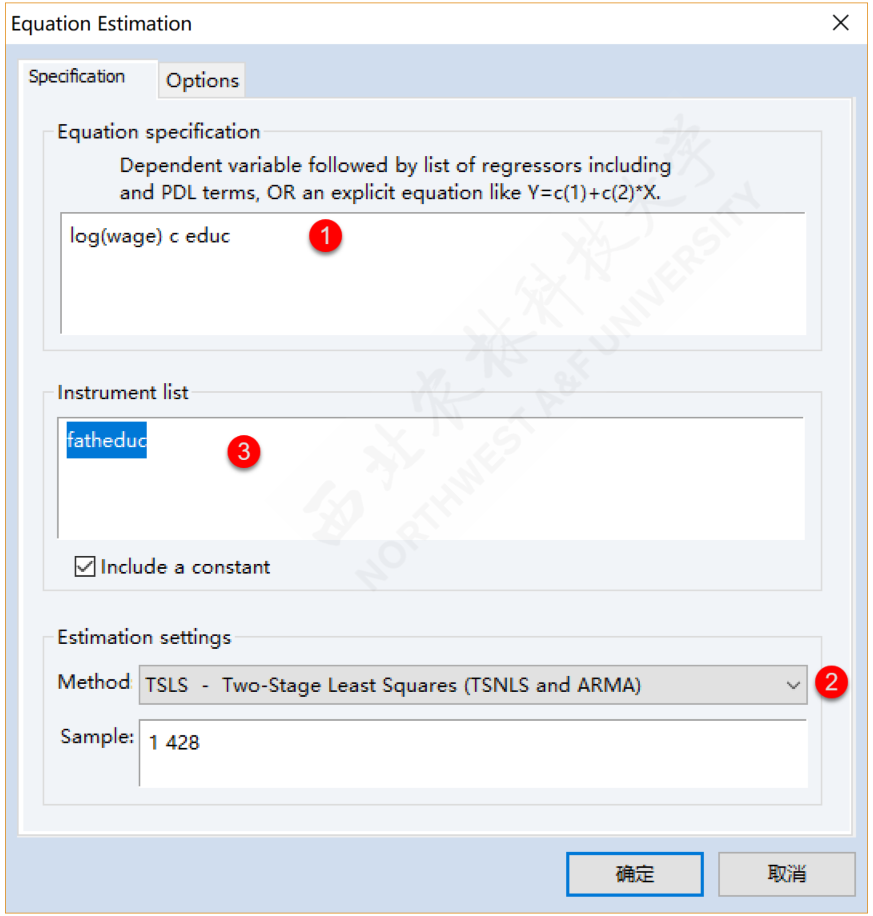
$$log(wage) = \lambda_1 + \lambda_2 educ | fatheduc + \epsilon_i$$

提问：

- 手工分步计算与软件自动计算有哪些不同？
- 判定系数和系数标准误差为什么会不同？

工具变量法回归（IV）：EViews软件自动计算

EViews软件下工具变量法的实现：



工具变量法回归（IV）：EViews软件自动计算

EViews软件下工具变量法的结果：

Equation: EQ_IV   Workfile: MROZ::wage\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(WAGE)
Method: Two-Stage Least Squares
Date: Time:
Sample: 1 428
Included observations: 428
Instrument specification: FATHEDUC
Constant added to instrument list

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.441103	0.446102	0.988795	0.3233
EDUC	0.059173	0.035142	1.683850	0.0929

R-squared	0.093438	Mean dependent var	1.190173
Adjusted R-squared	0.091310	S.D. dependent var	0.723198
S.E. of regression	0.689390	Sum squared resid	202.4601
F-statistic	2.835351	Durbin-Watson stat	1.968194
Prob(F-statistic)	0.092943	Second-Stage SSR	221.9799
J-statistic	1.51E-41	Instrument rank	2

11.3 多元回归模型的IV估计

新的符号表达体系

对于多元回归模型，我们可以记为：

$$
\begin{aligned}
\log(wage) &= \beta_1 + \beta_2 educ + \lambda_1 exper + u_i \\
Y_{1i} &= \beta_1 + \beta_2 Y_{2i} + \lambda_1 Z_{1i} + u_i \\
E(u_i) &= 0; \quad cov(Y_{2i}, u_i) \neq 0; \\
cov(Z_{1i}, u_i) &= 0; \quad cov(Y_{2i}, Z_{1i}) = 0
\end{aligned}
$$

- **内生变量**（endogenous variable）：用符号 Y_i表达。例如，因变量（工资水平）Y_{1i}显然是内生变量；而其中的一个自变量（教育年数）Y_{2i}我们在这里认为也是内生的——也即允许它跟随机干扰项 u_i**相关**。
- **外生变量**（exogenous variable）：用符号 Z_i表达。例如另一个自变量（工作经历）Z_{1i}则认为它是外生的——也即它跟随机干扰项 u_i**不相关**。

约简方程

$$\log(wage) = \beta_1 + \beta_2 educ + \lambda_1 exper + u_i$$
$$Y_{1i} = \beta_1 + \beta_2 Y_{2i} + \lambda_1 Z_{1i} + u_i$$

如果给内生变量 Y_{2i} 找到一个理想工具变量 Z_{2i}，则可构建如下**约简方程**：

$$\begin{aligned} Y_{2i} &= \pi_0 + \pi_1 Z_{1i} + \pi_2 Z_{2i} + v_i \\ E(v_i) &= 0; \quad cov(Y_{2i}, Z_{1i}) = 0; \\ cov(Z_{1i}, v_i) &= 0; \quad cov(Z_{2i}, v_i) = 0 \end{aligned}$$

- **约简方程**（reductive equation）：是指一个**内生变量**对全部**外生变量**的回归方程。
- 因为 $Z_{1i}; Z_{2i}$ 为外生变量都为**外生变量**，所以满足CLRM假设，上述**偏相关分析模型**可以直接使用OLS方法而得到**BLUE**。**偏相关分析模型**可以用于检验 Y_{2i} 与 Z_{2i} 是否相关，也即检验：$H_0: \pi_2 = 0; \quad H_1: \pi_2 \neq 0$。具体可以采用通常的**t检验**方法。

提问1：我们能不能检验 Z_{2i} 与 u_i 相关？能不能检验 Z_{1i} 与 u_i 相关？

提问2：还能不能构造别的**简约方程**？

两阶段最小二乘法（2SLS）

$$
\begin{aligned}
\log(wage) &= \beta_1 + \beta_2 educ + \lambda_1 exper + u_i \\
Y_{1i} &= \beta_1 + \beta_2 Y_{2i} + \lambda_1 Z_{1i} + u_i
\end{aligned}
$$

第一阶段OLS回归：对约简方程进行回归。

$$
\begin{aligned}
Y_{2i} &= \pi_0 + \pi_1 Z_{1i} + \pi_2 Z_{2i} + v_i \qquad \leftarrow [\pi_2 \neq 0] \\
\hat{Y}_{2i} &= \pi_0 + \pi_1 Z_{1i} + \pi_2 Z_{2i}
\end{aligned}
$$

第二阶段OLS回归：基于估计值进行回归。

$$
Y_{1i} = \alpha_0 + \alpha_1 Z_{1i} + \beta_2 \hat{Y}_{2i} + \epsilon_i
$$

已婚女性的教育回报案例

我们继续对已婚女性的教育回报案例进行分析讨论。

原始数据

已婚女性的教育回报数据n=(428)

wage	educ	exper	fatheduc	motheduc
3.3540	12	14	7	12
1.3889	12	5	7	7
4.5455	12	15	7	12
1.0965	12	6	7	7
4.5918	14	7	14	12
4.7421	12	33	7	14
8.3333	16	11	7	14
7.8431	12	35	3	3

