

Chương V: HỒI QUI VỚI BIẾN GIẢ



Nội Dung

1. Bản chất của biến giả
2. Hồi qui với 1 biến định lượng & 1 biến định tính
3. Hồi qui với 1 biến định lượng và 2 biến định tính
4. Kiểm định tính ổn định cấu trúc các mô hình HQ – Kiểm định CHOW

I. Bản chất biến giả

1/ Biến định lượng → giá trị quan sát thể hiện bằng số.

VD: thu nhập, giá cả, lãi suất, ...

2/ Biến định tính → có hay không có 1 tính chất hoặc các mức độ một tiêu thức → hồi qui: biến giả.

VD: giới tính, dân tộc, tôn giáo, khu vực bán hàng, ...

3/ Lượng hoá biến định tính → biến giả (Dummy variables)

VD1: Năng suất của 2 công nghệ sản xuất (A và B)

Zi	0	1	1	0	0	1	0	1	1	0
Yi	28	32	35	27	25	37	29	34	33	30

Yi – Năng suất (tấn SP/ngày)

Zi = 1 → Công nghệ A ; Zi = 0 → công nghệ B

(1). Mô hình hồi quy: $Y_i = \beta_1 + \beta_2 X$

Hàm HQ: $Y_i = 27,8 + 6,4Z_i$, $R^2 = 0,7758$

- Với công nghệ A ($Z_i = 1$) $\rightarrow Y_i = 34,2$
- Công nghệ B ($Z = 0$) $\rightarrow Y_i = 27,8$

(2). Nếu mã hóa ngược lại:

$Z = 1$ (Công nghệ B) ; $Z = 0$ (Công nghệ A) ?

$$\rightarrow Y_i = 34,2 - 6,4Z_i$$

$$(A: Z = 0): Y = 34,2$$

$$(B: Z = 1): Y = 27,8$$

II.1. Hồi qui với 1 biến định lượng, 1 biến định tính. Biến định tính có 2 phạm trù

VD: $Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + U_i$

Với: Y_i : tiền lương công nhân ngành cơ khí (ngàn đ/tháng)

