

CH8 習題演練

陳家威¹

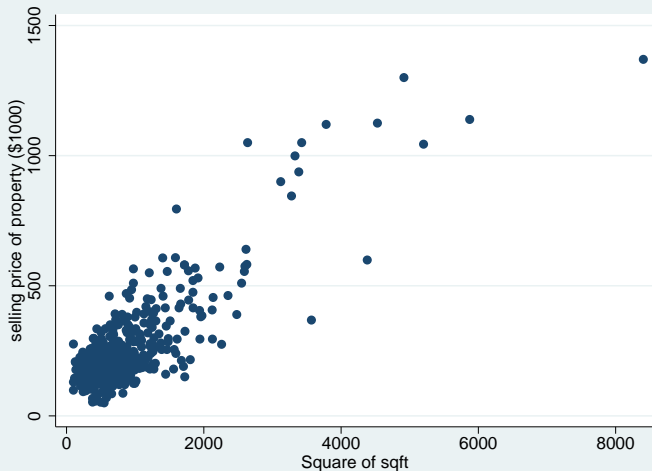
DECEMBER 1, 2022

¹R10323045@ntu.edu.tw

異質性變異數複習

異質變異數問題 HETEROSKEDASTICITY

- 當我的變異數隨 X 而改變
- 估計還是不偏且一致
- 但是估計值的標準誤差 (SE) 會有誤，導致檢定會出問題



哪裡出問題？

- 假設誤差項 e_i 的變異數同為 σ^2 之下，估計值 $\hat{\beta}_2$ 的變異數為

$$\text{var}(\hat{\beta}_2 \mid \mathbf{x}) = \frac{\sigma^2}{\sum_{i=1}^N (x_i - \bar{x})^2} = \sigma^2 (X'X)^{-1} \quad (8.6)$$

- 但實際上每一筆資料的誤差項，變異數不一樣，為 σ_i^2 ，用上面的就錯了，因為 β_2 的變異數這時會變成

$$\text{var}(\hat{\beta}_2 \mid \mathbf{x}) = \left[\sum_{i=1}^N (x_i - \bar{x}) \right]^{-1} \left[\sum_{i=1}^N (x_i - \bar{x}) \sigma_i^2 \right] \left[\sum_{i=1}^N (x_i - \bar{x}) \right]^{-1} \quad (8.8)$$

解決方法

1. 既然在有異質變異數之下用 OLS，估計值的變異數長這麼醜：

$$\text{var}(\hat{\beta}_2 \mid \mathbf{x}) = \left[\sum_{i=1}^N (x_i - \bar{x}) \right]^{-1} \left[\sum_{i=1}^N (x_i - \bar{x}) \sigma_i^2 \right] \left[\sum_{i=1}^N (x_i - \bar{x}) \right]^{-1}$$

那我乾脆就把上面的算出來吧：異質性穩健標準誤差
(heteroskedasticity-robust standard errors, HR)

2. 既然知道這樣會有異質變異數問題，那不然我改變一下資料，把變異數變成一樣不就好了：廣義最小平方法²

²課本翻成「一般化最小平方法」，但他們怎麼會覺得一般化 (GLS) 跟普通 (OLS) 一般人分辨得出來？...

檢測的方法

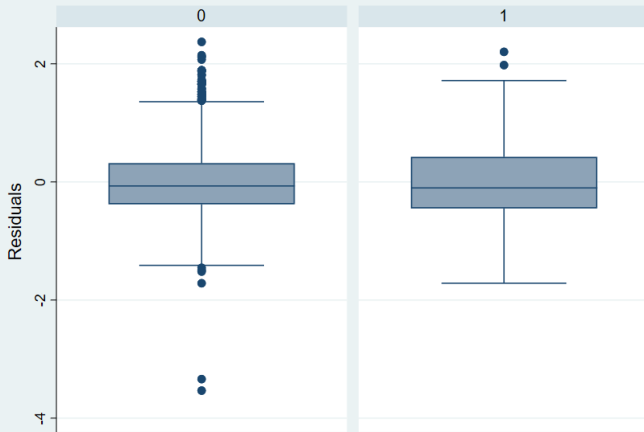
1. 分兩種，看他們的殘差項發散程度不一樣 — GQ test
檢定「醜人多作怪」
2. 懷疑殘差項跟 S 個 X 有關連，就拿殘差跟那些回歸
 - ▶ $NR^2 \sim \chi^2(S)$ — NR2 test
 - ▶ 聯合檢定殘差回歸的係數 — F test
3. 將所有變數、變數的平方、交乘項，全部納入回歸 — White Test

程式範例 — 8-15

- 墨西哥嫖妓交易資料
- 解釋變數： $\ln(\text{交易價格})$
- 被解釋變數：BAR、STREET、SCHOOL、AGE、RICH、ALCOHOL、ATTRACTIVE