Showing 1 to 8 of 428 entries

Previous 1 2 3 4 5 … 54 Next

偏误OLS回归：精炼报告

如果认为工作经历 $exper$是**外生变量**变量；且认为已婚女性的受教育年数 $educ$为**内生变量**。直接构建如下的“**偏误模型**”，并坚持采用OLS方法，估计结果为：

$$log(wage) = +\beta_1 + \beta_2 educ + \beta_3 exper + \beta_4 I(exper^2) + u_i$$

$$\begin{array}{lllll}
\widehat{log(wage)} = & -0.52 & +0.11educ & +0.04exper & -0.00I(exper^2) \\
\text{(t)} & (-2.6282) & (7.5983) & (3.1549) & (-2.0628) \\
\text{(se)} & (0.1986) & (0.0141) & (0.0132) & (0.0004) \\
\text{(fitness)} & R^2 = 0.1568; \bar{R}^2 = 0.1509 & & & \\
 & F^* = 26.29;\ p = 0.0000 & & &
\end{array}$$

偏误OLS回归：EViews报告

下面给出EViews软件分析报告：

Equation: EQ_OLS Workfile: MROZ::wage2\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(WAGE)
Method: Least Squares
Date: Time:
Sample: 1 428
Included observations: 428

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.522041	0.198632	-2.628179	0.0089
EDUC	0.107490	0.014146	7.598332	0.0000
EXPER	0.041567	0.013175	3.154906	0.0017
EXPER^2	-0.000811	0.000393	-2.062834	0.0397

R-squared	0.156820	Mean dependent var	1.190173
Adjusted R-squared	0.150854	S.D. dependent var	0.723198
S.E. of regression	0.666420	Akaike info criterion	2.035509
Sum squared resid	188.3051	Schwarz criterion	2.073445
Log likelihood	-431.5990	Hannan-Quinn criter.	2.050492
F-statistic	26.28615	Durbin-Watson stat	1.960988
Prob(F-statistic)	0.000000		

两阶段回归法（2SLS）：第一阶段

假设父亲受教育年数 $fatheduc$和母亲受教育年数 $motheduc$都是已婚女性受教育年数 $educ$的理想工具变量。

采用工具变量法的对如下约简方程：

$$educ = +\pi_1 + \pi_2 exper + \pi_3 exper^2 + \pi_4 fatheduc + \pi_5 motheduc + v_i$$

对以上约简方程进行第一阶段OLS回归，估计结果为：

$$\begin{array}{llllll}
\widehat{educ} = & +9.10 & +0.05exper & -0.00I(exper^2) & +0.19fatheduc & +0.16motheduc \\
(\text{t}) & (21.3396) & (1.1236) & (-0.8386) & (5.6152) & (4.3906) \\
(\text{se}) & (0.4266) & (0.0403) & (0.0012) & (0.0338) & (0.0359) \\
(\text{fitness}) & R^2 = 0.2115; \bar{R}^2 = 0.2040 & & & & \\
 & F^* = 28.36;\ p = 0.0000 & & & &
\end{array}$$

两阶段回归法（2SLS）：第二阶段

利用前述第一阶段回归得到的 $\widehat{educ}$ 以及原来的**外生变量** $exper$，我们可以构建如下的第二阶段回归模型：

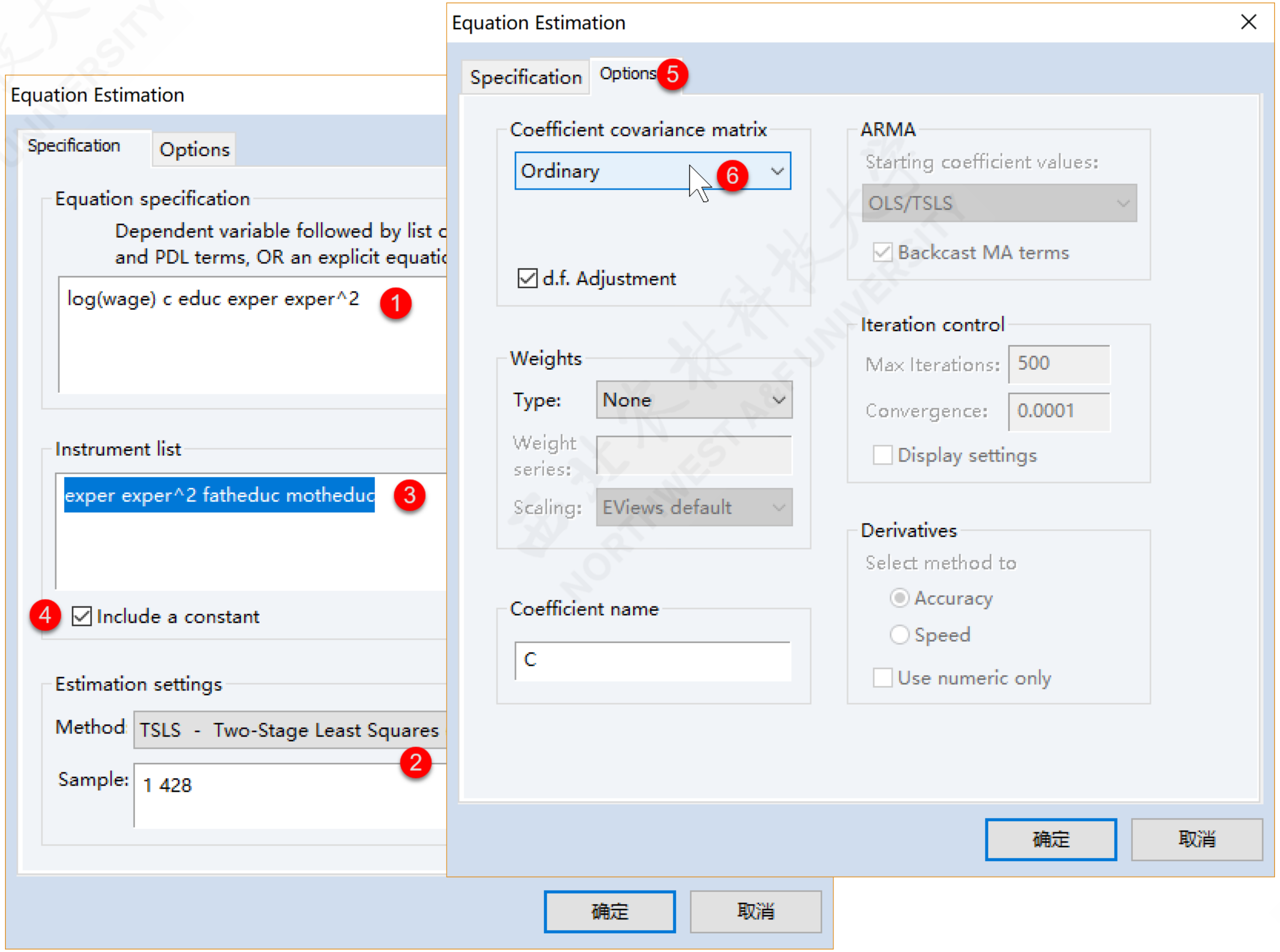
$$log(wage) = \alpha_1 + \alpha_2 exper_i + \alpha_3 exper_i^2 + \beta_1 \widehat{educ}_i + \epsilon_i$$

采用OLS方法对以上模型进行估计，得到如下结果：

$$\begin{array}{lllll}
\widehat{log(wage)} = +0.05 & & +0.06 educ.hat & +0.04 exper & -0.00 I(exper^2) \\
(\text{t}) & (0.1146) & (1.8626) & (3.1361) & (-2.1344) \\
(\text{se}) & (0.4198) & (0.0330) & (0.0141) & (0.0004) \\
(\text{fitness}) & R^2 = 0.0498; \bar{R}^2 = 0.0431 & & & \\
 & F^* = 7.40; \quad p = 0.0001 & & &
\end{array}$$

