

CH10, CH11 習題演練

陳家威¹

DECEMBER 20, 2022

¹R10323045@ntu.edu.tw

複習

什麼時候（最需要）考慮內生性

- Y 跟 X 被共同因子決定

- ▶ 薪資與工作決定 — 能力
- ▶ 嬰兒體重與母親抽菸 — 健康意識
- ▶ 道路出事率與車種 — 使用者族群

- Y 跟 X 同時被決定

- ▶ 價格與數量 — 需求 & 供給

工具變數與 2SLS

解決方法

1. 找一個影響 X 但不影響 Y 的變數
2. 做一次回歸把「乾淨的 X 」過濾出來
3. 再一次回歸，把 Y 對「乾淨的 X 」做回歸

可以用 `ivregress` 指令輕鬆做到

```
1 ivregress 2sls y x1 x2 (x3 = z1 z2 z3), first
```

- 解釋變數為 y
- $x3$ 為內生變數，有些因子共同影響 $x3$ 與 y
- $z1\ z2\ z3$ 影響 $x3$ 但不影響 y

聯立模型

如果模型長這樣

$$q_i = \alpha_1 + \alpha_2 p_i + \alpha_3 A_i + \alpha_4 B_i + u_i \quad \text{供給}$$

$$q_i = \beta_1 + \beta_2 p_i + \beta_3 C_i + \beta_4 D_i + v_i \quad \text{需求}$$

則 p, q 被共同決定，所以也內生性問題。

1. 先估計縮減式 $p_i = A + B + C + D$ 得到第一階段預測 \hat{p}_i
2. 估計供給：

$$q_i = \alpha_1 + \alpha_2 \hat{p}_i + \alpha_3 A_i + \alpha_4 B_i + u_i$$

3. 估計需求：

$$q_i = \beta_1 + \beta_2 \hat{p}_i + \beta_3 C_i + \beta_4 D_i + v_i$$

作法一、

```
1 ivregress 2sls q (p=C D) A B, first
2
3 ivregress 2sls q (p=A B) C D, first
```

作法二、用 3SLS 來估計聯立模型

```
1 reg3 (q p A B)(q p C D), endog(q p)
```

雞肉市場

考慮雞肉市場的供需

$$\ln(Q_t) = \alpha_1 + \alpha_2 \ln(P_t) + \alpha_3 \ln(Y_t) + \alpha_4 \ln(PB_t) + \alpha_5 POPGRO_t + e_t^d$$

$$\ln(QPROD_t) = \beta_1 + \beta_2 \ln(P_t) + \beta_3 \ln(PF_t) + \beta_4 TIME_t + \beta_5 \ln(QPROD_{t-1}) + e_t^s$$

■ 需求：

- ▶ 價錢
- ▶ 人均收入
- ▶ 牛肉價格
- ▶ 人口成長率

■ 供給：

- ▶ 價格
- ▶ 飼料價格
- ▶ 年份指數
- ▶ 上一期的供給量

先看需求

內生變數：消費量、價格

內生變數

由模型決定出來的變數稱為內生變數。在這裡價格與數量，是市場供需調整後定下的。

非模型決定則為外生變數

一個反向思考的方式為，其他被認為是外生變數的，有沒有可能其實有內生性？

例如有無可能「某些原因同時使雞肉供給數量減少，也造成出生率下降」？

先看需求

內生變數：消費量、價格

內生變數

由模型決定出來的變數稱為內生變數。在這裡價格與數量，是市場供需調整後定下的。

非模型決定則為外生變數

一個反向思考的方式為，其他被認為是外生變數的，有沒有可能其實有內生性？

例如有無可能「某些原因同時使雞肉供給數量減少，也造成出生率下降」？

內外生的判斷

內外生的判斷通常需要一些經濟理論模型與經濟直覺，也需要一些「故事」來 motivate 這樣的想法。本身並沒有一個絕對的對錯，但常常如果沒想到有哪些共同決定 Y 與 X 的故事，就會出現不正確的因果推論。

單純地進行估計

常見的錯誤：單純把結構式進行
OLS

需求線負斜率並不顯著

	ln_q
ln_p	-0.156 (0.0825)
ln_y	0.987*** (0.0630)
ln_pb	-0.158 (0.0897)
popgro	0.168*** (0.0326)
_cons	-6.197*** (0.635)

那些因素共同影響均衡需求量與價格？— 供給線！跟供給有關的有

- (價格)
- 飼料價格
- 年份指數
- 上一期的供給量

將這些變數作為工具變數，做兩階段估計

```
1 ivregress 2sls ln_q (ln_p = ln_pf time qprod_l lexpts_l)  
    ln_y ln_pb popgro
```