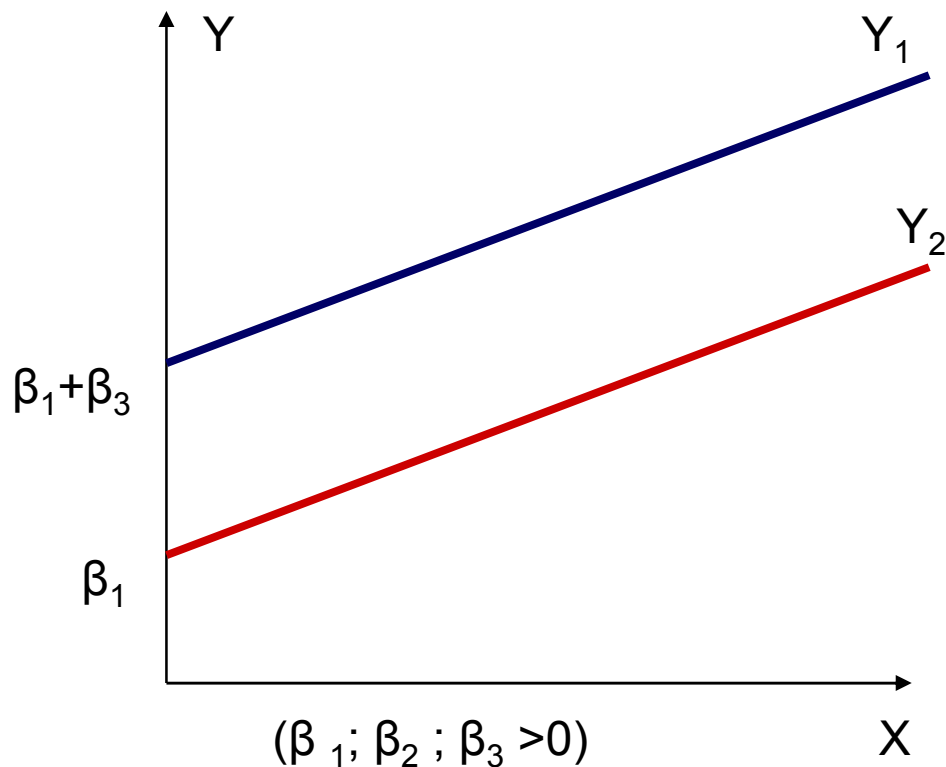
X_i : Học bậc

$D_i = 1$: khu vực tư nhân

$D_i = 0$: khu vực quốc doanh

- $Y_i = \beta_1 + \beta_2 X_i + U_i \rightarrow$ lương công nhân cơ khí quốc doanh
- $Y_i = \beta_1 + \beta_2 X_i + \beta_3 + U_i = (\beta_1 + \beta_3) + \beta_2 X_i + U_i \rightarrow$ lương công nhân cơ khí tư nhân
 - * β_3 : mức chênh lệch tiền lương công nhân cùng bậc thợ làm việc ở 2 khu vực
 - * β_2 : tốc độ tăng lương theo bậc thợ

Trường hợp 1: tung độ gốc khác nhau (lương khởi điểm khác nhau); hệ số góc bằng nhau (tốc độ tăng lương như nhau)



$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + U_i$$

a/ $D_i = 1 \rightarrow$ khu vực tư nhân
 $\rightarrow Y_1 = \beta_1 + \beta_2 X_i + \beta_3 + U_i$

Hay: $Y_1 = (\beta_1 + \beta_3) + \beta_2 X_i + U_i$

b/ $D_i = 0 \rightarrow$ khu vực quốc doanh

$\rightarrow Y_2 = \beta_1 + \beta_2 X_i + U_i$

c/ $(\beta_1 + \beta_3) > \beta_1 \rightarrow$ lương khởi điểm tư nhân > quốc doanh

Trường hợp 2: tung độ gốc bằng nhau (lương khởi điểm như nhau); hệ số góc khác nhau (tốc độ tăng lương khác nhau)

$\beta_1 = \text{nhau} \rightarrow$ sử dụng mô hình:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i D_i + U_i$$

Biến XD : biến tương tác, biểu thị ảnh hưởng đồng thời cả bậc thợ lẫn khu vực đối với tiền lương.

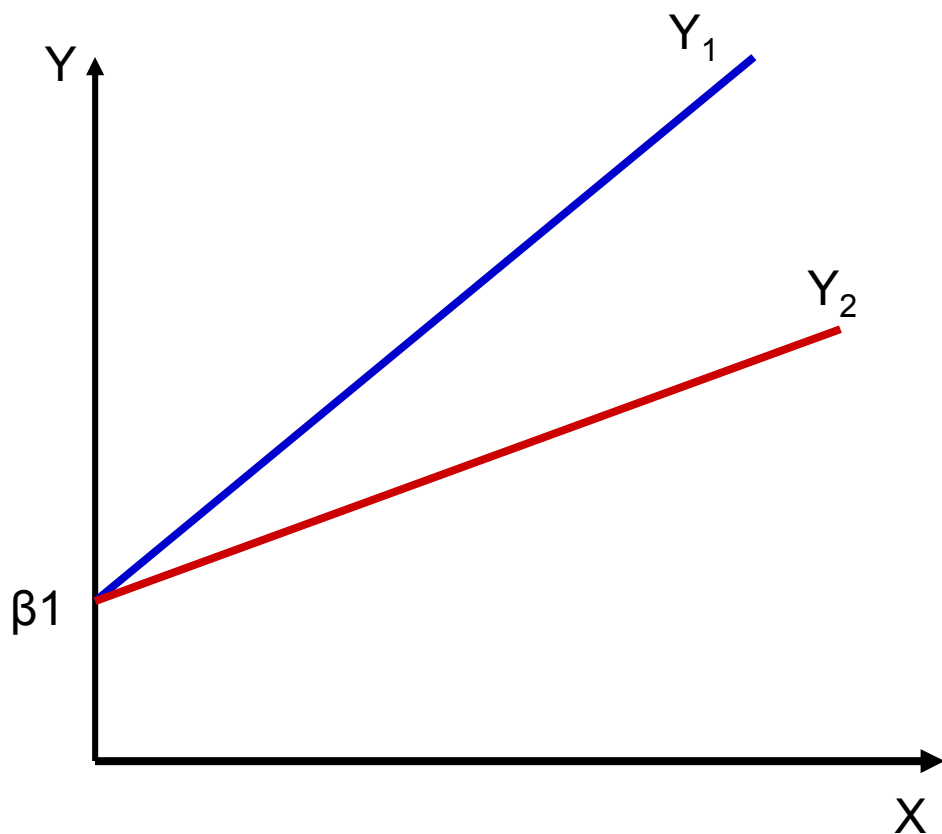
* Tiền lương trung bình công nhân cơ khí quốc doanh:

$$E(Y/X_i; D_i = 0): Y_2 = \beta_1 + \beta_2 X_i + U_i$$

* Tiền lương trung bình công nhân cơ khí tư nhân:

