



Chương IV – Hồi qui với biến giả



Chương IV – Hồi qui với biến giả



- 1. Biến định tính – Kỹ thuật biến giả**
- 2. Các mô hình có chứa biến giả**
- 3. Thủ tục so sánh 2 hồi qui**



Chương IV – Hồi qui với biến giả



1. Biến định tính – Kỹ thuật biến giả:

1.1. Biến định tính (*Qualitative variables*):

- Ví dụ:

Chỉ tiêu \leftarrow Thu nhập, Số người của hộ và ***Khu vực sống (TT/NT), Giới tính chủ hộ, ...***

Lượng cầu \leftarrow Giá, Giá hàng hóa liên quan và ***Xuất xứ sản phẩm, Thị trường, Chính sách khuyến mại, ...***

Sản lượng \leftarrow Vốn, Lao động và ***Loại hình DN, Dây chuyền sản xuất, ...***

- Khái niệm

- Biến giả (*dummy variables, indicator variables, seasonal variables, categorical variables*)



Chương IV – Hồi qui với biến giả



1. Biến định tính – Kỹ thuật biến giả:

1.2. Kỹ thuật biến giả (*Dummy variables*):

- Bản chất của biến giả: các biến nhận giá trị = 1 đại diện cho quan sát có xuất hiện của thuộc tính và = 0 đại diện cho quan sát không xuất hiện thuộc tính được nghiên cứu.
- Có 2 kỹ thuật áp dụng biến giả trong mô hình kinh tế lượng: một biến định tính có m phạm trù (thuộc tính): A_1, A_2, \dots, A_m
 - ✓ Kỹ thuật 1: được mô tả bằng $(m-1)$ biến giả D_1, D_2, \dots, D_{m-1} với các giá trị 0 và 1 → **hồi qui bình thường có hệ số chặn**
 - ✓ Kỹ thuật 2: sử dụng m biến giả D_1, D_2, \dots, D_m với các giá trị 0 và 1 → **hồi qui không có hệ số chặn**



Chương IV – Hồi qui với biến giả



1. Biến định tính – Kỹ thuật biến giả:

1.2. Kỹ thuật biến giả (*Dummy variables*):

- Kỹ thuật 1:

$$D_1 = \begin{cases} 1 & \text{Quan sát tương ứng với } A_1 \\ 0 & \text{Quan sát tương ứng với khác } A_1 \end{cases}$$

$$D_2 = \begin{cases} 1 & \text{Quan sát tương ứng với } A_2 \\ 0 & \text{Quan sát tương ứng với khác } A_2 \end{cases}$$

...

$$D_{(m-1)} = \begin{cases} 1 & \text{Quan sát tương ứng với } A_{m-1} \\ 0 & \text{Quan sát tương ứng với khác } A_{m-1} \end{cases}$$



Chương IV – Hồi qui với biến giả



1. Biến định tính – Kỹ thuật biến giả:

1.2. Kỹ thuật biến giả (*Dummy variables*):

- Thuộc tính A_1, A_2, \dots, A_{m-1} : thuộc tính so sánh (*compared categories*)
- Thuộc tính A_m (thuộc tính bị thiếu): thuộc tính cơ sở (*base, benchmark, reference, omitted category*)
- Bẫy đa cộng tuyến hoàn hảo
- Chú ý: việc lựa chọn thuộc tính cơ sở khác nhau sẽ dẫn đến ước lượng hệ số chặn của mô hình khác nhau nhưng không làm thay đổi các kết luận trong nghiên cứu



Chương IV – Hồi qui với biến giả



1. Biến định tính – Kỹ thuật biến giả:

1.2. Kỹ thuật biến giả (*Dummy variables*):

- Ví dụ: Nghiên cứu sự phụ thuộc của thu nhập Y (1000 USD/năm) vào số năm đi học X và khu vực làm việc Dummy (1 = tư nhân, 0 = nhà nước)
- Chú ý: phần mềm Eviews không cho phép sử dụng kí hiệu D để đặt tên biến (nếu dùng D_1, D_2, \dots thì được chấp nhận)

X	6	7	8	9	10	12	14	15	17	18
Y	4.45	5.77	5.97	7.33	7.32	7.82	11.02	10.67	13.62	13.53
Dummy	1	1	0	1	1	0	1	0	1	0