两阶段最小二乘法（2SLS,不调整方差）：EViews实现

EViews软件的具体设置为：



EViews软件的分析结果：

Equation: EQ_2SLS Workfile: MROZ::wage2\

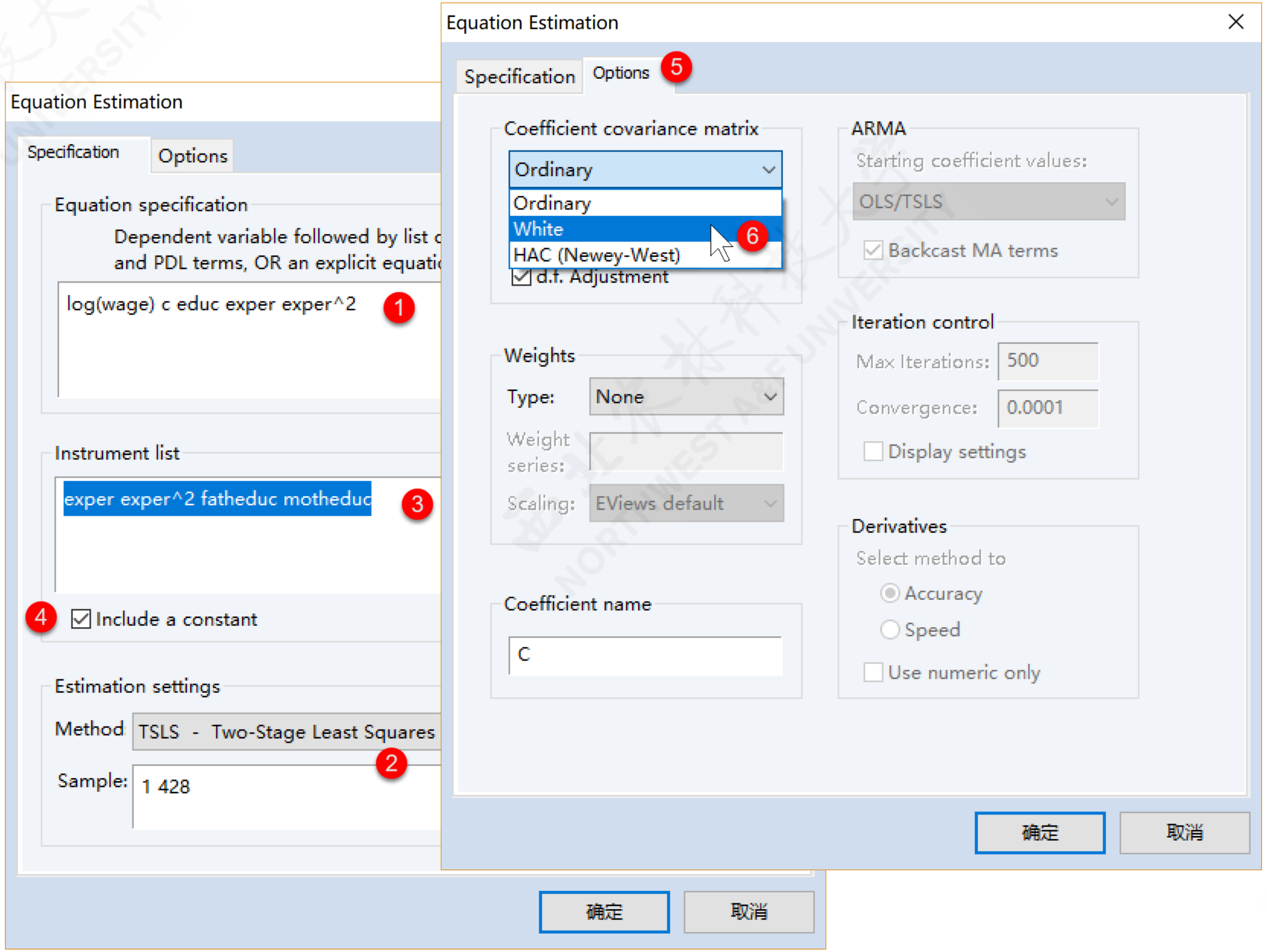
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(WAGE)
Method: Two-Stage Least Squares
Date: Time:
Sample: 1 428
Included observations: 428
Instrument specification: EXPER EXPER^2 FATHEDUC MOTHEDUC
Constant added to instrument list

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.048100	0.400328	0.120152	0.9044
EDUC	0.061397	0.031437	1.953024	0.0515
EXPER	0.044170	0.013432	3.288329	0.0011
EXPER^2	-0.000899	0.000402	-2.237993	0.0257

R-squared	0.135708	Mean dependent var	1.190173
Adjusted R-squared	0.129593	S.D. dependent var	0.723198
S.E. of regression	0.674712	Sum squared resid	193.0200
F-statistic	8.140709	Durbin-Watson stat	1.945659
Prob(F-statistic)	0.000028	Second-Stage SSR	212.2096
J-statistic	0.374538	Instrument rank	5
Prob(J-statistic)	0.540541		

EViews软件设置：



Equation Estimation
Specification
Options
Equation specification
Dependent variable followed by list o
and PDL terms, OR an explicit equatio
log(wage) c educ exper exper^2
1
Instrument list
exper exper^2 fatheduc motheduc
3
4
Include a constant
Estimation settings
Method
TSLS - Two-Stage Least Squares
2
Sample:
1 428
确定
取消
Equation Estimation
Specification
Options
5
Coefficient covariance matrix
Ordinary
Ordinary
White
HAC (Newey-West)
6
d.f. Adjustment
Weights
Type:
None
Weight series:
Scaling:
EViews default
Coefficient name
C
ARMA
Starting coefficient values:
OLS/TSLS
Backcast MA terms
Iteration control
Max Iterations:
500
Convergence:
0.0001
Display settings
Derivatives
Select method to
Accuracy
Speed
Use numeric only
确定
取消

两阶段最小二乘法（2SLS,怀特矫正）：EViews结果

EViews软件分析结果：

Equation: EQ_2SLS_ADJ Workfile: MROZ::wage2\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(WAGE)
Method: Two-Stage Least Squares
Date: Time:
Sample: 1 428
Included observations: 428
White heteroskedasticity-consistent standard errors & covariance
Instrument specification: EXPER EXPER^2 FATHEDUC MOTHEDUC
Constant added to instrument list

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.048100	0.429798	0.111914	0.9109
EDUC	0.061397	0.033339	1.841609	0.0662
EXPER	0.044170	0.015546	2.841202	0.0047
EXPER^2	-0.000899	0.000430	-2.090220	0.0372

R-squared	0.135708	Mean dependent var	1.190173
Adjusted R-squared	0.129593	S.D. dependent var	0.723198
S.E. of regression	0.674712	Sum squared resid	193.0200
F-statistic	8.140709	Durbin-Watson stat	1.945659
Prob(F-statistic)	0.000028	Second-Stage SSR	212.2096
J-statistic	0.374538	Instrument rank	5
Prob(J-statistic)	0.540541		

11.4 工具变量法的一些讨论

工具变量法回归中的判定系数

1. 与OLS中的估况不同，由于IV的SSR实际上可能大于SST。所以IV估计中的 R^2 可能为负。尽管报告IV估计的 R^2 也不会有什么害处，但也不是很有用。
2. 当自变量 X_i 与随机干扰项 v_i **相关**时，因变量 Y_i 的方差分解成 $\lambda_2^2 var(X_i) + var(v_i)$，因此**判定系数** R^2 没有合理的解释。进一步地，工具变量回归下 R^2 也**不能**用于联合约束的**F检验**。
3. 如果目标是为了得到最大的 R^2，我们将总是使用**OLS**。如果采用工具变量法（IV），拟合优度 R^2 已经不是其考虑的方面了。
4. **两阶段最小二乘法**（2SLS）是GLS方法的一种，它是需要利用额外的信息（工具变量）。很多时候2SLS还需要考虑对**系数方差矩阵**的矫正——**怀特矫正**（White方法）或**HAC矫正**（Neway-West方法）。

低劣工具变量条件下IV的性质

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$Y_i = \alpha_1 + \alpha_2 Z_i + v_i$$

最小二乘法（OLS）估计下：

$$\hat{\beta}_2|_{OLS}^{plim} = \beta_2 + corr(u_i, X_i) \cdot \frac{\sigma_{u_i}}{\sigma_{X_i}}$$

工具变量法（IV）估计下：

$$\hat{\alpha}_2|_{IV}^{plim} = \alpha_2 + \frac{corr(Z_i, v_i)}{corr(Z_i, X_i)} \cdot \frac{\sigma_{v_i}}{\sigma_{X_i}}$$

- 如果 $corr(Z_i, X_i) = 0.2$，要使得IV比OLS具有更小的渐近偏误，$corr(Z_i, v_i)$ 必须小于 $corr(X_i, u_i)$的1/5.