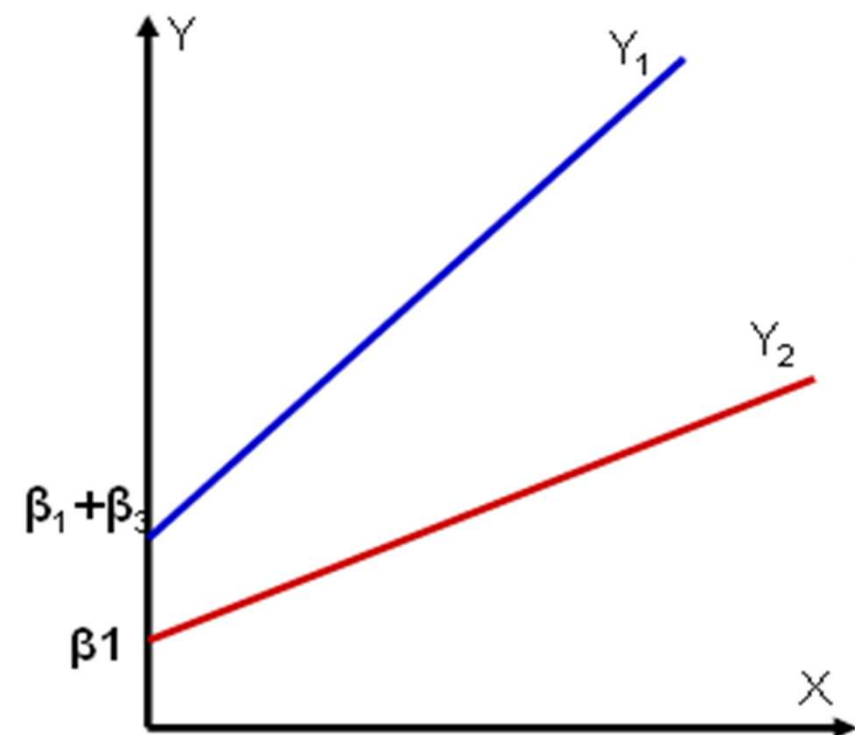
$$E(Y/X_i; D_i = 1) Y_1 = \beta_1 + (\beta_2 + \beta_3) X_i + U_i$$

* Nếu giả thiết $H_0: \beta_3 = 0$ bị bác bỏ \rightarrow tốc độ tăng lương 2 khu vực khác nhau, minh họa qua biểu đồ bên.



$(\beta_1; \beta_2; \beta_3 > 0)$

Trường hợp 3: tung độ gốc khác nhau (lượng khởi điểm khác nhau); hệ số góc khác nhau (tốc độ tăng lượng khác nhau)



$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + \beta_4 X_i D_i + U_i$$

a/Tiền lương trung bình công nhân cơ khí

quốc doanh:

$$E(Y/X_i; D_i = 0): Y_{2i} = \beta_1 + \beta_2 X_i + U_i$$

b/Tiền lương trung bình công nhân cơ khí tư

$$\text{nhân: } E(Y/X_i; D_i = 1): Y_{1i} = (\beta_1 + \beta_3) + (\beta_2 + \beta_4) X_i + U_i$$

- Giả thiết $H_0: \beta_3 = \beta_4 = 0 \rightarrow$ lương 2 khu vực như nhau

- Có ít nhất 1 trong 2 hệ số khác 0 và có ý nghĩa \rightarrow lương 2 khu vực khác nhau

- Chỉ β_4 khác 0, có ý nghĩa \rightarrow tốc độ tăng lương khác nhau

- Chỉ β_3 khác 0, có ý nghĩa \rightarrow tốc độ tăng lương như nhau, lương khởi điểm khác nhau

II.2. Hồi qui với 1 biến định lượng, 1 biến định tính. Biến định tính có nhiều hơn 2 phạm trù