Dependent Variable: Y

Method: Least Squares

Sample: 1 10

Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.978267	0.567355	-1.724259	0.1283
X	0.790624	0.039001	20.2721	0.0000
Dummy	0.928382	0.320818	2.893798	0.0232
R-squared	0.983906	Mean dependent var	8.75	
Adjusted R-squared	0.979307	S.D. dependent var	3.256269	
S.E. of regression	0.468413	Akaike info criterion	1.564394	
Sum squared resid	1.535877	Schwarz criterion	1.655169	
Log likelihood	-4.821968	F-statistic	213.9676	
Durbin-Watson stat	1.240617	Prob(F-statistic)	0.000001	



Chương IV – Hồi qui với biến giả



1. Biến định tính – Kỹ thuật biến giả:

1.2. Kỹ thuật biến giả (*Dummy variables*):

- Kỹ thuật 2:

$$D_1 = \begin{cases} 1 & \text{Quan sát tương ứng với } A_1 \\ 0 & \text{Quan sát tương ứng với khác } A_1 \end{cases}$$

$$D_2 = \begin{cases} 1 & \text{Quan sát tương ứng với } A_2 \\ 0 & \text{Quan sát tương ứng với khác } A_2 \end{cases}$$

...

$$D_m = \begin{cases} 1 & \text{Quan sát tương ứng với } A_m \\ 0 & \text{Quan sát tương ứng với khác } A_m \end{cases}$$



Chương IV – Hồi qui với biến giả



1. Biến định tính – Kỹ thuật biến giả:

1.2. Kỹ thuật biến giả (*Dummy variables*):

- Ví dụ: Nghiên cứu sự phụ thuộc của chi tiêu CONS (1000 USD/năm) vào khu vực sống và làm việc D1 (1 = Miền Tây, 0 = khác), D2 (1 = Miền Đông Bắc, 0 = khác) và D3 (1 = Miền Nam, 0 = khác) của giáo viên Mỹ năm 1986

CONS	2,3	3,7	3	3,3	3,8	2,5	3,4	8,3	3,6	4,67	2,93	3,4
D1	1	1	0	0	0	0	0	1	0	0	1	0
D2	0	0	0	1	1	0	0	0	1	1	0	1
D3	0	0	1	0	0	1	1	0	0	0	0	0

Dependent Variable: CONS

Method: Least Squares

Sample: 1 12

Included observations: 12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D1	4.3075	0.813765	5.2933	0.0005
D2	3.754	0.727853	5.157634	0.0006
D3	2.966667	0.939654	3.157189	0.0116
R-squared	0.114523	Mean dependent var	3.741667	
Adjusted R-squared	-0.082249	S.D. dependent var	1.564462	
Log likelihood	-21.14593	Durbin-Watson stat	1.730088	



Chương IV – Hồi qui với biến giả



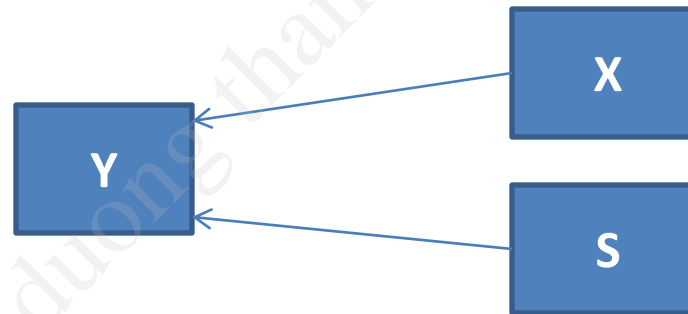
2. Các mô hình có chứa biến giả:

Y
Biến phụ thuộc
Biến lượng

X
Biến độc lập
Biến lượng

S
Biến độc lập
Biến giả
 $1 = A_1$ và $0 = A_2$

2.1. Mô hình điều chỉnh hệ số chặn:



$$PRM : Y_i = \beta_1 + \beta_2 X_i + \beta_3 S_i + U_i$$

$$(A_1) : Y_i = (\beta_1 + \beta_3) + \beta_2 X_i + U_i$$

$$(A_2) : Y_i = \beta_1 + \beta_2 X_i + U_i$$



Chương IV – Hồi qui với biến giả



2. Các mô hình có chứa biến giả:

2.1. Mô hình điều chỉnh hệ số chặn:

Y – điểm số môn bóng rổ (thang 100)