11.5 解释变量的内生性检验

（豪斯曼检验）

内生性检验的内容

工具变量（instrument variable）具备如下**两个性质**：

$$
\begin{aligned}
Cov(Z_i, u_i) &= 0 && \text{(instrumental exogeneity)} \\
Cov(Z_i, X_i) &\neq 0 && \text{(instrumental relevance)}
\end{aligned}
$$

我们先来检验**第二个**条件：一个变量 X_i是否真的是**内生性**的？

> 因为 u_i不能直接观测得到，所以第一个条件的检验往往不是很直接。

豪斯曼检验：偏误模型和2SLS

给定如下的“偏误回归模型”：

$$Y_{1i} = \beta_1 + \beta_2 Y_{2i} + \lambda_1 Z_{1i} + \lambda_2 Z_{2i} + u_i \qquad \text{(stucture eq.)}$$

- 仅有1个疑似**内生变量** Y_{2i}；2个**外生变量** $Z_{1i}; Z_{2i}$。
- 假定已经找到了变量 Y_{2i} 的两个**理想工具变量** $Z_{3i}; Z_{4i}$。

我们可以利用**两阶段最小二乘法**（2SLS）方法进行估计：

$$\begin{aligned}
Y_{2i} &= \pi_0 + \pi_1 Z_{1i} + \pi_2 Z_{2i} + \pi_3 Z_{3i} + \pi_4 Z_{4i} + v_i && \text{(1st eq. / reduce eq.)}\\
Y_{1i} &= \alpha_1 + \alpha_2 \hat{Y}_{2i} + \gamma_1 Z_{1i} + \gamma_2 Z_{2i} + \epsilon_i && \text{(2st eq.)}
\end{aligned}$$

豪斯曼检验：基本思想

豪斯曼检验（Hausman Test）的基本思想是：

- 如果疑似**内生变量** Y_{2i} 与随机干扰项 u_i **不相关**。那么以上模型符合**CLRM假设**，就可以直接使用**OLS估计**。而且使用OLS估计和使用**2SLS估计**结果应该是**一致的**（why？）。
- 如果疑似**内生变量** Y_{2i} 与随机干扰项 u_i **相关**。那么使用OLS估计和使用**2SLS估计**结果应该是有**差异的**（why？）。
- 如果有差异，将表明 Y_{2i} 必定是**内生的**（其中 Z_{ki} $(k \in 1,2,3,4)$ 仍旧保持**外生的**）。而且，理论上可以证明 Z_{ki} $(k \in 1,2,3,4)$ 与 u_i 不相关的**充要条件**是 v_i 与 u_i 不相关，也即：

> 若 $cov(u_i, v_i) = 0$，则 $cov(Z_{ki}, u_i) = 0 \quad (k \in 1,2,3,4)$。

豪斯曼检验：分析模型

为了证实上述观点，**豪斯曼**（Hausman）提出构建如下的回归模型，并检验 $H_0:\delta=0; H_1:\delta\neq 0$。

$$u_i=\delta v_i+\varepsilon_i \qquad \leftarrow [cov(v_i,\varepsilon_i)=0; E(\varepsilon_i)=0]$$

$$\begin{aligned}Y_{1i}&=\beta_1+\beta_2 Y_{2i}+\lambda_1 Z_{1i}+\lambda_2 Z_{2i}+\delta v_i+\epsilon_i\\ &\leftarrow [v_i=Y_{2i}-\pi_0-\pi_1 Z_{1i}-\pi_2 Z_{2i}+\pi_3 Z_{3i}-\pi_4 Z_{4i}=Y_{2i}-E(Y_{2i})]\end{aligned}$$

因为 v_i不能观测得到，实际上使用的是**约简方程**的残差 $\hat{v}_i$，最终估计如下的**豪斯曼检验方程**：

$$\begin{aligned}Y_{1i}&=\beta_1+\beta_2 Y_{2i}+\lambda_1 Z_{1i}+\lambda_2 Z_{2i}+\delta\hat{v}_i+\varepsilon_i\\ &\leftarrow \left[\hat{v}_i=Y_{2i}-\hat{\pi}_0-\hat{\pi}_1 Z_{1i}-\hat{\pi}_2 Z_{2i}+\hat{\pi}_3 Z_{3i}-\hat{\pi}_4 Z_{4i}=Y_{2i}-\hat{Y}_{2i}\right]\end{aligned}$$

豪斯曼检验：操作过程

步骤1：明确外生变量（Z_{1i}, Z_{2i}）；疑似内生变量（Y_{2i}）及其工具变量（$Z_{3i}; Z_{4i}$）：

$$
\begin{aligned}
Y_{1i} &= \beta_1 + \beta_2 Y_{2i} + \lambda_1 Z_{1i} + \lambda_2 Z_{2i} + u_i && \text{(stucture eq.)} \\
Y_{2i} &= \pi_0 + \pi_1 Z_{1i} + \pi_2 Z_{2i} + \pi_3 Z_{3i} + \pi_4 Z_{4i} + v_i && \text{(1st eq. / reduce eq.)} \\
Y_{1i} &= \alpha_1 + \alpha_2 \hat{Y}_{2i} + \gamma_1 Z_{1i} + \gamma_2 Z_{2i} + \epsilon_i && \text{(2st eq.)}
\end{aligned}
$$

步骤2：采用**OLS**方法估计约简方程并得到残差 $\hat{v}_i$：

$$
\hat{v}_i = Y_{2i} - \hat{\pi}_0 - \hat{\pi}_1 Z_{1i} - \hat{\pi}_2 Z_{2i} + \hat{\pi}_3 Z_{3i} - \hat{\pi}_4 Z_{4i} = Y_{2i} - \hat{Y}_{2i}
$$

豪斯曼检验：操作过程

步骤3：采用**OLS**方法估计豪斯曼检验方程：

$$Y_{1i} = \beta_1 + \beta_2 Y_{2i} + \lambda_1 Z_{1i} + \lambda_2 Z_{2i} + \delta \hat{v}_i + \varepsilon_i$$

步骤4：做出判断并得到**内生性检验**的结论：

给出检验假设： $H_0 : \delta = 0; H_1 : \delta \neq 0$。在给定置信水平 α，对豪斯曼检验方程中的 δ进行t检验（需要进行**异方差矫正**）：

- 如果t检验**显著**，则拒绝原假设 H_0，表明 Y_{2i}是**内生的**。
- 如果t检验结果**不显著**，则不能拒绝原假设 H_0，表明 Y_{2i}是**外生的**。

已婚女性的教育回报案例

我们继续对已婚女性的教育回报案例进行分析讨论。

案例：明确变量和模型

如果认为工作经历 $exper$是**明确外生变量**。认为已婚女性的受教育年数 $educ$为**疑似内生变量**，如果找到它的2个**工具变量**分别为母亲受教育年数 $motheduc$和父亲受教育年数 $fatheduc$:

$$
\begin{aligned}
log(wage_i) &= \beta_1 + \beta_2 educ_i + \lambda_1 exper_i + \lambda_2 exper_i^2 + u_i && \text{(stucture eq.)} \\
educ_i &= \pi_0 + \pi_1 exper_i + \pi_2 exper_i^2 \\
&\quad + \pi_3 motheduc_i + \pi_4 fatheduc_i + v_i && \text{(1st eq. / reduce eq.)} \\
log(wage_i) &= \alpha_1 + \alpha_2 \widehat{educ_i} + \gamma_1 exper_i + \gamma_2 exper_i^2 + \epsilon_i && \text{(2st eq.)}
\end{aligned}
$$

案例：估计简约方程

采用**OLS方法估计约简方程**并得到残差 $\hat{\upsilon}_i$：

$$\hat{\upsilon}_i = Y_{2i} - \hat{\pi}_0 - \hat{\pi}_1 Z_{1i} - \hat{\pi}_2 Z_{2i} + \hat{\pi}_3 Z_{3i} - \hat{\pi}_4 Z_{4i} = Y_{2i} - \hat{Y}_{2i}$$

