

CHƯƠNG 4

PHÂN TÍCH HỒI QUY

VỚI BIẾN ĐỊNH TÍNH

i dung

1. Khái niệm biến giả
2. Mô hình có chứa biến độc lập là biến giả
3. Mô hình chứa biến giả và biến tương tác
4. Trường hợp biến định tính có nhiều phạm trù

1. Khái niệm biến giả

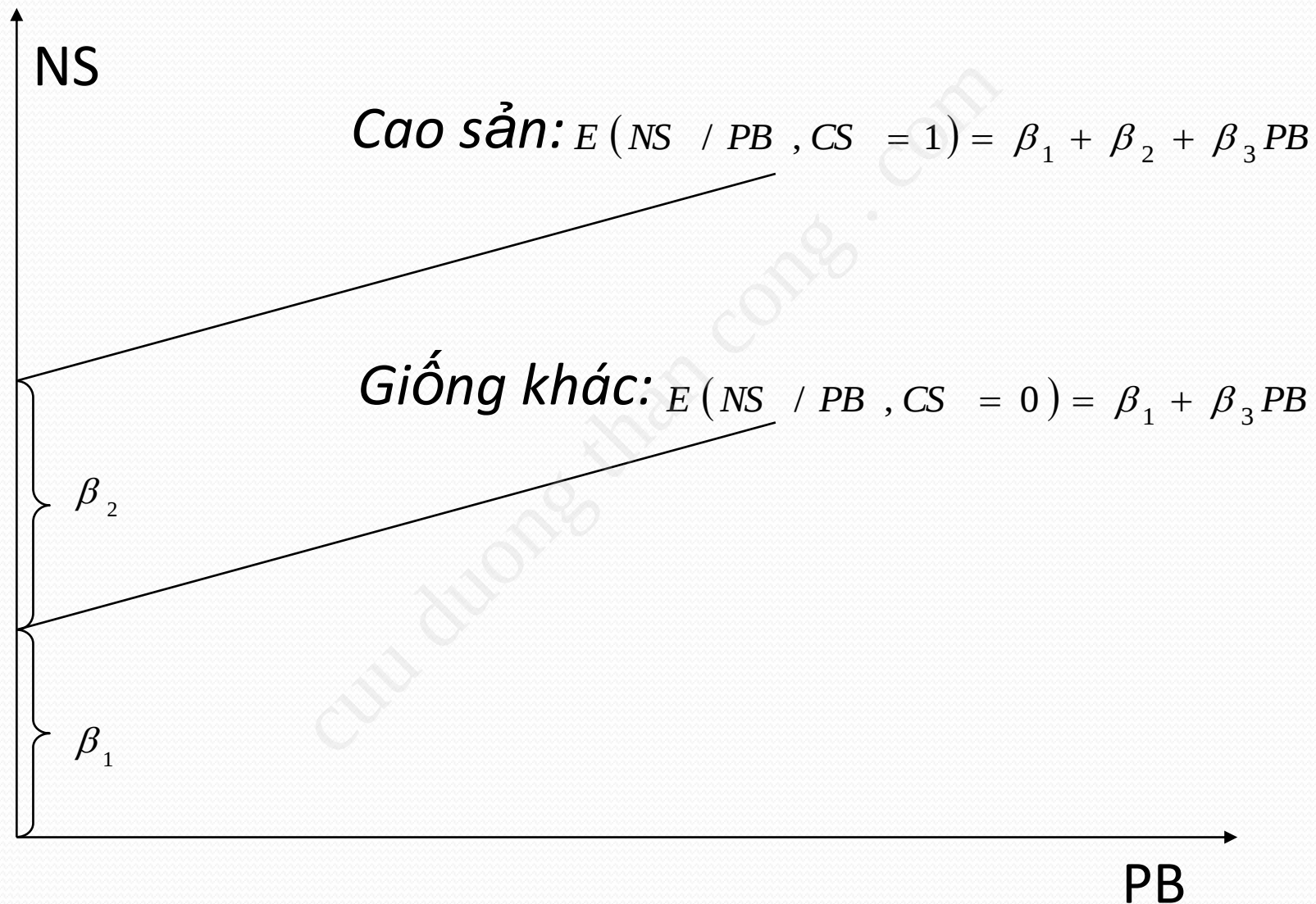
- Biến phụ thuộc không chỉ chịu tác động của các biến định lượng mà còn vào các yếu tố định tính.
- Việc đưa các thông tin mang tính định tính vào mô hình được thực hiện thông qua biến giả (dummy variable), nhận hai giá trị là 0 và 1.
- Giới tính có hai phạm trù là nam và nữ, ta dùng biến giả “Nữ” để lượng hóa như sau:
 - Nữ =1 nếu một người là nữ
 - Nữ =0 nếu một người **không** phải nữ (nam).
- Thông tin về trường học của sinh viên:
 - KTQD=1 nếu sinh viên thuộc trường ĐH KTQD
 - KTQD=0 nếu sinh viên **không** thuộc trường ĐH KTQD