X – chiều cao sinh viên (inch)

S – giới tính (1 = nam, 0 = nữ)

Dependent Variable: Y				
Method: Least Squares				
Sample: 1 10				
Included observations: 10				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-12.732	26.1192	-0.487458	0.6408
X	1.397143	0.396308	3.525396	0.0097
S	6.764571	2.190834	3.087669	0.0176



Chương IV – Hồi qui với biến giả



2. Các mô hình có chứa biến giả:

2.1. Mô hình điều chỉnh hệ số chặn:

Biến phụ thuộc trong 2 trường hợp chênh lệch nhau là β_3 đơn vị. Các phân tích hồi quy sẽ tập trung vào thông tin này.

(?) Viết hồi quy mẫu cho 2 trường hợp điểm số môn bóng rổ của sinh viên nam và nữ

(?) Giới tính có ảnh hưởng đến điểm số môn bóng rổ hay không (khi chiều cao của sinh viên nam và nữ là ngang nhau.

(?) Điểm môn bóng rổ của sinh viên nam và nữ chênh lệch nhau trong khoảng nào

(?) Điểm của học viên nữ thấp hơn của học viên nam trung bình là 6 điểm. Nhận xét ý kiến này



Chương IV – Hồi qui với biến giả



2. Các mô hình có chứa biến giả:

2.1. Mô hình điều chỉnh hệ số chặn:

(*) Chú ý: trường hợp biến định tính có nhiều phạm trù hoặc có nhiều biến định tính cần được đưa vào phân tích trong mô hình, khi đó mô hình sẽ xuất hiện nhiều biến giả.

Các phân tích có thể phải sử dụng đến các kiểm định khác ngoài kiểm định T:

- kiểm định sự phù hợp của hàm hồi qui**
- kiểm định ràng buộc giữa các hệ số hồi qui**
- kiểm định thu hẹp hồi qui**



2. Các mô hình có chứa biến giả:

2.1. Mô hình điều chỉnh hệ số chặn:

(*) Ví dụ: CE – expenditure consumption

$D_1 = 1$ với quý 1, $= 0$ với quý khác

$D_2 = 1$ với quý 2, $= 0$ với quý khác

$D_3 = 1$ với quý 3, $= 0$ với quý khác

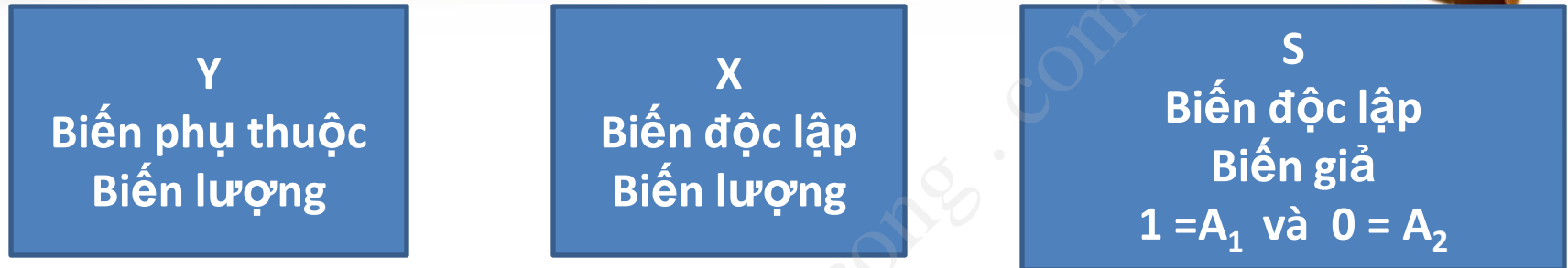
Dependent Variable: CE				
Included observations: 44				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	50421.45	928.3103	54.31531	0.0000
D1	-4624.455	1312.829	-3.522511	0.0011
D2	-3694.273	1312.829	-2.813979	0.0076
D3	-2243.545	1312.829	-1.708940	0.0952
R-squared	0.261033	F-statistic		4.709868
Log likelihood	-413.7583	Prob(F-statistic)		0.006586