对以上**约简方程**进行OLS回归，估计结果为：

$$\begin{array}{llllll}
\widehat{educ} = & +9.10 & +0.05exper & -0.00I(exper^2) & +0.19fatheduc & +0.16motheduc \\
\text{(t)} & (21.3396) & (1.1236) & (-0.8386) & (5.6152) & (4.3906) \\
\text{(se)} & (0.4266) & (0.0403) & (0.0012) & (0.0338) & (0.0359) \\
\text{(fitness)} & R^2 = 0.2115; \bar{R}^2 = 0.2040 & & & & \\
 & F^* = 28.36;\ p = 0.0000 & & & &
\end{array}$$

案例：估计简约方程的残差数据

已婚女性的教育回报数据n=(428)

wage	educ	fatheduc	motheduc	exper	educ.hat	vi.hat
3.3540	12	7	12	14	12.7560	-0.7560
1.3889	12	7	7	5	11.7336	0.2664
4.5455	12	7	12	15	12.7720	-0.7720
1.0965	12	7	7	6	11.7677	0.2323
4.5918	14	14	12	7	13.9146	0.0854
4.7421	12	7	14	33	13.0294	-1.0294
8.3333	16	7	14	11	13.0112	2.9888
7.8431	12	3	3	35	10.4908	1.5092

Showing 1 to 8 of 428 entries

Previous 1 2 3 4 5 … 54 Next

案例：估计豪斯曼检验方程

为了检验 $educ_i$ 是否为**内生变量**，构建如下**豪斯曼检验模型**：

$$log(wage_i) = \beta_1 + \beta_2 educ_i + \lambda_1 exper_i + \lambda_2 exper_i^2 + \delta \hat{v}_i + \varepsilon_i$$

对以上**豪斯曼检验方程**进行OLS回归，估计结果为：

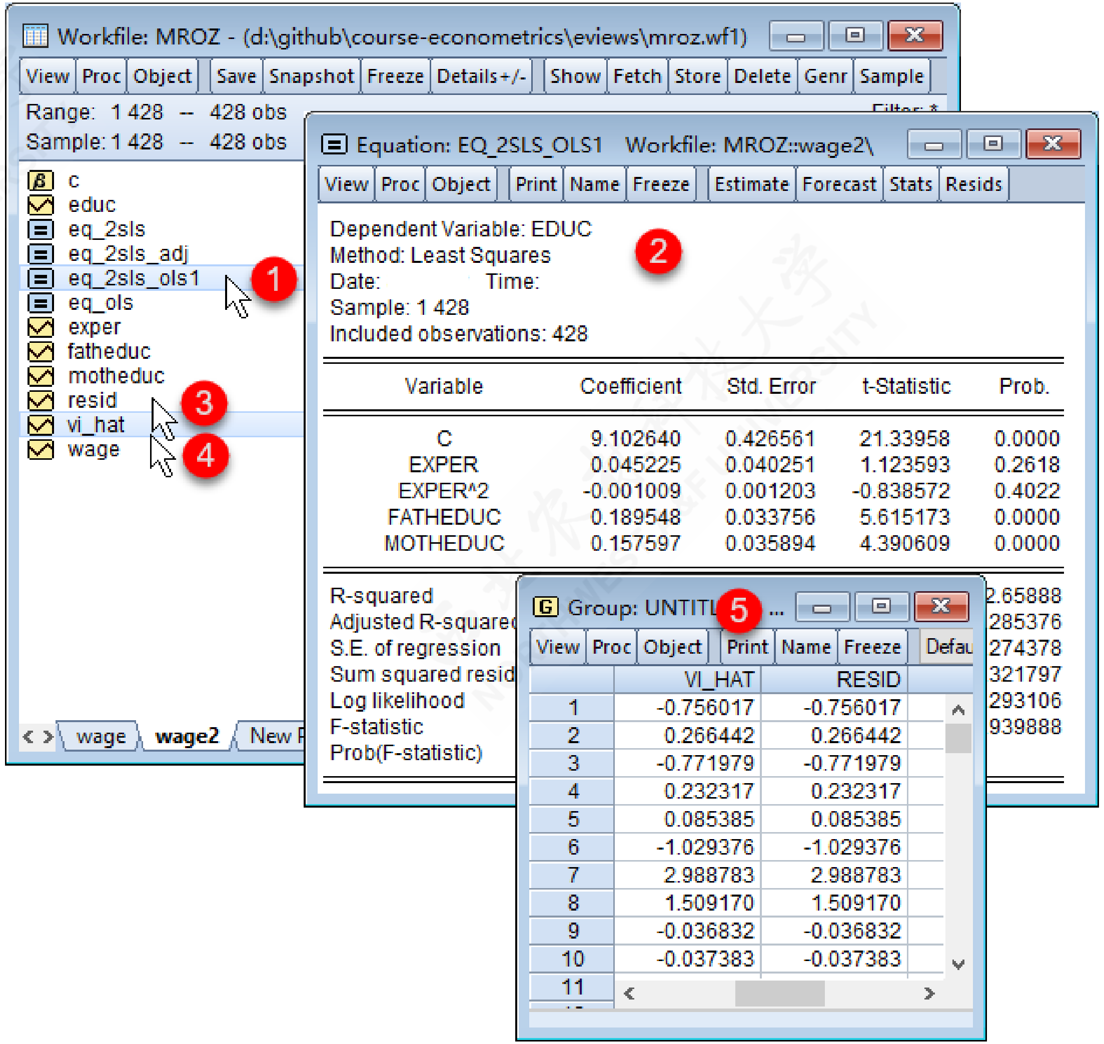
$$\begin{array}{llllll}
\widehat{log(wage)} = & +0.05 & +0.06 educ & +0.04 exper & -0.00 I(exper^2) & +0.06 vi.hat \\
\text{(t)} & (0.1219) & (1.9815) & (3.3363) & (-2.2706) & (1.6711) \\
\text{(se)} & (0.3946) & (0.0310) & (0.0132) & (0.0004) & (0.0348) \\
\text{(fitness)} & R^2 = 0.1624; \bar{R}^2 = 0.1544 & & & & \\
 & F^* = 20.50;\ p = 0.0000 & & & &
\end{array}$$