- **Ví dụ 2:** Thu nhập bác sỹ theo thâm niên (biến định lượng) và nơi công tác (biến định tính) gồm thành phố, đồng bằng và miền núi
→ 3 phạm trù.
- Dùng mô hình: $Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_{1i} + \beta_4 D_{2i} + u_i$
Với: Y_i : thu nhập (tr đ/năm)
 X_i : thâm niên (năm)
 $D_{1i} = 1 \rightarrow$ công tác ở thành phố
 $D_{1i} = 0 \rightarrow$ công tác nơi khác
 $D_{2i} = 1 \rightarrow$ công tác vùng đồng bằng
 $D_{2i} = 0 \rightarrow$ nơi khác
- **Miền núi:** $E(Y/X_i; D_{1i} = 0, D_{2i} = 0): Y_1 = \beta_1 + \beta_2 X_i + U_i$
- **Đồng bằng:** $E(Y/X_i; D_{1i} = 0; D_{2i} = 1): Y_2 = \beta_1 + \beta_2 X_i + \beta_4 + U_i$
- **Thành phố:** $E(Y/X_i; D_{1i} = 1; D_{2i} = 0): Y_3 = \beta_1 + \beta_2 X_i + \beta_3 + U_i$

III. Hồi qui với 1 biến định lượng 2 biến định tính

n : số biến giả; k : số biến định tính;
 n_i : số phạm trù của biến định tính thứ i .

$$n = \sum_{i=1}^k (n_i - 1)$$

VD 3: Thu nhập bác sỹ theo thâm niên (biến định lượng), nơi công tác (biến định tính) gồm thành phố, đồng bằng và miền núi \rightarrow 3 phạm trù và thêm chuyên môn (biến định tính) gồm BS Tây y, Đông y và Xét nghiệm.

Dùng mô hình: $Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_{1i} + \beta_4 D_{2i} + \beta_5 D_{3i} + \beta_6 D_{4i} + U_i$

Với: Y_i : thu nhập (tr đ/năm)

X_i : thâm niên (năm)

$D_{1i} = 1 \rightarrow$ công tác ở thành phố

$D_{1i} = 0 \rightarrow$ nơi khác

$D_{2i} = 1 \rightarrow$ vùng đồng bằng

$D_{2i} = 0 \rightarrow$ nơi khác

$D_{3i} = 1 \rightarrow$ BS Tây y

$D_{3i} = 0 \rightarrow$ chuyên môn khác

$D_{4i} = 1 \rightarrow$ BS Đông y

$D_{4i} = 0 \rightarrow$ chuyên môn khác

Ví dụ: $E_1(Y/D_{1i} = 1; D_{2i}=0; D_{3i}=1; D_{4i}=0): Y_1 = \beta_1 + \beta_2 X_i + \beta_3 + \beta_5 + u_i$

Bác sỹ thâm niên X_i , công tác thành phố, chuyên môn Tây y.

$$Y = b_1 + b_2 X + b_3 + b_5$$

$$= (b_1 + b_3 + b_5) + b_2 X \text{ (TP \& Tây Y)}$$

$$Y = b_1 + b_2 X + b_4 + b_6$$

$$= (b_1 + b_4 + b_6) + b_2 X$$

(Đồng bằng & Đông Y)

→ Chênh lệch về thu nhập:

$$(b_1 + b_3 + b_5) - (b_1 + b_4 + b_6) = (b_3 + b_5) - (b_4 + b_6)$$

IV. Kiểm định tính ổn định cấu trúc của các mô hình hồi qui – Kiểm định CHOW

- Xét hai hay nhiều hồi qui có khác nhau không. Nếu khác, khác tung độ gốc hay hệ số góc hay cả hai.
- Các bước:**
 - 1/ Kết hợp các quan sát của cả 2 mẫu: $n = n_1 + n_2$
 - từ mẫu n , ước lượng $Y_i = \alpha_1 + \alpha_2 X_i + U_i$
 - Tính RSS với bậc tự do $(n_1 + n_2 - k)$ với k - số tham số
 - 2/ Ước lượng riêng từng mô hình, tính RSS_1 và RSS_2 với bậc tự do lần lượt $(n_1 - k)$ và $(n_2 - k)$
 - Đặt: $\overline{RSS} = RSS_1 + RSS_2$
 - 3/ Tính giá trị kiểm định

$$F_0 = \frac{(RSS - \overline{RSS}) / k}{\overline{RSS} / (n_1 + n_2 - 2k)}$$

Nếu $F_0 > F$ tới hạn $= F_{\alpha; (2; n_1 + n_2 - 2k)}$

- bác bỏ giả thiết cho rằng 2 đường hồi qui như nhau
- Các quan sát ở 2 nhóm không thể gộp với nhau