選定供給線中的外生變數當 IV 進行兩階段估計

需求線負斜率顯著

	ln_q	ln_q
ln_p	-0.156 (0.0825)	-0.255* (0.125)
ln_y	0.987*** (0.0630)	0.932*** (0.0867)
ln_pb	-0.158 (0.0897)	-0.0990 (0.0897)
popgro	0.168*** (0.0326)	0.223*** (0.0375)
_cons	-6.197*** (0.635)	-5.708*** (0.876)

檢定 2SLS

用以下指令來檢定第一階段

```
1 estat firststage
```

First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	F(4,31)	Prob > F
ln_p	0.9063	0.8852	0.3727	4.60547	0.0049

Minimum eigenvalue statistic = 4.60547

Critical Values	# of endogenous regressors:	1
Ho: Instruments are weak	# of excluded instruments:	4

2SLS relative bias	5%	10%	20%	30%
	16.85	10.27	6.71	5.34
2SLS Size of nominal 5% Wald test	10%	15%	20%	25%
	24.58	13.96	10.26	8.31
LIML Size of nominal 5% Wald test	5.44	3.87	3.30	2.98

一般選用 $F > 10$ 作為好的 IV 的標準 — 無法拒絕是一個弱 IV

也可以土法煉鋼去檢驗

```
1 reg ln_p ln_pf time qprod_l lexpts_l ln_y ln_pb popgro  
2 test ln_pf time qprod_l lexpts_l
```

第一階段的變數

在進行第一階段估計時，除了工具變數以外，其他外生變數也要一併納入估計。

而檢驗則只需要做工具變數的聯合檢定。

時間序列上的現象

老師跳過了時間序列的部分，透過這題稍微補充。

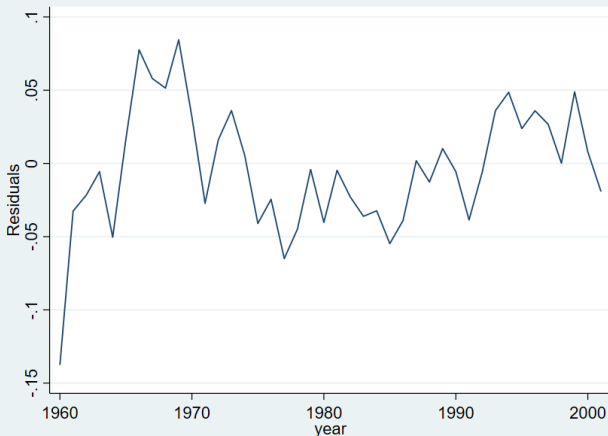


Figure: 殘差項的折線圖

時間序列上面的變異數異質性

異質性

- $e_i \sim \mathcal{N}(0, \sigma_i^2), \sigma_i^2 = h(x_i\beta)$

- $e_t = \rho e_{t-1} + \nu_t$

兩種都違反 OLS 的假設，第二種 $cov(e_t, e_{t-1}) \neq 0$ ，稱為自相關 (autocorrelation)。這時模型出現了自相關誤差 (autoregressive error)。

自相關誤差解決方法

與一般異質性問題類似，可以兩種做法

1. 將錯就錯，但好好把有自相關時的變異數計算出來 — Heteroskedasticity and Autocorrelation Consistent(HAC) Standard Error
2. 將變數做改變 — GLS

以這題為例，示範計算兩階段估計時，遇到有自相關的時候的穩健做法 — HAC 穩健標準誤差。

```
1 ivregress 2sls ln_q (ln_p = ln_pf time qprod_l lexpts_l)
   ln_y ln_pb popgro, ///
2 vce(hac nw 2) first
```

NEWKEY-WEST HAC 標準誤差

```
1 ivregress 2sls ln_q (ln_p = ln_pf time qprod_l lexpts_l)
   ln_y ln_pb popgro, ///
2 vce(hac nw 2) first
```

- 指令中的 vce 告訴 Stata 要特別計算標準誤差。
- hac 表示要考慮異質性 (heteroskedasticity) 與自相關 (autocorrelation) 的問題
- nw 表示要用 Newey-West (1987) 的計算方法 (不用管)
- 2 表示在考慮自相關時，要考慮兩期的落後，也就是 $e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \nu_t$ 。期數選擇超出範圍
- 如果不指定數字，就是用 N-2 期 lag (容易 over fit)