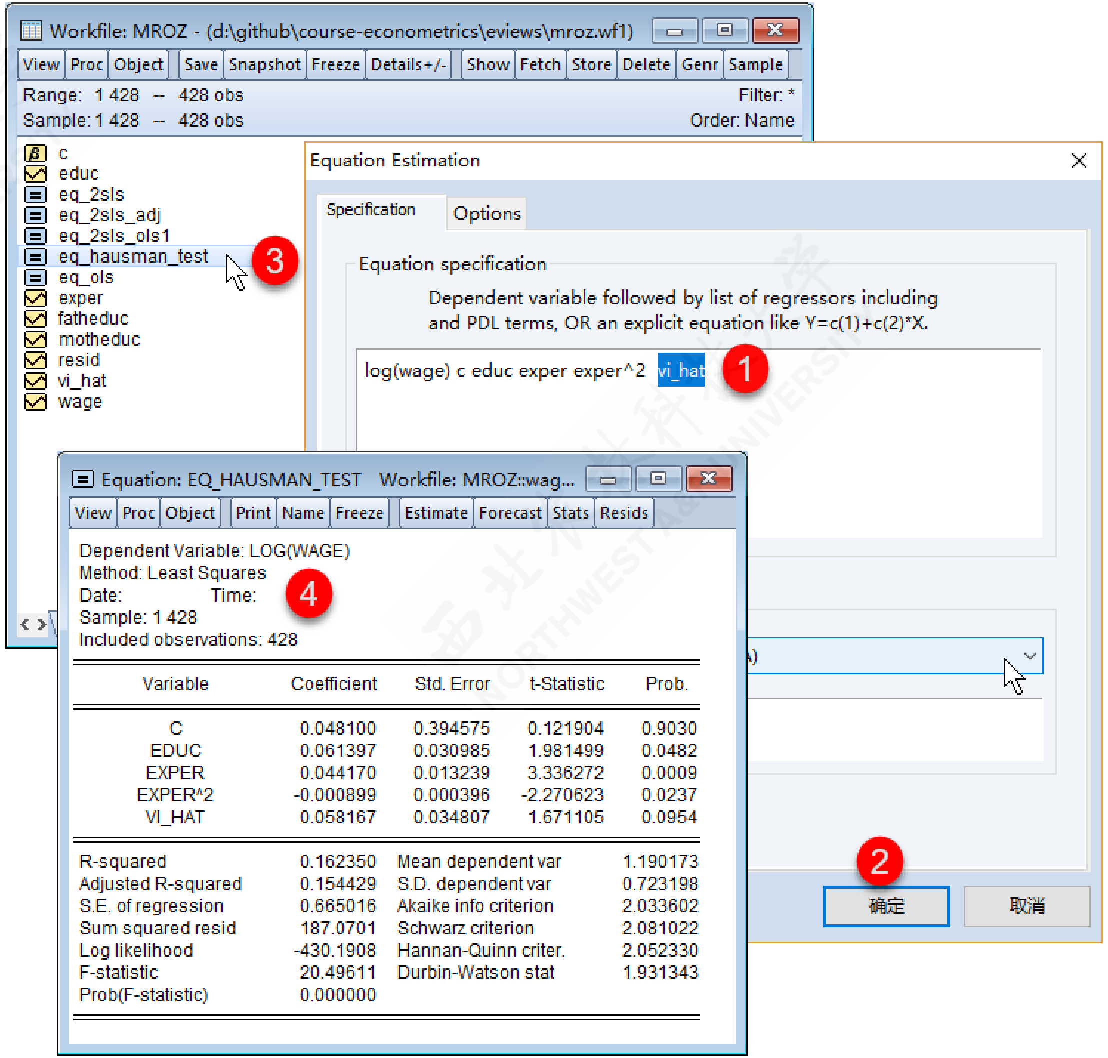
案例：得出豪斯曼检验结论

可以看到 $\hat{v}_i$（也即`vi.hat`）前的系数 δ对应的t样本统计量值为`1.6711`。

- 若给定置信水平 $\alpha=0.05$，查t表可知 $t_{1-\alpha/2}(n-k)=t_{0.975}(428-5)=1.9656$。因此 $t^*<t_{1-\alpha/2}(n-k)$，不能拒绝原假设 H_0，认为模型中的 $educ$是**外生的**。
- 若给定置信水平 $\alpha=0.1$，查t表可知 $t_{1-\alpha/2}(n-k)=t_{0.95}(428-5)=1.6485$。因此 $t^*>t_{1-\alpha/2}(n-k)$，则拒绝原假设 H_0，接受备择假设 H_1，认为模型中的 $educ$是**内生的**。
- 你如何看待以上的结论？

案例：豪斯曼检验EViews操作1





Dependent Variable: LOG(WAGE)
Method: Least Squares
Date: Time:
Sample: 1 428
Included observations: 428

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.048100	0.394575	0.121904	0.9030
EDUC	0.061397	0.030985	1.981499	0.0482
EXPER	0.044170	0.013239	3.336272	0.0009
EXPER^2	-0.000899	0.000396	-2.270623	0.0237
VI_HAT	0.058167	0.034807	1.671105	0.0954

R-squared	0.162350	Mean dependent var	1.190173
Adjusted R-squared	0.154429	S.D. dependent var	0.723198
S.E. of regression	0.665016	Akaike info criterion	2.033602
Sum squared resid	187.0701	Schwarz criterion	2.081022
Log likelihood	-430.1908	Hannan-Quinn criter.	2.052330
F-statistic	20.49611	Durbin-Watson stat	1.931343
Prob(F-statistic)	0.000000		

案例：豪斯曼检验EViews结果

Equation: EQ_HAUSMAN_TEST Workfile: MROZ::wag...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(WAGE)
Method: Least Squares
Date: Time:
Sample: 1 428
Included observations: 428

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.048100	0.394575	0.121904	0.9030
EDUC	0.061397	0.030985	1.981499	0.0482
EXPER	0.044170	0.013239	3.336272	0.0009
EXPER^2	-0.000899	0.000396	-2.270623	0.0237
VI_HAT	0.058167	0.034807	1.671105	0.0954

R-squared	0.162350	Mean dependent var	1.190173
Adjusted R-squared	0.154429	S.D. dependent var	0.723198
S.E. of regression	0.665016	Akaike info criterion	2.033602
Sum squared resid	187.0701	Schwarz criterion	2.081022
Log likelihood	-430.1908	Hannan-Quinn criter.	2.052330
F-statistic	20.49611	Durbin-Watson stat	1.931343
Prob(F-statistic)	0.000000		

11.6 工具变量的外生性检验

外生性检验的内容

工具变量（instrument variable）具备如下**两个性质**：

$$\begin{aligned} Cov(Z_i, u_i) &= 0 \qquad \text{(instrumental exogeneity)} \\ Cov(Z_i, X_i) &\neq 0 \qquad \text{(instrumental relevance)} \end{aligned}$$

我们现在来检验**第一个**条件：工具变量 Z_i是否真的是**外生性**的？

- 因为 u_i不能直接观测得到，所以第一个条件的检验往往不是很直接。
- 如果只有一个**工具变量**，第一个条件往往不能检验。（为什么？）

外生性检验：基本思路

考虑如下分析情形：有明确的**外生变量**（Z_{1i}, Z_{2i}）；有**疑似内生变量**（Y_{2i}）及其2个**工具变量**（$Z_{3i}; Z_{4i}$）。考虑如下的一个**偏误模型**（结构模型）：

$$Y_{1i} = \beta_1 + \beta_2 Y_{2i} + \lambda_1 Z_{1i} + \lambda_2 Z_{2i} + u_i \qquad \text{(stucture eq.)}$$

- 仅利用工具变量 Z_{3i}时的**2SLS**分析为：

$$Y_{2i} = \acute{\pi}_0 + \acute{\pi}_1 Z_{1i} + \acute{\pi}_2 Z_{2i} + \acute{\pi}_3 Z_{3i} + \acute{v}_i \qquad \text{(1st eq. / reduce eq.)}$$

$$Y_{1i} = \acute{\alpha}_1 + \acute{\alpha}_2 \acute{\hat{Y}}_{2i} + \acute{\gamma}_1 Z_{1i} + \acute{\gamma}_2 Z_{2i} + \acute{\epsilon}_i \qquad \text{(2st eq.)}$$

- 仅利用工具变量 Z_{4i}时的**2SLS**分析为：

$$Y_{2i} = \grave{\pi}_0 + \grave{\pi}_1 Z_{1i} + \grave{\pi}_2 Z_{2i} + \grave{\pi}_4 Z_{4i} + \grave{v}_i \qquad \text{(1st eq. / reduce eq.)}$$

$$Y_{1i} = \grave{\alpha}_1 + \grave{\alpha}_2 \grave{\hat{Y}}_{2i} + \grave{\gamma}_1 Z_{1i} + \grave{\gamma}_2 Z_{2i} + \grave{\epsilon}_i \qquad \text{(2st eq.)}$$

外生性检验：基本思路

- 如果：所有的 Z_{ki} $(k \in 1,2,3,4)$都是**外生的**；而且**疑似内生变量**（Y_i）确实与变量 Z_{3i}, Z_{4i}部分相关。
- 那么：两次**2SLS**估计得到的 $\acute{\alpha}_2$和 $\grave{\alpha}_2$就都是真值 α_2的**一致估计量**。
- 因此：我们可以基于这两个估计量的差（$\acute{\alpha}_2 - \grave{\alpha}_2$）来检验 Z_{3i}和 Z_{4i}是否都是**外生的**。——但是我们**很难区分**是哪一个或全部两个都是外生的？
- 然而：假设我们选择工具变量（Z_{3i}, Z_{4i}）的逻辑是相同的，如果一个**不是外生的**，那么另一个往往也**不是外生的**。

