

微观计量经济学

第一次上机课：工具变量估计

赵震宇

xishanyu2.github.io

1 本章概要

2 IV、2SLS 与 GMM

3 数据案例：药品支出

4 工具变量方法的实践

目录

- 1 本章概要
- 2 IV、2SLS 与 GMM
- 3 数据案例：药品支出
- 4 工具变量方法的实践

核心要点

- 问题：在模型 $y_i = \mathbf{x}_i' \boldsymbol{\beta} + u_i$ 中，若 $\text{Cov}[\mathbf{x}_i, u_i] \neq \mathbf{0}$ ，则 OLS 不一致。
- 解决方案：假设存在工具变量 $\mathbf{z}_i$ ，满足 $\text{Cov}[\mathbf{z}_i, u_i] = \mathbf{0}$ 。
- 若工具变量个数等于回归元个数：
 - 工具变量法 (IV)
- 若工具变量个数大于回归元个数：
 - 两阶段最小二乘 (2SLS)
 - 广义矩估计 (GMM)
- 假设检验：
 - 解释变量内生性
 - 弱工具变量问题

公式汇总

Table 6.1. IV estimators and their asymptotic variances

Estimator	Definition and estimate of the VCE
IV (just-identified)	$\hat{\beta}_{IV} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$ $\hat{V}(\hat{\beta}) = (\mathbf{Z}'\mathbf{X})^{-1}\hat{\mathbf{S}}(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}$
2SLS	$\hat{\beta}_{2SLS} = \{\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\}^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$ $\hat{V}(\hat{\beta}) = (\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\hat{\mathbf{S}}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$ $\times \{\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\}^{-1}$
Optimal GMM	$\hat{\beta}_{OGMM} = (\mathbf{X}'\mathbf{Z}\hat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\hat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{y}$ $\hat{V}(\hat{\beta}_{OGMM}) = (\mathbf{X}'\mathbf{Z}\hat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{X})^{-1}$

Stata 命令

IV (恰好识别)	<code>ivregress 2sls</code>
2SLS (过度识别)	<code>ivregress 2sls</code>
GMM (过度识别)	<code>ivregress gmm</code>
过度识别约束检验	<code>estat overid; xtoverid</code>
Hausman 检验 (同方差)	<code>hausman</code>
Hausman 检验 (异方差)	<code>estat endogenous</code>
弱工具变量	<code>estat firststage; condivreg; ivreg2</code>
静态面板 IV	<code>xtivreg; xthaustaylor</code>
动态面板 IV	<code>xtabond; xtdpdsys; xtdpd</code>
非线性 GMM	<code>gmm</code>

目录

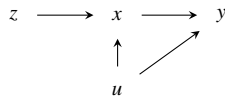
- 1 本章概要
- 2 IV、2SLS 与 GMM
- 3 数据案例：药品支出
- 4 工具变量方法的实践

工具变量的两个条件

假设存在工具变量 z ，满足：

- z 的变化不直接导致 y 的变化；
- z 的变化与 x 的变化相关。

路径关系为： z 不直接影响 y ，但 z 与 x 相关， x 再决定 y 。



- ① 外生性： z 与误差 u 不相关；
- ② 相关性： z 与回归元 x 相关。

总体矩与样本矩

模型为 $y_i = \mathbf{x}_i' \beta + u_i$ ，由于 $\text{Cov}[\mathbf{x}_i, u_i] \neq \mathbf{0}$ ，OLS 不一致。

假设存在工具变量 $\mathbf{z}_i$ ，使得 $\text{Cov}[\mathbf{z}_i, u_i] = \mathbf{0}$

在 $E[u_i | \mathbf{z}_i] = 0$ 下， $\text{Cov}[\mathbf{z}_i, u_i] = 0 \Rightarrow E[\mathbf{z}_i u_i] = \mathbf{0}$

因此总体矩条件为：

$$E[\mathbf{z}_i(y_i - \mathbf{x}_i' \beta)] = \mathbf{0}$$

对应的样本矩条件为：

$$\frac{1}{N} \sum_{i=1}^N \mathbf{z}_i(y_i - \mathbf{x}_i' \beta) = \mathbf{0}$$