Số liệu với biến giả về giới tính của người lao động

Thứ tự	Thu nhập	Chi tiêu	Giới tính	Nữ
1	24	20	<i>Nữ</i>	1
2	30	25	<i>Nam</i>	0
3	28	26	<i>Nữ</i>	1
..	..	..	..	..
n	37	30	<i>Nữ</i>	1

2. Mô hình có chứa biến độc lập là biến giả

- Năng suất lúa bị tác động bởi: Lượng phân bón; giống lúa (cao sản/ không phải cao sản); Yếu tố khác.
- Lượng hóa yếu tố giống lúa (mang tính chất định tính):
 - $CS=1$: nếu ruộng lúa thuộc giống cao sản
 - $CS=0$: nếu ruộng lúa thuộc giống lúa khác;
- Mô hình kinh tế lượng: $NS = \beta_1 + \beta_2 CS + \beta_3 PB + u$
 - Với ruộng lúa thuộc giống cao sản ($CS=1$)
$$NS_i = (\beta_1 + \beta_2) + \beta_3 PB_i + u_i$$
 - Với ruộng lúa thuộc giống lúa khác ($CS=0$)
$$NS_i = \beta_1 + \beta_3 PB_i + u_i$$
- Ý nghĩa của hệ số ước lượng?



- Một cách tổng quát, biến định tính Z có hai phạm trù và có tác động đến biến phụ thuộc Y.
- Gọi D là biến giả thể hiện biến định tính Z, trong đó D nhận giá trị 1 nếu quan sát thuộc nhóm 1 của biến Z, bằng 0 nếu quan sát thuộc nhóm còn lại.
- Bài toán hồi quy bội có chứa biến giả D như sau:

$$Y = \beta_1 + \beta_2 D + \beta_3 X_3 + \dots + \beta_k X_k + u$$

- Hệ số β_2 thể hiện sự khác biệt giữa giá trị trung bình của Y trong nhóm các quan sát thuộc nhóm 1 với các quan sát thuộc nhóm còn lại, khi các biến X_j là như nhau.

- Ví dụ 4.1. Ta có kết quả ước lượng sau đây:
$$\text{Wage} = 2565 + 157.76\text{Urban} + 57.88\text{Edu} + e$$
$$n = 935, R^2 = 0.14$$
- Trong đó:
 - Wage: mức lương (nghìn đồng/tháng)
 - Urban: biến giả nhận giá trị 1 nếu người lao động ở thành thị, 0 nếu người lao động ở nông thôn,
 - Edu: số năm đi học của người lao động.
- Ý nghĩa của các hệ số?
- Viết hàm hồi qui mẫu đối với lao động ở khu vực nông thôn.

3. Mô hình với biến giả và biến tương tác

- Hàm hồi quy đối với hai phạm trù không chỉ khác nhau về hệ số chặn mà còn khác nhau về hệ số góc.
- Để đưa sự khác biệt giữa các hệ số góc này vào mô hình, ta đưa thêm biến tương tác giữa biến giả và biến độc lập khác của mô hình. Chẳng hạn:

$$Wage = \beta_1 + \beta_2 Urban + \beta_3 Edu + \beta_4 Urban \times Edu + u$$

- Tích Urban x Edu được gọi là biến tương tác.
- Mô hình với các lao động thành thị:

$$Wage = (\beta_1 + \beta_2) + (\beta_3 + \beta_4) Edu + u$$

- Mô hình với các lao động ở nông thôn:

$$Wage = \beta_1 + \beta_3 Edu + u$$

- **Ý nghĩa của các hệ số?**

- Ví dụ 4.2. Sử dụng 935 quan sát như trong ví dụ 4.1 thu được kết quả hồi quy sau:

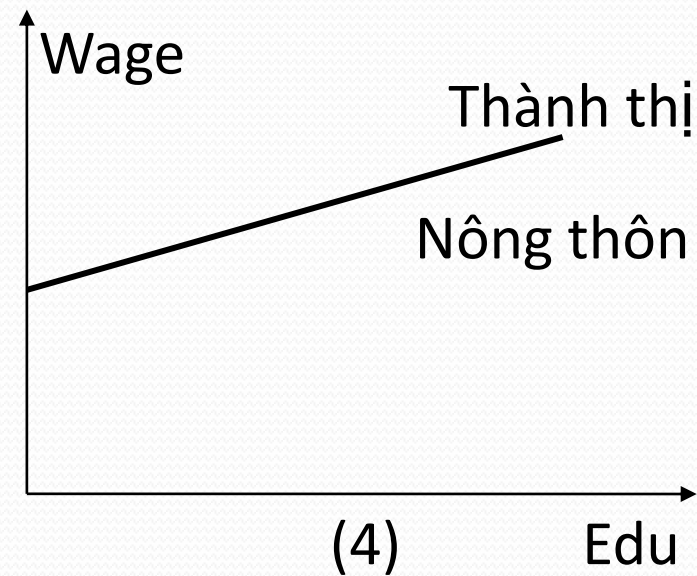
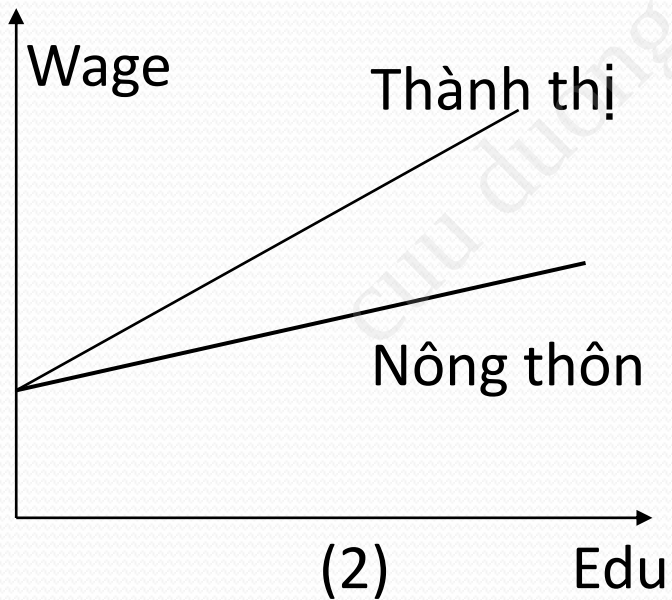
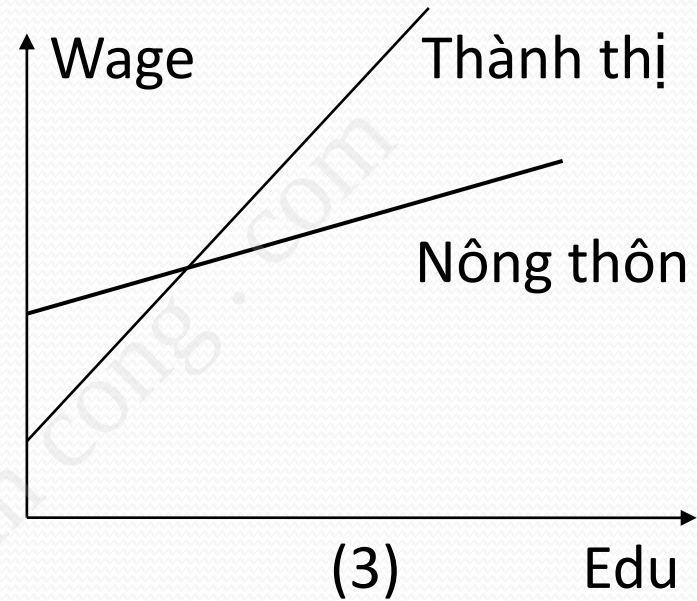
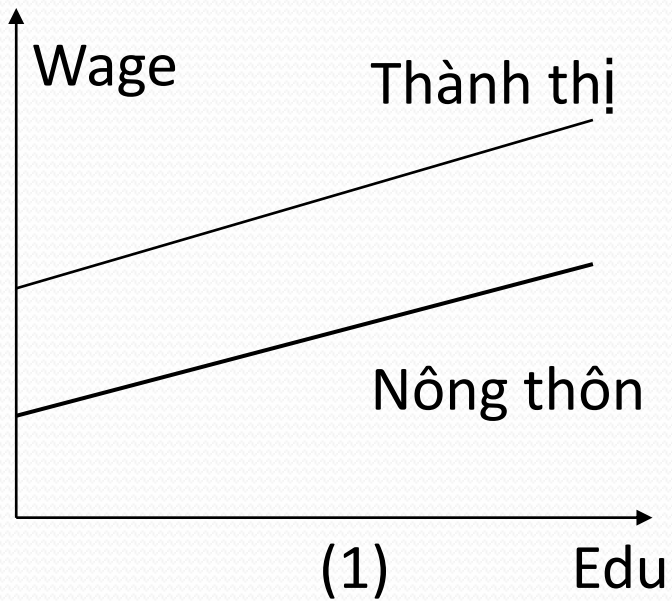
$$\text{Wage} = 2815.83 - 190.27\text{Urban} + 38.91\text{Edu} + 26.14\text{UrbanxEdu} + e$$

se (143.23) (169.39) (10.69) (12.56)