外生性检验：操作过程

步骤1：用两阶段法估计**结构方程**（使用2个工具变量），获得2SLS的残差 $\hat{u}_i$：

$$
\begin{aligned}
Y_{1i} &= \beta_1 + \beta_2 Y_{2i} + \lambda_1 Z_{1i} + \lambda_2 Z_{2i} + u_i && \text{(stucture eq.)} \\
Y_{2i} &= \pi_0 + \pi_1 Z_{1i} + \pi_2 Z_{2i} + \pi_3 Z_{3i} + \pi_4 Z_{4i} + v_i && \text{(1st eq. / reduce eq.)} \\
Y_{1i} &= \alpha_1 + \alpha_2 \hat{Y}_{2i} + \gamma_1 Z_{1i} + \gamma_2 Z_{2i} + \epsilon_i && \text{(2st eq.)}
\end{aligned}
$$

步骤2：构建外生性检验模型。将残差 $\hat{u}_i$对所有外生变量进行回归，并获得判定系数 R^2。

$$
\hat{u}_i = \lambda_0 + \lambda_1 Z_{1i} + \lambda_2 Z_{2i} + \lambda_3 Z_{3i} + \lambda_4 Z_{4i} + \varepsilon_i
$$

外生性检验：操作过程

步骤3：构建**卡方统计量**，根据原假设（ H_0 :全部工具变量都是**外生的**）/备择假设（ H_1 :部分工具变量不是**外生的**），做出判断并得到假设检验结论。

$$\chi^{2^*} = nR^2 \simeq \chi^2(q) \qquad \leftarrow [q = (n_{iv} - n_{env})]$$

- 卡方分布的自由度为 $q = (n_{iv} - n_{env})$，其中 n_{iv}指**工具变量**的个数（也即 $(mothedu_i, fatheduc_i)$），n_{env}指**内生变量**的个数（也即 edu_i）——（不包括结构模型中的因变量 $log(wage_i)$）。
- 如果卡方统计量值**大于**查表值，也即 $\chi^{2^*} > \chi^2_{1-\alpha}(q)$，则拒绝原假设 H_0，接收备择假设 H_1，表明**至少**部分工具变量**不是外生的**。
- 如果卡方统计量值**小于**查表值，也即 $\chi^{2^*} < \chi^2_{1-\alpha}(q)$，则接受原假设 H_0，表明**全部**工具变量**是外生的**。

已婚女性的教育回报案例

我们继续对已婚女性的教育回报案例进行**外生性检验**的分析讨论。

案例：明确变量和模型

如果认为工作经历 $exper$是**明确外生变量**。认为已婚女性的受教育年数 $educ$为**疑似内生变量**，如果找到它的2个**工具变量**分别为母亲受教育年数 $motheduc$和父亲受教育年数 $fatheduc$:

$$
\begin{aligned}
log(wage_i) &= \beta_1 + \beta_2 educ_i + \lambda_1 exper_i + \lambda_2 exper_i^2 + u_i && \text{(stucture eq.)} \\
educ_i &= \pi_0 + \pi_1 exper_i + \pi_2 exper_i^2 \\
&\quad + \pi_3 motheduc_i + \pi_4 fatheduc_i + v_i && \text{(1st eq. / reduce eq.)} \\
log(wage_i) &= \alpha_1 + \alpha_2 \widehat{educ}_i + \gamma_1 exper_i + \gamma_2 exper_i^2 && \text{(2st eq.)}
\end{aligned}
$$

案例：两阶段最小二乘法估计（2SLS）

采用**2SLS方法**估计前述**结构方程**并得到残差 $\hat{u}_i$，R软件估计结果为：

```
Call:
ivreg(formula = log(wage) ~ educ + exper + I(exper^2) | exper +
    I(exper^2) + fatheduc + motheduc, data = mroz)

Residuals:
    Min      1Q  Median      3Q     Max
-3.0986 -0.3196  0.0551  0.3689  2.3493

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0481003  0.4003281   0.120  0.90442
educ         0.0613966  0.0314367   1.953  0.05147 .
exper        0.0441704  0.0134325   3.288  0.00109 **
I(exper^2)  -0.0008990  0.0004017  -2.238  0.02574 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6747 on 424 degrees of freedom
Multiple R-Squared: 0.1357,    Adjusted R-squared: 0.1296
```

案例：2SLS的残差数据

已婚女性的教育回报数据n=(428)

wage	educ	fatheduc	motheduc	exper	educ.hat	vi.hat	ui.hat
3.3540	12	7	12	14	12.7560	-0.7560	-0.0169
1.3889	12	7	7	5	11.7336	0.2664	-0.6547
4.5455	12	7	12	15	12.7720	-0.7720	0.2690
1.0965	12	7	7	6	11.7677	0.2323	-0.9254
4.5918	14	14	12	7	13.9146	0.0854	0.3515
4.7421	12	7	14	33	13.0294	-1.0294	0.2930
8.3333	16	7	14	11	13.0112	2.9888	0.7127
7.8431	12	3	3	35	10.4908	1.5092	0.8300

Showing 1 to 8 of 428 entries

Previous 1 2 3 4 5 … 54 Next

案例：构建外生性检验模型

将残差 $\hat{u}_i$对所有外生变量进行回归，并获得判定系数 R^2。

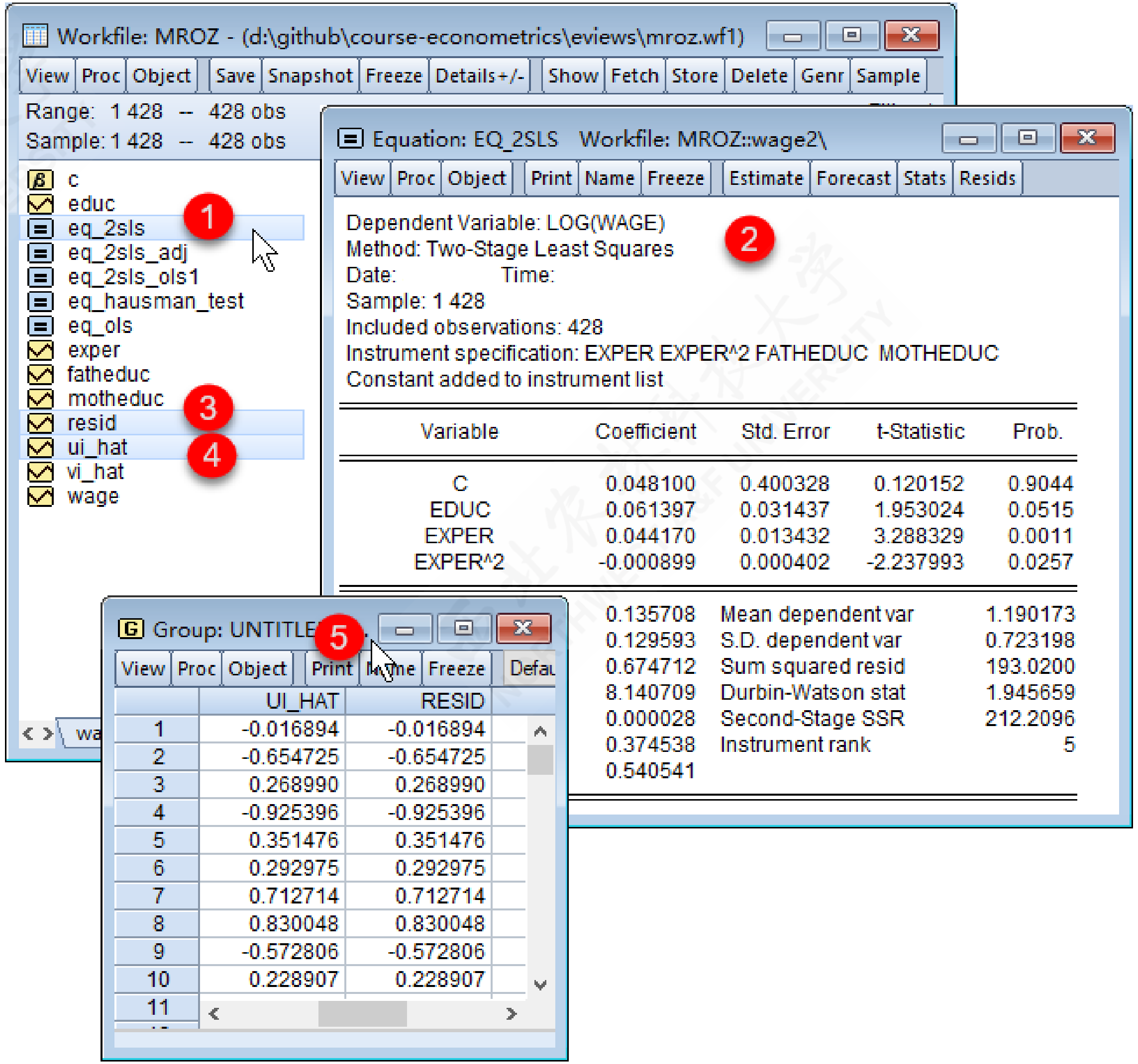
$$\hat{u}_i = \lambda_0 + \lambda_1 exper_i + \lambda_2 exper_i^2 + \lambda_3 montheduc_i + \lambda_4 fatheduc_i + \varepsilon_i$$