視覺化殘差像之間的自相關

一般來說會有兩種

■ Autocorrelation function —

$$\frac{\text{corr}(y_t, y_{t+h})}{\sqrt{\text{var}(y_t)\text{var}(y_{t+h})}}$$

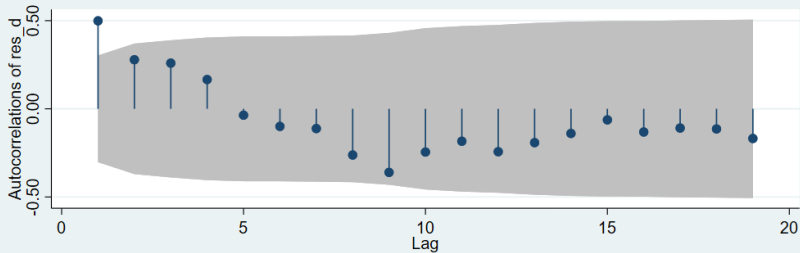
$$y_t = \hat{\rho}_{t-k}^{AC} y_{t-k} + \mu_t$$

■ Partial autocorrelation function —

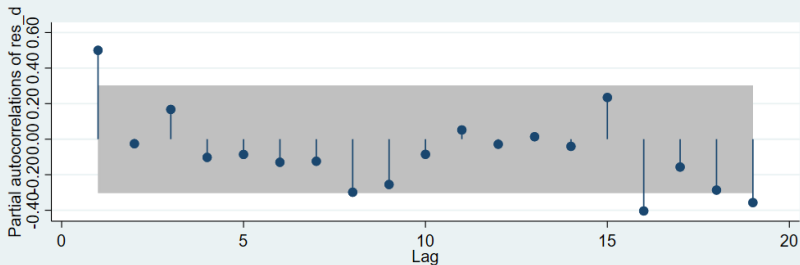
$$\frac{\text{corr}(y_t, y_{t+h} \mid y_{t \dots t+h-1})}{\sqrt{\text{var}(y_t \mid y_{t \dots t+h-1})\text{var}(y_{t+h} \mid y_{t \dots t+h-1})}}$$

$$y_t = \alpha + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \hat{\rho}_{t-k}^{PAC} y_{t-k} + \mu_t$$

選不同的 h ，都有對應的 AC, PAC，可以連同標準誤差畫出來



Bartlett's formula for MA(q) 95% confidence bands



95% Confidence bands [se = $1/\sqrt{n}$]

$$\ln(QPROD_t) = \beta_1 + \beta_2 \ln(P_t) + \beta_3 \ln(PF_t) + \beta_4 TIME_t + \beta_5 \ln(QPROD_{t-1}) -$$

內生變數：供給量、價格

外生變數：

- 飼料價格
- 年份指數
- 上一期的供給量

單純地進行估計

跟前面一樣，先看看常見的錯誤：
單純把結構式進行 OLS

供給線斜率不顯著為正

	ln_qprod
ln_p	0.0252 (0.0671)
ln_pf	-0.0999* (0.0421)
time	0.0113* (0.00503)
L.ln_qprod	0.727*** (0.104)
_cons	2.154** (0.782)

第一階段的顯著性

工具變數：

- 人均收入
- 牛肉價格
- 人口成長率
- 上一期的出口

第一階段聯合檢定

```
1 test ln_y ln_pb popgro L.  
lexpts
```

	ln_p
ln_y	0.856 (0.630)
ln_pb	0.219 (0.234)
popgro	-0.0231 (0.118)
L.lexpts	2.322** (0.709)
ln_pf	0.177 (0.108)
time	-0.0505* (0.0216)
L.ln_qprod	-0.141 (0.327)
_cons	-5.412 (6.297)

作法二

也可以直接在 ivregress

```
1 eststo est_2SLS : ivregress 2sls ln_qprod (ln_p = ln_y  
    ln_pb popgro L.lexpts) ///  
2     ln_pf time L.ln_qprod, first  
3  
4 estat first
```

First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	F(4,31)	Prob > F
ln_p	0.9063	0.8852	0.2594	2.71428	0.0478

Minimum eigenvalue statistic = 2.71428

Critical Values	# of endogenous regressors:	1
Ho: Instruments are weak	# of excluded instruments:	4

2SLS relative bias	5%	10%	20%	30%
	16.85	10.27	6.71	5.34
	10%	15%	20%	25%

供給函數的 2SLS

```
1 eststo est_2SLS : ivregress 2sls ln_qprod ///  
2   (ln_p = ln_y ln_pb popgro L.lexpts) ///  
3   ln_pf time L.ln_qprod, first
```

結果

	est_b	est_2SLS	est_2SLS_HAC
ln_p	0.0252 (0.0671)	0.0446 (0.123)	0.0446 (0.0891)
ln_p ^f	-0.0999* (0.0421)	-0.105* (0.0486)	-0.105 (0.0571)
time	0.0113* (0.00503)	0.0120* (0.00589)	0.0120* (0.00599)
L.ln_qprod	0.727*** (0.104)	0.718*** (0.110)	0.718*** (0.132)
_cons	2.154** (0.782)	2.214** (0.800)	2.214* (1.001)

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$