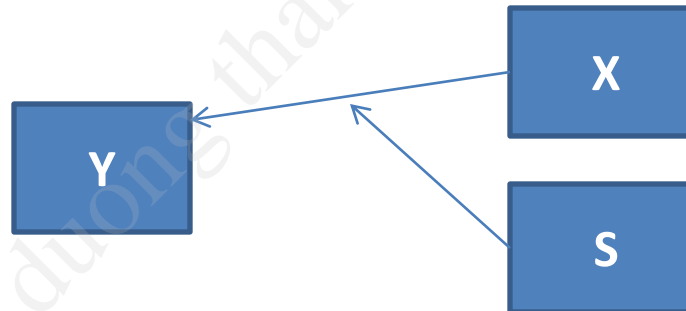
Chương IV – Hồi qui với biến giả



2. Các mô hình có chứa biến giả:



2.2. Mô hình điều chỉnh hệ số góc:



$$PRM : Y_i = \beta_1 + \beta_2 X_i + \beta_3 (S * X)_i + U_i$$

$$(A_1) : Y_i = \beta_1 + (\beta_2 + \beta_3) X_i + U_i$$

$$(A_2) : Y_i = \beta_1 + \beta_2 X_i + U_i$$



Chương IV – Hồi qui với biến giả



2. Các mô hình có chứa biến giả:

2.2. Mô hình điều chỉnh hệ số góc:

Y – thu nhập (1000 USD/năm)

X – số năm đi học (năm)

F – khu vực làm việc (1 = tư nhân, 0 = nhà nước)

Dependent Variable: Y

Method: Least Squares

Sample: 1 10

Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.337530	0.462845	-0.729251	0.4895
X	0.743830	0.036528	20.36304	0.0000
X*F	0.072874	0.024738	2.945851	0.0215



Chương IV – Hồi quy với biến giả



2. Các mô hình có chứa biến giả:

2.2. Mô hình điều chỉnh hệ số góc:

Biến phụ thuộc trong 2 trường hợp chênh lệch nhau là $\beta_3 * X_i$ đơn vị. Các phân tích hồi quy sẽ tập trung vào thông tin này.

(?) Viết hồi quy mẫu cho 2 trường hợp nhân viên làm việc trong khu vực nhà nước và tư nhân

(?) Khu vực làm việc có ảnh hưởng đến thu nhập của nhân viên không?

(?) Đối với khu vực tư nhân, số năm đi học tăng lên 1 thì thu nhập trung bình có xu hướng tăng thêm trong khoảng nào, biết hiệp phương sai của 2 hệ số góc $= 7.62 * 10^{-5}$?



Chương IV – Hồi qui với biến giả



2. Các mô hình có chứa biến giả:

2.3. Mô hình điều chỉnh cả 2 hệ số:

$$PRM : Y_i = \beta_1 + \beta_2 X_i + \beta_3 S_i + \beta_4 (S * X)_i + U_i$$

$$(A_1) : Y_i = (\beta_1 + \beta_3) + (\beta_2 + \beta_4) X_i + U_i$$

$$(A_2) : Y_i = \beta_1 + \beta_2 X_i + U_i$$

Ví dụ:

Saving – tiết kiệm người dân Mỹ từ 1970 – 1995

Income – thu nhập người dân Mỹ cùng thời kỳ

Dum – biến giả (0 = trước năm 1982, 1 = từ 1982)



Chương IV – Hồi qui với biến giả



Dependent Variable: SAVING

Method: Least Squares

Sample: 1970 1995

Included observations: 26

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.022197	20.16398	0.050694	0.9600
INCOME	0.080328	0.014496	5.541293	0.0000
DUM	152.4725	33.08197	4.608930	0.0001
DUM*INCOME	-0.065466	0.015982	-4.096229	0.0005
R-squared	0.881943	Mean dependent var		162.0885
Adjusted R-squared	0.865844	S.D. dependent var		63.20446
Log likelihood	-116.4127	F-statistic		54.78344
Durbin-Watson stat	1.648475	Prob(F-statistic)		0.000000



Chương IV – Hồi quy với biến giả



2. Các mô hình có chứa biến giả:

2.2. Mô hình điều chỉnh hệ số góc:

Biến phụ thuộc trong 2 trường hợp chênh lệch nhau là $\beta_3 * X_i$ đơn vị. Các phân tích hồi quy sẽ tập trung vào thông tin này.