$$\begin{array}{llll}
\widehat{ui.hat} = & +0.01 & -0.00exper & +0.00I(exper^2) \\
\text{(t)} & (0.0776) & (-0.0014) & (0.0018) \\
\text{(se)} & (0.1413) & (0.0133) & (0.0004) \\
\text{(cont.)} & -0.01motheduc & +0.01fatheduc & \\
\text{(t)} & (-0.5558) & (0.5173) & \\
\text{(se)} & (0.0119) & (0.0112) & \\
\text{(fitness)} & R^2 = 0.0009; & \bar{R}^2 = -0.0086 & \\
 & F^* = 0.09; & p = 0.9845 &
\end{array}$$

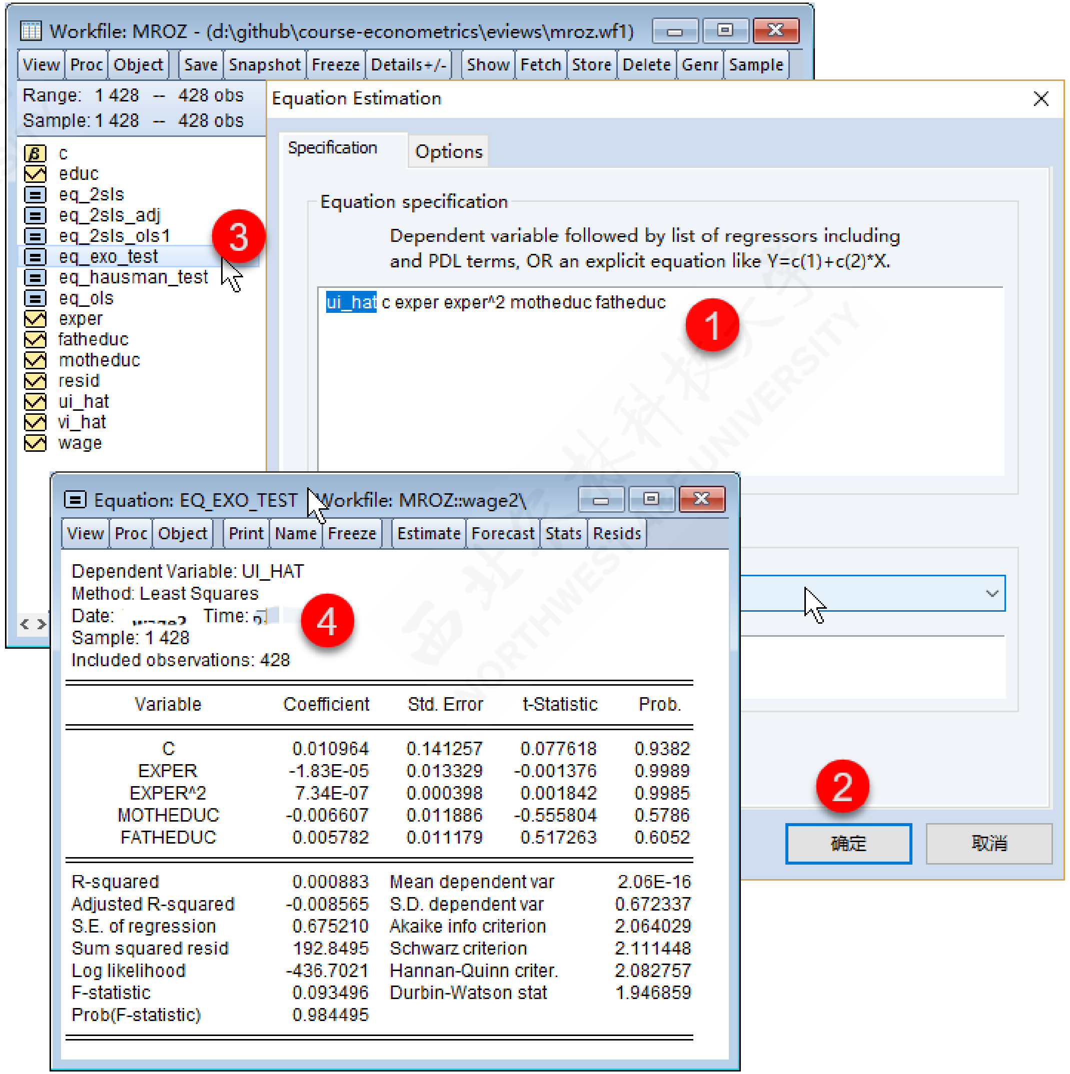
案例：外生性检验验结论

- 原假设（ H_0 :全部工具变量都是**外生的**）/备择假设（ H_1 :部分工具变量不是**外生的**）.
- 构建并计算**卡方统计量**。容易发现，判定系数 $R^2=0.0009$。可以计算得到样本卡方统计量为 $\chi^{2^*}=nR^2=0.3781$。
- 查卡方分布。容易计算出卡方分布的自由度为： $q=(n_{iv}-n_{env})=1$，其中 $n_{iv}=2$（**工具变量**的个数，也即 $(mothedu_i, fatheduc_i)$。 $n_{env}=1$（**内生变量**的个数，也即 edu_i）。若给定置信水平 $\alpha=0.05$，查卡方分布表可知 $\chi^2_{1-\alpha}(q)=\chi^2_{0.95}(1)=3.8415$。
- 给出假设检验结论。因为 $\chi^{2^*}<\chi^2_{1-\alpha}(q)$，不能拒绝原假设 H_0，认为模型中的工具变量 $(motheduc_i, fatheduc_i)$都是**外生的**。

案例：工具外生性检验EViews操作1



案例工具外生性检验EViews操作2



Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.010964	0.141257	0.077618	0.9382
EXPER	-1.83E-05	0.013329	-0.001376	0.9989
EXPER^2	7.34E-07	0.000398	0.001842	0.9985
MOTHEDUC	-0.006607	0.011886	-0.555804	0.5786
FATHEDUC	0.005782	0.011179	0.517263	0.6052

R-squared	0.000883	Mean dependent var	2.06E-16
Adjusted R-squared	-0.008565	S.D. dependent var	0.672337
S.E. of regression	0.675210	Akaike info criterion	2.064029
Sum squared resid	192.8495	Schwarz criterion	2.111448
Log likelihood	-436.7021	Hannan-Quinn criter.	2.082757
F-statistic	0.093496	Durbin-Watson stat	1.946859
Prob(F-statistic)	0.984495		

案例：工具外生性检验EViews结果

Equation: EQ_EXO_TEST Workfile: MROZ::wage2\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: UI_HAT
Method: Least Squares
Date: Time:
Sample: 1 428
Included observations: 428

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.010964	0.141257	0.077618	0.9382
EXPER	-1.83E-05	0.013329	-0.001376	0.9989
EXPER^2	7.34E-07	0.000398	0.001842	0.9985
MOTHEDUC	-0.006607	0.011886	-0.555804	0.5786
FATHEDUC	0.005782	0.011179	0.517263	0.6052

R-squared	0.000883	Mean dependent var	2.06E-16
Adjusted R-squared	-0.008565	S.D. dependent var	0.672337
S.E. of regression	0.675210	Akaike info criterion	2.064029
Sum squared resid	192.8495	Schwarz criterion	2.111448
Log likelihood	-436.7021	Hannan-Quinn criter.	2.082757
F-statistic	0.093496	Durbin-Watson stat	1.946859
Prob(F-statistic)	0.984495		

本章结束