IV 估计：恰好识别

在恰好识别情形下（工具变量个数 = 回归元个数），求解 k 个方程、 k 个未知数，得到工具变量估计量：

$$\hat{\beta}_{IV} = \left(\sum_{i=1}^N \mathbf{z}_i \mathbf{x}_i' \right)^{-1} \left(\sum_{i=1}^N \mathbf{z}_i y_i \right) = (\mathbf{Z}' \mathbf{X})^{-1} \mathbf{Z}' \mathbf{y}$$

Stata 命令示例：

```
ivregress 2sls y x_exog (x_endog = z)
```

通常只有一个回归元 $\mathbf{x}_i$ 是内生的（即与 u_i 相关），此时 $\mathbf{z}_i$ 中有一个变量是该内生回归元的工具变量，其余变量为外生变量；外生变量作为自身的工具变量。

GMM 估计：过度识别

在过度识别情形下（工具变量个数 > 回归元个数），不能直接求解 $\frac{1}{N} \sum_{i=1}^N \mathbf{z}_i(y_i - \mathbf{x}'_i\beta) = \mathbf{0}$ ，GMM 估计量最小化如下二次型：

$$Q(\beta) = \left[\frac{1}{N} \sum_{i=1}^N \mathbf{z}_i(y_i - \mathbf{x}'_i\beta) \right]' \mathbf{W}_N \left[\frac{1}{N} \sum_{i=1}^N \mathbf{z}_i(y_i - \mathbf{x}'_i\beta) \right]$$

其中 $\mathbf{W}_N$ 为对称、满秩、且不依赖于 β 的权重矩阵。

由 $\partial Q(\beta)/\partial \beta = 0$ 得到 GMM 估计量：

$$\hat{\beta}_{GMM} = (\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{y}$$

最优 GMM 与 2SLS (同方差)

- 当权重矩阵 $\mathbf{W}_N$ 是 $(\text{Var}[\mathbf{Z}'\mathbf{u}])^{-1}$ 的一致估计时, $\hat{\beta}_{GMM}$ 的方差最小。
- 在恰好识别情形 ($r = K$) 下, 对任意 $\mathbf{W}_N$, GMM 等价于 IV。
- 同方差误差下:

$$\text{Var}[\mathbf{Z}'\mathbf{u}] = \sigma^2 \sum_{i=1}^N \mathbf{z}_i \mathbf{z}_i'$$

2SLS 估计量设定:

$$\mathbf{W}_N = \left(\sum_{i=1}^N \mathbf{z}_i \mathbf{z}_i' \right)^{-1}$$

因此, $\hat{\beta}_{2SLS} = [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1} \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$

Stata 命令: **ivregress 2sls**; 建议使用稳健 VCE。

最优 GMM (异方差)

- 异方差误差下:

$$\text{Var}[\mathbf{Z}'\mathbf{u}] = \sum_{i=1}^N \sigma_i^2 \mathbf{z}_i \mathbf{z}_i'$$

“最优” GMM 设定:

$$\mathbf{W}_N = \left(\sum_{i=1}^N \hat{u}_i^2 \mathbf{z}_i \mathbf{z}_i' \right)^{-1}, \quad \hat{u}_i = y_i - \mathbf{x}_i' \hat{\beta}_{2SLS}$$

Stata 命令: `ivregress gmm`

关于 2SLS 的进一步说明

2SLS 之所以称为“两阶段”，是因为它可以分两步计算。

设 y_1 部分取决于内生变量 y_2 ，结构式方程： $y_{1i} = \beta_1 y_{2i} + \mathbf{x}'_{1i} \beta_2 + u_i$ 。

简约式方程： $y_{2i} = \mathbf{x}'_{1i} \pi_1 + \mathbf{x}'_{2i} \pi_2 + v_i$ ，其中 $\mathbf{x}_2$ 是 y_2 的一个或多个工具变量， $\mathbf{z}_i = (\mathbf{x}'_{1i}, \mathbf{x}'_{2i})'$ 。