有無質性？

先從直覺上，漂亮的性工作者，價錢或許得以持續開很高，而長相較為不佳者，價錢的變動或許較高。



Graphs by 1 if the sex worker is attractive; 0 otherwise

針對 ATTRACTIVE 測試異質變異數 — NR2 TEST

作法一、土法煉鋼

1. 拿殘差項平方對 ATTRACTIVE 回歸
2. 算出 $nr2 = N \times R^2$
3. 自由度 = 1
4. 算出 1% 顯著水準下的臨界值 $\text{invchi2tail}(1, 0.01)$
5. p-value : $\text{chi2tail}(1, nr2)$

```
1 predict res_a, res
2 gen res_a2=res_a^2
3
4 eststo het_nr2 : reg res_a2 attractive
5 scalar nr2=e(N)*e(r2)
6
7 di "NR2 value:" nr2
8 di "Critical value: " invchi2tail(1, 0.01)
9 di "P value :" chi2tail(1, nr2)
```

作法二、內建指令進行 NR2 test

```
1 est restore est_a  
2 estat hettest attractive, iid
```

注意！

在土法煉鋼算 NR2 test 時，要用的是殘差對變數的回歸。而在使用內建指令時，不用特別進行殘差項的回歸，而是要將原本的回歸變成目前的回歸（也就是要 `est restore est_a`）。

作法一、土法煉鋼

1. 拿殘差項平方對 ATTRACTIVE 回歸
2. (聯合) 檢定殘差項回歸的係數

```
1 predict res_a, res
2 gen res_a2=res_a^2
3 eststo het_nr2 : reg res_a2 attractive
4
5 test attractive
```

作法二、內建指令進行 F test

```
1 est restore est_a  
2 estat hettest attractive, fstat
```

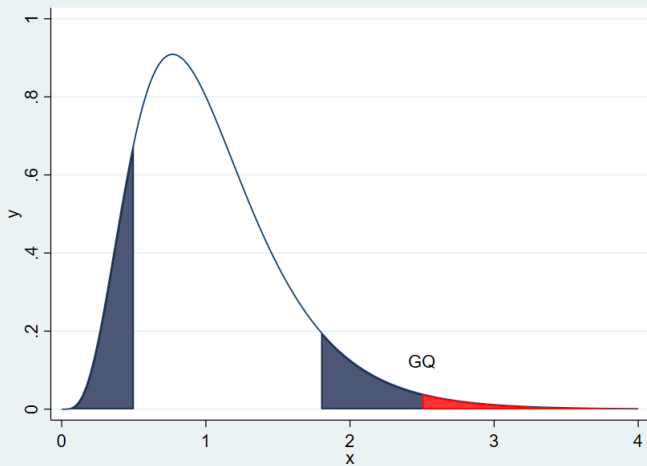
WHITE TEST

```
1 est restore est_a  
2 estat imtest, white
```

在 NR2 以及 F test 當中，可以任意選擇，你認為那些東西會影響變異數。但是 white test 則是將所有變數的組合與交乘項都考慮進去，對殘差項回歸，因此不用設定哪些會有影響。

GQ test 是另外一個檢定兩群體有無異質性差異的方式。

1. 針對醜人回歸，找出殘差的變異數 σ_0^2 。可以用 $\frac{SSE_0}{N_0-K}$ 估計
2. 針對美人回歸，找出殘差的變異數 σ_1^2 。可以用 $\frac{SSE_1}{N_1-K}$ 估計
3. 算出 $GQ = \frac{\sigma_0^2}{\sigma_1^2}$
4. 找出 $F_{(N_0-K, N_1-K)}$ 在左右兩尾的臨界值為何
5. 看 GQ 是不是在臨界值以外。若在外面，拒絕虛無假設，相信兩樣本誤差的變異數不一樣。




```

1
2  eststo est_c_att0 : reg lnprice bar street school age rich
      alcohol if attractive == 0
3  scalar sigma2_0=e(rss)/e(df_r)
4  scalar df_0 = e(df_r)
5
6  eststo est_c_att1 : reg lnprice bar street school age rich
      alcohol if attractive == 1
7  scalar sigma2_1=e(rss)/e(df_r)
8  scalar df_1 = e(df_r)
9
10 scalar gq = sigma2_1/sigma2_0
11 di "GQ          :" gq
12
13 di "L Critical value      :" invF(df_1, df_0, 0.025)
14 di "R Critical value      :" invFtail(df_1, df_0, 0.025)
15
16 di "P value      :" Ftail(df_1, df_0, max(gq, 1/gq) )
17

```

處理異質性變異數 — 穩健標準誤差

穩健標準誤差的好處在於，我們單純從殘差項來調整估計參數的標準誤差，不用考慮這個殘差項跟哪些變數有關係（因此名為穩健）

```
1 reg lnprice bar street school age rich alcohol attractive,  
   r
```