(?) Viết hồi quy mẫu cho 2 trường hợp nhân viên làm việc trong khu vực nhà nước và tư nhân

(?) Khu vực làm việc có ảnh hưởng đến thu nhập của nhân viên không?

(?) Đối với khu vực tư nhân, số năm đi học tăng lên 1 thì thu nhập trung bình có xu hướng tăng thêm trong khoảng nào, biết hiệp phương sai của 2 hệ số góc $= 7.62 * 10^{-5}$?



Chương IV – Hồi qui với biến giả



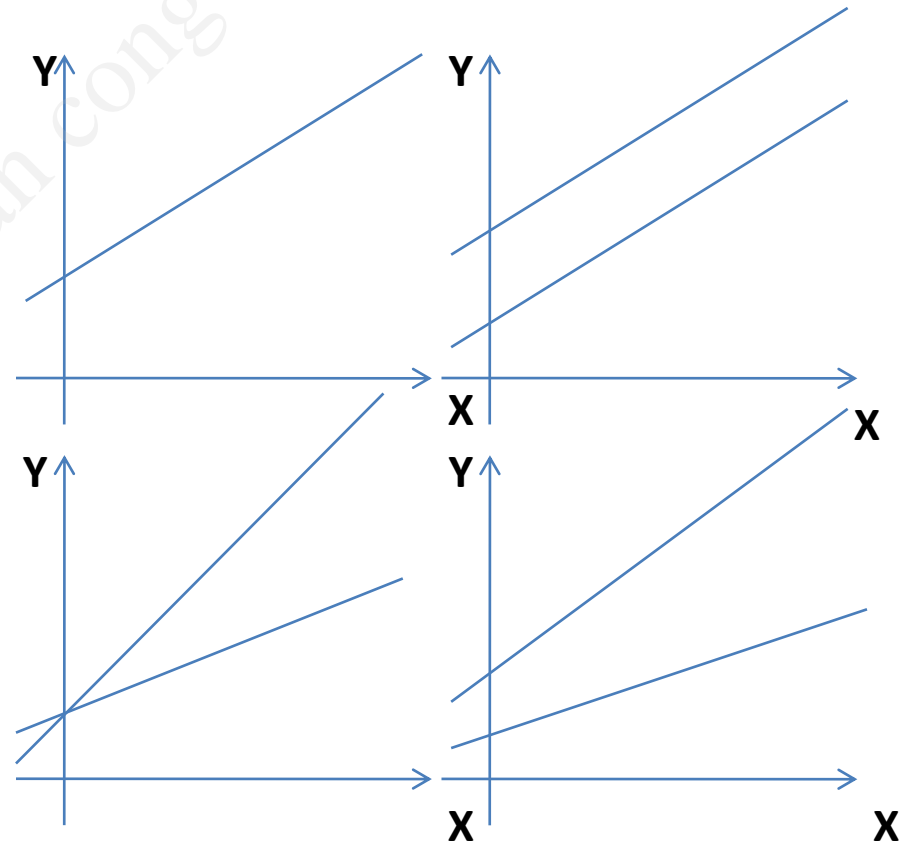
3. Thủ tục so sánh 2 hồi qui

3.1. Các trường hợp khác biệt giữa 2 hồi qui:

Trên thực tế, khi thời gian và môi trường thay đổi, quan hệ giữa các biến số

trong mô hình có thể thay đổi, ví dụ: quan hệ phụ thuộc của chi tiêu vào thu nhập trước và sau thời kỳ đổi mới, ...

→ Nhu cầu so sánh 2 hồi qui





Chương IV – Hồi qui với biến giả



3. Thủ tục so sánh 2 hồi qui

3.2. Kỹ thuật biến giả:

$$Dum = \begin{cases} 1 & \text{với } n_1 \text{ quan sát} \\ 0 & \text{với } n_2 \text{ quan sát} \end{cases}$$

Sử dụng chính mô hình điều chỉnh cả 2 hệ số để phân

tích:
$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 Dum_i + \beta_4 (Dum * X)_i + U_i$$

$$(sample - n_1) : Y_i = (\beta_1 + \beta_3) + (\beta_2 + \beta_4) X_i + U_i$$

$$(sample - n_2) : Y_i = \beta_1 + \beta_2 X_i + U_i$$

Ưu điểm: cho biết hai hồi qui khác nhau ở đâu, hệ số góc hay hệ số chặn hoặc cả hai hệ số.