- 直接对 y_1 关于 y_2 和 $\mathbf{x}_1$ 做 OLS 不一致；
- 2SLS 计算步骤：
 - ① 第一阶段：用 y_2 对 $\mathbf{x}_1, \mathbf{x}_2$ 做 OLS，得到预测值 $\hat{y}_2$ ；
 - ② 第二阶段：用 y_1 对 $\hat{y}_2$ 和 $\mathbf{x}_1$ 做 OLS。
- 但这一过程不能直接推广到非线性模型。

目录

- 1 本章概要
- 2 IV、2SLS 与 GMM
- 3 数据案例：药品支出**
- 4 工具变量方法的实践

数据与变量

研究美国老年人的药品支出 `ldrugexp`，核心解释变量为是否拥有私人健康保险的虚拟变量 `hi_empunion`，并控制外生变量。

```
use mus06data.dta
global x2list tothcr age female blhisp linc
keep if linc != .
describe ldrugexp hi_empunion $x2list
```

Variable name	Storage type	Display format	Value label	Variable label
<code>ldrugexp</code>	float	%9.0g		<code>log(drugexp)</code>
<code>hi_empunion</code>	byte	%8.0g		<code>Insured thro emp/union</code>
<code>tothcr</code>	byte	%8.0g		<code>Total chronic cond</code>
<code>age</code>	byte	%8.0g		<code>Age</code>
<code>female</code>	byte	%8.0g		<code>Female</code>
<code>blhisp</code>	float	%9.0g		<code>Black or Hispanic</code>
<code>linc</code>	float	%9.0g		<code>log(income)</code>

描述统计

```
summarize ldrugexp hi_empunion $x2list
```

Variable	Obs	Mean	Std. dev.	Min	Max
ldrugexp	10,089	6.481361	1.362052	0	10.18017
hi_empunion	10,089	.3821984	.4859488	0	1
totchr	10,089	1.860938	1.292858	0	9
age	10,089	75.05174	6.682109	65	91
female	10,089	.5770641	.4940499	0	1
blhisp	10,089	.1635445	.36988	0	1
linc	10,089	2.743275	.9131433	-6.907755	5.744476

样本为 65 岁及以上人群，38% 拥有雇主或工会提供的健康保险。

OLS 估计

```
regress ldrugexp hi_empunion $x2list, robust
```

Linear regression

Number of obs = **10,089**
 F(6, 10082) = **376.85**
 Prob > F = **0.0000**
 R-squared = **0.1770**
 Root MSE = **1.236**

ldrugexp	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
hi_empunion	.0738788	.0259848	2.84	0.004	.0229435	.1248141
totchr	.4403807	.0093633	47.03	0.000	.4220268	.4587346
age	-.0035295	.001937	-1.82	0.068	-.0073264	.0002675
female	.0578055	.0253651	2.28	0.023	.0080848	.1075262
blhisp	-.1513068	.0341264	-4.43	0.000	-.2182013	-.0844122
linc	.0104815	.0137126	0.76	0.445	-.0163979	.037361
_cons	5.861131	.1571037	37.31	0.000	5.553176	6.169085

OLS 结论：拥有私人保险使药品支出约增加 7.4%。若 hi_empunion 内生，则 OLS 不一致。

工具变量选择

为 `hi_empunion` 寻找有效工具变量，需要满足：

- 与误差项 u_i 不相关，即不直接进入 `ldrugexp` 方程；
- 与内生变量 `hi_empunion` 相关。

可能的工具变量 1（需要假设收入的直接作用已由 `linc` 充分控制）：

$$\text{ssiratio} = \frac{\text{社会保障收入}}{\text{其他来源收入}}$$

可能的工具变量 2（需要假设企业规模不直接影响 `ldrugexp`）：

$$\text{multlc} = 1\{\text{企业有多个经营地点}\}$$

工具变量与内生变量的相关性

```
describe ssiratio multlc
summarize ssiratio multlc
correlate hi_empunion ssiratio multlc
```