- Ý nghĩa của các hệ số ước lượng

Kiểm định sự khác biệt của hai hàm hồi qui của hai nhóm

- Giả sử chúng ta muốn biết mối quan hệ giữa học vấn và mức lương giữa lao động thành thị và lao động nông thôn có khác nhau hay không.
- Có 4 trường hợp có thể xảy ra:



- Ta có thể kiểm định điều này qua hai cách:
 - Kiểm định Chow
 - Kiểm định sử dụng biến giả

Kiểm định Chow

- Giả sử các quan sát ứng với lao động nông thôn được sắp xếp từ 1 đến n_1 , và các quan sát ứng với lao động thành thị từ $n_1 + 1$ đến n .
 - $Wage_i = \beta_1 + \beta_2 Edu_i + u_i \quad (i = 1, n_1) \quad (1a)$
 - $Wage_i = \alpha_1 + \alpha_2 Edu_i + u_i \quad (i = n_1+1, n) \quad (1b)$
- Mô hình gồm hai phương trình $\{(1a), (1b)\}$ là mô hình không có ràng buộc gì về sự giống nhau giữa hai nhóm lao động, nó gồm n phương trình tương ứng với n quan sát.
- Nếu quan hệ giữa Edu và Wage là như nhau giữa hai nhóm lao động thì ta có $\alpha_1 = \beta_1; \alpha_2 = \beta_2$

- Xét cặp giả thuyết:

- $H_0 : \alpha_1 = \beta_1; \alpha_2 = \beta_2; \quad H_1: \text{không phải } H_0.$

- Mô hình có ràng buộc sẽ là:

- $Wage_i = \beta_1 + \beta_2 Edu_i + u_i \quad (i=1, n) \quad (2)$

- Thực hiện kiểm định Chow

- Ước lượng mô hình (2) cho các tập số liệu sau:

- Nhóm lao động khu vực nông thôn, thu được RSS_1
 - Nhóm lao động khu vực thành thị, thu được RSS_2
 - Nhóm bao gồm lao động cả hai khu vực, thu được RSS

- Thống kê kiểm định:

$$F = \frac{[RSS - (RSS_1 + RSS_2)] / k}{(RSS_1 + RSS_2) / (n - 2k)}$$

$$W_\alpha = \{F, F > F_\alpha^{(k, n-2k)}\}$$

- Hạn chế của kiểm định Chow là:
 - Nếu hai mô hình là khác nhau thì kiểm định này không chỉ ra được là khác nhau ở hệ số nào, ở hệ số chặn, hệ số góc, hay cả hai loại hệ số?
 - Kiểm định Chow cũng đòi hỏi số quan sát ở mỗi nhóm phải đủ lớn, vì trong quá trình kiểm định chúng ta có thực hiện hồi quy cho mỗi nhóm.
- Phương pháp sử dụng biến giả có lợi thế là có thể khắc phục được cả hai hạn chế trên.

Kiểm định sử dụng biến giả

- Ví dụ 4.3. Trở lại kết quả ước lượng trong ví dụ 4.2 với mô hình:

$$\text{Wage} = \beta_1 + \beta_2 \text{Urban} + \beta_3 \text{Edu} + \beta_4 \text{Urban} \times \text{Edu} + u$$

- Để xét xem có sự khác biệt giữa hai nhóm lao động hay không, chúng ta kiểm định cặp giả thuyết sau:

$$H_0 : \beta_2 = \beta_4 = 0; H_1 : \beta_2^2 + \beta_4^2 \neq 0$$

- (Thực hiện thông qua kiểm định F – thu hẹp hồi qui)

4. Trường hợp biến định tính có nhiều phạm trù

- Trong phần trên ta xét trường hợp biến định tính có **hai** phạm trù, và khi đó có thể "số hóa" bằng cách sử dụng **một** biến giả 0 – 1.
- Trong thực tế, nhiều trường hợp các biến định tính có chứa nhiều hơn hai phạm trù. Ví dụ, Biến loại hình doanh nghiệp có ba phạm trù: doanh nghiệp nhà nước, doanh nghiệp tư nhân, doanh nghiệp có vốn đầu tư nước ngoài. Ta có thể sử dụng hai biến giả:
 - $NN=1$ nếu doanh nghiệp là doanh nghiệp nhà nước;
 - $NN=0$ nếu không phải là doanh nghiệp nhà nước.
 - $TN=1$ nếu doanh nghiệp là doanh nghiệp tư nhân
 - $TN=0$ nếu không phải là doanh nghiệp tư nhân.