Nhược điểm: mô hình sẽ phức tạp khi có nhiều hệ số góc



Chương IV – Hồi qui với biến giả



3. Thủ tục so sánh 2 hồi qui

3.3. Kiểm định CHOW:

$$Y_i = \beta_1 + \beta_2 X_i + U_i \quad (1)$$

B1: hồi qui (1) với n_1 quan sát $\rightarrow RSS_1$

B2: hồi qui (1) với n_2 quan sát $\rightarrow RSS_2$

Tính $\overline{RSS} = RSS_1 + RSS_2$

B3: hồi qui (1) với $n = n_1 + n_2 \rightarrow RSS$

B4: Kiểm định: $\begin{cases} H_0 : 2 \text{ hồi qui như nhau} \\ H_1 : 2 \text{ hồi qui khác nhau} \end{cases}$

$$F_{qs} = \frac{(\overline{RSS} - RSS) / k}{RSS / (n - 2k)}$$

Miền bác bỏ H_0

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

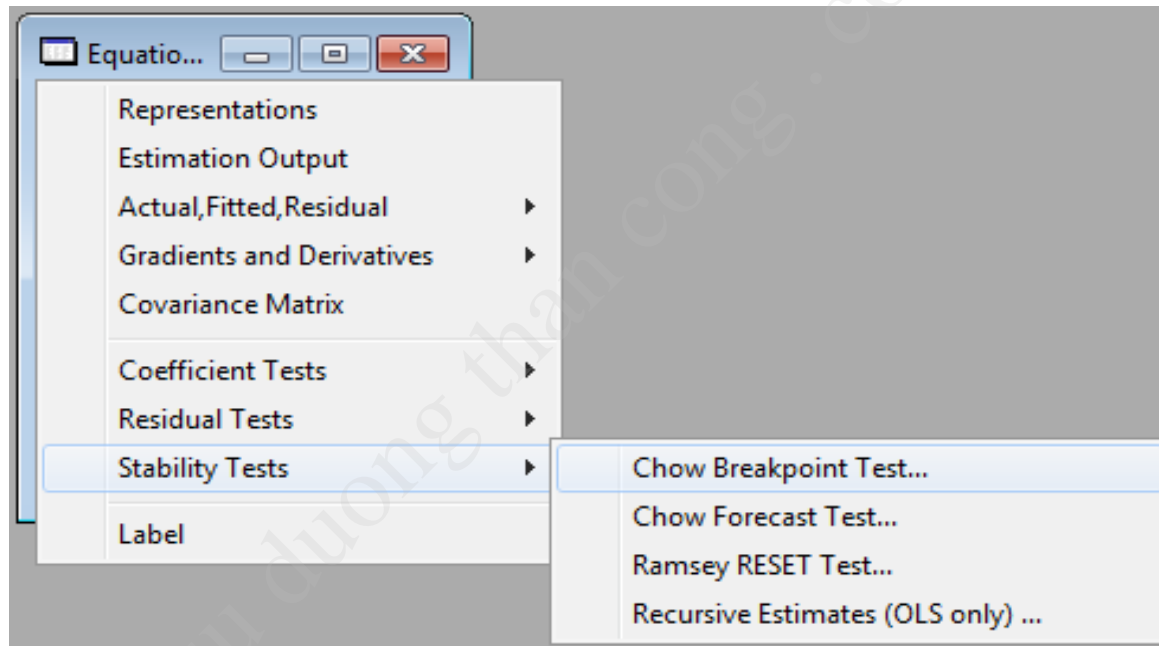


Chương IV – Hồi qui với biến giả



3. Thủ tục so sánh 2 hồi qui

3.3. Kiểm định CHOW:



CuuDuongThanCong.com

Chow Breakpoint Test: 1982

F-statistic	10.68948	Probability	0.000571
Log likelihood ratio	17.65224	Probability	0.000147