(obs=10,089)

	hi_emp~n	ssiratio	multlc
hi_empunion	1.0000		
ssiratio	-0.2124	1.0000	
multlc	0.1198	-0.1904	1.0000

相关性较低，提示可能存在弱工具变量问题。

IV 估计：一个工具变量

```
ivregress 2sls ldrugexp (hi_empunion = ssiratio) $x2list, robust
```

Instrumental-variables 2SLS regression	Number of obs	=	10,089
	Wald chi2(6)	=	2000.86
	Prob > chi2	=	0.0000
	R-squared	=	0.0640
	Root MSE	=	1.3177

ldrugexp	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
hi_empunion	-.8975913	.2211268	-4.06	0.000	-1.330992	-.4641908
totchr	.4502655	.0101969	44.16	0.000	.43028	.470251
age	-.0132176	.0029977	-4.41	0.000	-.0190931	-.0073421
female	-.020406	.0326114	-0.63	0.531	-.0843232	.0435113
blhisp	-.2174244	.0394944	-5.51	0.000	-.294832	-.1400167
linc	.0870018	.0226356	3.84	0.000	.0426368	.1313668
_cons	6.78717	.2688453	25.25	0.000	6.260243	7.314097

Endogenous: **hi_empunion**Exogenous: **totchr age female blhisp linc ssiratio**

2SLS 估计：两个工具变量

```
ivregress 2sls ldrugexp (hi_empunion = ssiratio multlc) $x2list, robust
```

Instrumental-variables 2SLS regression

Number of obs = **10,089**
 Wald chi2(6) = **1955.36**
 Prob > chi2 = **0.0000**
 R-squared = **0.0414**
 Root MSE = **1.3335**

ldrugexp	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
hi_empunion	-.9899269	.2045907	-4.84	0.000	-1.390917	-.5889365
totchr	.4512051	.0103088	43.77	0.000	.4310001	.47141
age	-.0141384	.0029	-4.88	0.000	-.0198223	-.0084546
female	-.0278398	.0321743	-0.87	0.387	-.0909002	.0352207
blhisp	-.2237087	.0395848	-5.65	0.000	-.3012934	-.1461239
linc	.0942748	.0218841	4.31	0.000	.0513828	.1371668
_cons	6.875188	.2578855	26.66	0.000	6.369741	7.380634

Endogenous: **hi_empunion**

Exogenous: **totchr age female blhisp linc ssiratio multlc**

结果解读

OLS 和 IV 结果对比：

- 系数由 OLS 的 0.074 变为 IV 的 -0.898 ，符号发生改变；
- 标准误由 OLS 的 0.026 增至 IV 的 0.221；
- 这体现 OLS 估计的不精确性。

IV 和 2SLS 结果对比：

- 系数由 IV 的 -0.898 变为 2SLS 的 -0.990 ；
- 标准误由 0.221 降至 0.205；
- 增加有效工具变量可提高效率。

最优 GMM 估计：异方差情形

```
ivregress gmm ldrugexp (hi_empunion = ssiratio multlc) $x2list, robust
```

Instrumental-variables GMM regression	Number of obs	=	10,089
	Wald chi2(6)	=	1952.65
	Prob > chi2	=	0.0000
	R-squared	=	0.0406
GMM weight matrix: Robust	Root MSE	=	1.3341

ldrugexp	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
hi_empunion	-.9932795	.2046731	-4.85	0.000	-1.394431	-.5921275
totchr	.4509508	.0103104	43.74	0.000	.4307428	.4711588
age	-.0141509	.0029014	-4.88	0.000	-.0198375	-.0084644
female	-.0281716	.0321881	-0.88	0.381	-.0912592	.034916
blhisp	-.2231048	.0395972	-5.63	0.000	-.3007139	-.1454957
linc	.0944632	.0218959	4.31	0.000	.0515481	.1373783
_cons	6.877821	.2579974	26.66	0.000	6.372155	7.383486