VD 4: Thời kỳ 1: (1946 -1954) ; Thời kỳ 2 (1955 – 1963)

Với: Y – tiết kiệm, X thu nhập

Y1	0.36	0.21	0.08	0.2	0.1	0.12	0.41	0.5	0.43
X1	8.8	9.4	10	10.6	11	11.9	12.7	13.5	14.3
Y2	0.59	0.9	0.95	0.82	1.04	1.53	1.94	1.75	1.99
X2	15.5	16.7	17.7	18.6	19.7	21.1	22.8	23.9	25.2

$$Y_1 = -0,26625 + 0,047X_1 \quad RSS_1 = 0,13965$$

$$Y_2 = -1,75 + 0,15045 X_2 \quad RSS_2 = 0,19312$$

$$Y_{2,1} = -1,082 + 0,117845X \quad RSS_{2,1} = 0,5722266$$

$$RSS_{2,1} = 0,13965 + 0,19312 = 0,33277$$

$$F_0 = \frac{(0,5722266 - 0,33277) / 2}{0,33277 / (9 + 9 - 4)} = 5,037$$

$$F \text{ tới hạn} = F_{0,05;(2,14)} = 3,74$$

$F_0 > F \text{ tới hạn} \rightarrow$ bác bỏ giả thiết cho rằng HQ Y1 và Y2 như nhau. Nghĩa là hàm tiết kiệm ở 2 thời kỳ khác nhau có ý nghĩa thống kê.

Bài tập 1

Y	15	15	16	16	17	17	18	18	19	20
X	5	5	4	4	3	4	4	3	3	2
Z	1	1	1	0	1	1	0	1	0	1

Y	11	10	12	16	15	12	12	13	14	14
X	8	8	7	4	5	7	7	6	6	5
Z	1	0	1	0	1	0	1	0	1	0

Y: lượng hàng bán được (tấn/tháng) ; X: giá bán (ngàn đ/kg)

Z=0 → nông thôn; Z=1 → thành phố

1/ Tìm các hàm HQ:

* $Y_1 = \alpha_1 + \alpha_2 X$ * $Y_2 = \beta_1 + \beta_2 X + \beta_3 Z$. Ý nghĩa β_2 và β_3

2/ Dự báo lượng hàng bán được, dùng hàm Y_1 hay hàm Y_2 ?

3/ Dùng Y_1 dự báo lượng hàng bán được khi giá bán là 7 ngàn đ/kg, độ tin cậy 95%.

1/ Hàm 2 biến: $Y = 22,67 - 1,5345 X$

$$\bar{R}_{2\text{biên}}^2 = 0,9455$$

Hàm 3 biến: $Y = 22,66 - 1,5328 X + 0,0975 Z$

$$\bar{R}_{3\text{biên}}^2 = 0,9427$$

2/ Kiểm định β_3 ($H_0 : \beta_3 = 0$) $\rightarrow p_value = 0,75 > 0,05$ Chấp nhận $H_0 \rightarrow \beta_3$ không có ý nghĩa thống kê, biến Z không ảnh hưởng lên biến $Y \rightarrow$ **Mô hình hàm 2 biến là mô hình phù hợp hơn**

* R^2 hàm 2 biến $>$ R^2 hàm 3 biến

* β_3 không có ý nghĩa thống kê, biến Z không ảnh hưởng lên Y

3/ Dự báo giá trị trung bình, độ tin cậy 95% (**Dựa vào hàm 2 biến**)

Bài tập 2

1. Giới tính có ảnh hưởng mức lương?
2. Dự báo lương khởi điểm một giáo viên nam có 8 năm kinh nghiệm, độ tin cậy 95%.
3. Dự báo lương khởi điểm một giáo viên nữ có 9 năm kinh nghiệm, độ tin cậy 98%.

Lương k . Điểm (Y)	Thâm niên (X)	Giới tính (Z)
23,0	1	1
19,5	1	0
24,0	2	1
21,0	2	0
25,0	3	1
22,0	3	0
26,5	4	1

Lương k . Điểm (Y)	Thâm niên (X)	Giới tính (Z)
23,1	4	0
25,0	5	0
28,0	5	1
29,5	6	1
26,0	6	0
27,5	7	0
31,5	7	1
29,0	8	0

Lương – ngàn USD năm; giới tính: nam = 1, nữ = 0; thâm niên – số năm công tác