Thứ tự	Q	K	Hình thức	NN	TN
1	240	120	Nhà nước	1	0
2	300	125	FDI	0	0
3	208	126	Tư nhân	0	1
4	350	200	Tư nhân	0	1
..	..	..		..	..
N	320	130	Nhà nước	1	0

Hai biến giả đủ để phân biệt ba phạm trù của biến định tính “loại hình doanh nghiệp”, doanh nghiệp nhà nước (NN=1, TN=0), doanh nghiệp tư nhân (NN=0, TN=1), doanh nghiệp có vốn đầu tư nước ngoài (NN=0, TN=0)

- Một cách tổng quát, để đưa biến định tính có m phạm trù vào mô hình, ta sẽ cần $(m-1)$ biến giả 0-1.
- *Phạm trù cơ sở*: Với mỗi biến định tính, ta gọi phạm trù ứng với trường hợp mà tất cả các biến giả nhận giá trị bằng 0 là phạm trù cơ sở (phạm trù gốc), với ý nghĩa các phạm trù khác sẽ được so sánh với phạm trù này.
- Trong ví dụ trên, nhóm « doanh nghiệp có vốn đầu tư nước ngoài » (bao gồm doanh nghiệp FDI) là phạm trù cơ sở.

Bẫy biến giả

- Về kỹ thuật, ta có thể tạo ra m biến giả 0-1 để biểu diễn thông tin cho biến định tính với m phạm trù. Tuy nhiên khi đó tổng các biến giả sẽ bằng 1 và ta gặp vấn đề về đa cộng tuyến hoàn hảo, nghĩa là giả thiết 4 bị vi phạm.
- Vì vậy nên người ta thường chỉ sử dụng $(m-1)$ biến giả cho biến định tính có m phạm trù.
- Trường hợp nếu muốn sử dụng cả m biến giả thì cần sử dụng mô hình không có hệ số chặn.

Thứ tự	Q	K	Hình thức	NN	TN	FDI
1	240	120	Nhà nước	1	0	0
2	300	125	FDI	0	0	1
3	208	126	Tư nhân	0	1	0
4	350	200	Tư nhân	0	1	0
..	..	..		..	..	
N	320	130	Nhà nước	1	0	0

Hộp 4.2: Hồi quy với biến giả

- Với biến định tính có m phạm trù cần sử dụng $(m-1)$ biến giả.
- Hệ số của biến giả dùng để so sánh hệ số chặn của nhóm với nhóm cơ sở.
- Hệ số của biến tương tác dùng để so sánh hệ số góc của một nhóm với nhóm cơ sở, tức là đánh giá sự khác biệt trong tác động của biến độc lập lên biến phụ thuộc giữa một nhóm với nhóm cơ sở.
- Khi sử dụng biến giả cần lưu ý tránh bẫy biến giả.

Khi biến định tính là biến thứ bậc

- Một trường hợp đặc biệt của biến định tính là biến thứ bậc (ordinal), trong đó các giá trị có thể có của nó có thể sắp xếp được theo thứ tự.
- Ví dụ thông tin về mức độ hài lòng của các doanh nghiệp về một loại dịch vụ công thường được xếp hạng thành:
 - 1: rất không hài lòng,
 - 2: không hài lòng,
 - 3: bình thường,
 - 4: hài lòng, và
 - 5: rất hài lòng.
- Tương tự như vậy, thông tin về trình độ học vấn của người dân cũng có thể được phân làm các nhóm như: dưới phổ thông trung học, phổ thông trung học, đại học và trên đại học.
- Các biến này có cùng một đặc tính chung là có thể xếp theo một thứ tự tăng dần (hay giảm dần) theo một tính chất nào đó.

- Chúng ta thường nhận được thông tin loại này dưới dạng số, nhưng khi đưa trực tiếp các biến này vào mô hình dưới dạng biến số thì các kết luận thu được sẽ thiếu chính xác.
- Trong các trường hợp này chúng ta phải sử dụng biến giả.