Endogenous: **hi_empunion**

Exogenous: **totchr age female blhisp linc ssiratio multlc**

估计量比较

```
quietly regress ldrugexp hi_empunion $x2list, robust
estimates store OLS

quietly ivregress 2sls ldrugexp (hi_empunion = ssiratio multlc) $x2list, robust
estimates store IV

quietly ivregress 2sls ldrugexp (hi_empunion = ssiratio) $x2list, robust
estimates store TWOSLS

quietly ivregress gmm ldrugexp (hi_empunion = ssiratio multlc) $x2list, robust
estimates store GMM

estimates table OLS IV TWOSLS GMM, b(%9.4f) se(%9.3f) stats(N r2 F)
```

结论：OLS 与 IV/2SLS/GMM 在 `hi_empunion` 系数上差异很大，说明内生性问题不可忽视。

目录

- 1 本章概要
- 2 IV、2SLS 与 GMM
- 3 数据案例：药品支出
- 4 工具变量方法的实践**

关键问题

- 是否真的需要使用工具变量？
 - Hausman 内生性检验。
- 工具变量是否有效（与误差项不相关）？
 - 若模型过度识别，可做过度识别约束检验。
- 工具变量是否与内生回归元弱相关？
 - 弱工具变量会损失效率；
 - 若非常弱，会导致有限样本偏误和检验水平扭曲。
- 工具变量应该有多少个？
 - 工具变量个数至少不少于内生回归元个数；
 - 理论上越多越好，但过多可能带来有限样本偏误。

Hausman 检验

```
quietly ivregress 2sls ldrugexp (hi_empunion = ssiratio multlc) $x2list
estimates store IV
quietly regress ldrugexp hi_empunion $x2list
estimates store OLS
hausman IV OLS
```

注意：hausman 不能加 `vce(robust)`, `vce(cluster cvar)`；不要写成 `hausman OLS IV`

```
ivregress 2sls ldrugexp (hi_empunion = ssiratio) $x2list, robust
estat endogenous
```

Hausman 检验

	—— Coefficients ——		(b-B) Difference	sqrt(diag(V_b-V_B)) Std. err.
	(b) IV	(B) OLS		
hi_empunion	-.9899269	.0738788	-1.063806	.1904319
totchr	.4512051	.4403807	.0108243	.0043312
age	-.0141384	-.0035295	-.0106089	.0020441
female	-.0278398	.0578055	-.0856453	.018387
blhisp	-.2237087	-.1513068	-.0724019	.0188348
linc	.0942748	.0104815	.0837933	.0160062

b = Consistent under H_0 and H_a ; obtained from `ivregress`.

B = Inconsistent under H_a , efficient under H_0 ; obtained from `regress`.

Test of H_0 : Difference in coefficients not systematic

$$\begin{aligned}\chi^2(6) &= (b-B)'[(V_b-V_B)^{-1}](b-B) \\ &= \mathbf{31.21}\end{aligned}$$

Prob > $\chi^2 = \mathbf{0.0000}$

Hausman 检验

Tests of endogeneity

H_0 : Variables are exogenous

Robust score chi2(1) = 24.935 (p = 0.0000)

Robust regression F(1,10081) = 26.4333 (p = 0.0000)

由于 $p = 0.000$, 拒绝 H_0 , 认为 hi_empunion 是内生的, 因此需要使用 IV。

过度识别检验

```
ivregress gmm ldrugexp (hi_empunion = ssiratio multlc) $x2list, wmatrix(robust)
estat overid
```

Test of overidentifying restriction:

Hansen's J $\chi^2(1) = 1.04754$ ($p = 0.3061$)

这里有一个过度识别约束（2个工具变量对应1个内生变量）。结果 $p \approx 0.306$ ，不拒绝 H_0 。在恰好识别约束有效的前提下，可认为过度识别约束也有效。

弱工具变量

```
ivregress 2sls ldrugexp (hi_empunion = ssiratio) $x2list, robust
estat firststage
```