對，就這麼簡單

	(1)	(2)
	OLS	OLS with Robust SE
bar	0.216** (0.0786)	0.216* (0.0961)
street	-0.262*** (0.0794)	-0.262** (0.0968)
school	0.164*** (0.0238)	0.164*** (0.0244)
age	-0.0210*** (0.00145)	-0.0210*** (0.00130)
rich	0.292*** (0.0304)	0.292*** (0.0296)
alcohol	0.240*** (0.0358)	0.240*** (0.0377)
attractive	0.239*** (0.0316)	0.239*** (0.0374)
_cons	5.752*** (0.0913)	5.752*** (0.106)
<i>N</i>	3016	3016

Standard errors in parentheses

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

處理異質性變異數 — FGLS

假如我們已經知道變異數與哪些因素有關，則可以透過改變資料，或是在回歸中加入權重的方式，來處理異質變異數的問題——廣義最小平方法。

作法一、土法煉鋼

首先，我們假設變異數跟「年紀」與「美貌」有關

1. 先做一般的 OLS
2. 將殘差的平方取對數 $\ln(u_i^2)$ ，對「年紀」與「美貌」做回歸
3. 找出各資料的誤差的估計值，作為修正項 $\hat{h}_i = \exp(\ln(\hat{e}_i^2))$ ，
4. 將每一筆資料除上這個修正項，再進行一次回歸

```

1 est restore est_a          // 1
2
3 predict res_a, residual    // 2
4 gen ln_res_a2 = ln(res_a^2)
5 reg ln_res_a2 age attractive
6
7 predict ln_e_hat           // 3
8 gen e_hat=exp(ln_e_hat)
9
10 eststo est_FGLS_hand :    /// 4
11     reg lnprice bar street school age rich alcohol
    attractive ///
12     [aweight=1/e_hat]

```

FGLS 內建指令

```
1 eststo est_FGLS :    ///  
2     hetregress lnprice bar street school age rich alcohol  
   attractive, ///  
3     twostep het(age attractive)  
4
```

請向 Stata 工程師致敬！回歸人，我的超人

	(1) OLS	(2) OLS Robust	(3) FGLS by Hand	(4) FGLS	
	—	—	—	Inprice	Insigma2
bar	0.216** (0.0786)	0.216* (0.0961)	0.306*** (0.0779)	0.306*** (0.0779)	
street	-0.262*** (0.0794)	-0.262** (0.0968)	-0.172* (0.0784)	-0.172* (0.0784)	
school	0.164*** (0.0238)	0.164*** (0.0244)	0.141*** (0.0236)	0.141*** (0.0236)	
age	-0.0210*** (0.00145)	-0.0210*** (0.00130)	-0.0201*** (0.00141)	-0.0201*** (0.00141)	-0.00435 (0.00525)
rich	0.292*** (0.0304)	0.292*** (0.0296)	0.282*** (0.0296)	0.282*** (0.0296)	
alcohol	0.240*** (0.0358)	0.240*** (0.0377)	0.265*** (0.0352)	0.265*** (0.0352)	
attractive	0.239*** (0.0316)	0.239*** (0.0374)	0.240*** (0.0364)	0.240*** (0.0364)	0.410*** (0.118)
_cons	5.752*** (0.0913)	5.752*** (0.106)	5.635*** (0.0903)	5.635*** (0.0903)	-1.117*** (0.151)
N	3016	3016	3016	3016	

Standard errors in parentheses

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

樣本分群

剛剛只考慮了樣本分群，誤差變異數不同。但會不會係數也不一樣？

- 兩組樣本誤差是否一樣 — GQ test
- 兩組樣本係數是否一樣 — Chow test

	(1) Ugly	(2) Pretty
bar	0.498*** (0.0841)	-0.550** (0.197)
street	0.0242 (0.0842)	-1.142*** (0.233)
school	0.0939*** (0.0248)	0.452*** (0.0667)
age	-0.0177*** (0.00147)	-0.0411*** (0.00495)
rich	0.260*** (0.0303)	0.610*** (0.123)
alcohol	0.322*** (0.0366)	-0.103 (0.116)
_cons	5.362*** (0.0963)	7.143*** (0.257)
<i>N</i>	2600	416

Standard errors in parentheses

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

用 CHOW TEST 檢定兩組係數是否相同

```
1 eststo est_d : reg lnprice ( i.bar i.street i.school c.age  
    i.rich i.alcohol )##attractive  
2 testparm 1.attractive#1.* 1.attractive#c.*
```

其結果為：

$$F_{7,3002} = 29.95$$

$$Prob > F = 0.0000$$

兩者係數顯著不一樣。

這是一個可行的 CHOW TEST 嗎？

Chow test 的前提

Chow test 需要兩個群體的變異數是相同的！而前面已經檢定過，兩者變異數會不同！