First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	Robust F(1,10082)	Prob > F
hi_empunion	0.0761	0.0755	0.0179	65.7602	0.0000

第一阶段回归

```
ivregress 2sls ldrugexp (hi_empunion = ssiratio) $x2list, robust first
```

First-stage regressions

Number of obs = **10,089**
 F(6, 10082) = **119.18**
 Prob > F = **0.0000**
 R-squared = **0.0761**
 Adj R-squared = **0.0755**
 Root MSE = **0.4672**

hi_empunion	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
totchr	.0127865	.0036655	3.49	0.000	.0056015	.0199716
age	-.0086323	.0007087	-12.18	0.000	-.0100216	-.0072431
female	-.07345	.0096392	-7.62	0.000	-.0923448	-.0545552
blhisp	-.06268	.0122742	-5.11	0.000	-.08674	-.0386201
linc	.0483937	.0066075	7.32	0.000	.0354417	.0613456
ssiratio	-.1916432	.0236326	-8.11	0.000	-.2379678	-.1453186
_cons	1.028981	.0581387	17.70	0.000	.9150172	1.142944

第二阶段回归

Instrumental-variables 2SLS regression

Number of obs = **10,089**
 Wald chi2(6) = **2000.86**
 Prob > chi2 = **0.0000**
 R-squared = **0.0640**
 Root MSE = **1.3177**

ldrugexp	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
hi_empunion	-.8975913	.2211268	-4.06	0.000	-1.330992	-.4641908
totchr	.4502655	.0101969	44.16	0.000	.43028	.470251
age	-.0132176	.0029977	-4.41	0.000	-.0190931	-.0073421
female	-.020406	.0326114	-0.63	0.531	-.0843232	.0435113
blhisp	-.2174244	.0394944	-5.51	0.000	-.294832	-.1400167
linc	.0870018	.0226356	3.84	0.000	.0426368	.1313668
_cons	6.78717	.2688453	25.25	0.000	6.260243	7.314097

Endogenous: hi_empunion

Exogenous: totchr age female blhisp linc ssratio

小测

Q1. 如果工具变量与内生变量的相关性很弱（第一阶段 $F < 10$ ），会发生什么？

- A. 2SLS 估计量仍然一致，只是标准误略微增大。
- B. 2SLS 的有限样本偏误方向趋近于 OLS 偏误，且标准误巨大。
- C. 2SLS 自动退化为 OLS，两者给出完全相同的估计结果。
- D. 弱工具变量问题可以通过增大样本量完全消除。

小测

Q2. 关于排他性约束 (Exclusion Restriction), 以下哪项说法是错误的?

- A. 排他性约束要求工具变量只能通过内生变量 X 影响 Y , 不能有独立的影响渠道。
- B. 在恰好识别 (一个 Z 、一个内生 X) 时, 排他性约束可以用 J 检验来验证。
- C. 排他性约束的合理性必须依赖经济理论和制度知识来论证, 而非纯统计检验。
- D. 工具变量的排他性约束被轻微违反时, 弱 IV 会大幅放大这个违反造成的偏误。

思考题

3. 解释为什么 J 检验在恰好识别（一个内生变量、一个工具变量）时无法实施，它的零假设是什么？
4. 使用 Python 或 Stata，模拟一个含有内生性和有效工具变量的 DGP；比较 OLS 和 2SLS 的估计结果；改变第一阶段 F 的大小（调节 IV 的系数），观察弱 IV 对 2SLS 分布的偏误和方差的影响。
5. 用真实数据（如 CFPS 或 CHARLS）复制一个简单的 IV 分析。推荐选题：(a) 教育的回报——以义务教育法或配偶教育作为工具变量；(b) 家庭收入对子女教育的影响——以政策冲击作为工具变量。
6. 在 Angrist & Krueger (1991) 的经典论文中，出生季度被用作教育的工具变量。你觉得出生季度满足排他性约束吗？为什么有人质疑这一工具变量的有效性？如何体现“排他性约束需要制度论证而非统计检验”这一原则？