

Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2010

Các Phương Pháp Phân Tích Định Lượng

Gợi ý lời giải - Bài tập 9

HỒI QUI TUYẾN TÍNH ĐA BIẾN: KIỂM ĐỊNH GIẢ THUYẾT VÀ LỰA CHỌN MÔ HÌNH

Ngày Phát: Thứ Hai, 13/12/2010

Ngày Nộp: **8:20 sáng, Thứ Hai, 20/12/2010**

Bản in nộp tại Phòng Giáo Vụ

Bản điện tử gửi đến thầy Nguyễn Khánh Duy theo địa
chỉ duyнк@fets.vnn.vn

Bài 1

Bạn có thể sử dụng dữ liệu của Bảng 7.3 trong bộ dữ liệu của Gujarati về tổng sản lượng thực (Y , triệu NT\$ hay triệu TWD : Đô la mới của Đài Loan), ngày lao động (X_2 , triệu ngày) và nhập lượng vốn thực (X_3 , triệu NT\$) ở khu vực nông nghiệp; hoặc dữ liệu ở bài tập 7.18 ở khu vực công nghiệp trong Gujarati để làm bài này.

Có 2 hàm hồi quy

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \quad (1)$$

$$\ln Y_t = \alpha_1 + \alpha_2 \ln X_{2t} + \alpha_3 \ln X_{3t} + u_t \quad (2)$$

a. $\hat{\beta}_2$ và $\hat{\beta}_3$ có ý nghĩa thống kê riêng biệt không

Do P-value của $\hat{\beta}_2$ và $\hat{\beta}_3$ đều < 0.05 nên mỗi hệ số này đều có ý nghĩa thống kê riêng biệt tại mức ý nghĩa 5%

b. Chúng có khác 1 về mặt thống kê không?

+ **Đối với hệ số hồi quy riêng β_2 :**

Giả thuyết: $H_0: \beta_2 = 1$

$H_1: \beta_2 \neq 1$

Với mức ý nghĩa $\alpha = 5\%$ ta tính được giá trị tới hạn của kiểm định hai phía dựa trên phân phối t với $(n-k) = 15 - 3 = 12$ bậc tự do là: $t_{\alpha/2, n-k} = t_{0,025, 12} = 2.179$.

Trị thống kê kiểm định:

$$t_{\hat{\beta}_2} = (147.936 - 1) / 36.443 = 4.032$$

Vì $|t_{\hat{\beta}_2}| = 4.032 (> 2.179)$ nên ta bác bỏ H_0 . Do đó, β_2 khác 1 về mặt thống kê (với mức ý nghĩa thống kê 5%)

+ **Đối với hệ số hồi quy riêng β_3 :**

Giả thuyết: $H_0: \beta_3 = 1$

$H_1: \beta_3 \neq 1$

Trị thống kê kiểm định:

$$t_{\hat{\beta}_3} = (0.404 - 1) / 0.074 = -8.054$$

Với mức ý nghĩa $\alpha = 5\%$ ta tính được giá trị tới hạn của kiểm định hai phía dựa trên phân phối t với $(n-k) = 15 - 3 = 12$ bậc tự do là: $t_{\alpha/2, n-k} = t_{0.025, 12} = 2.179$

Vì $|t_{\hat{\beta}_3}| = 8.054 (> 2.179)$ nên bác bỏ H_0 . Do đó, β_3 khác 1 về mặt thống kê, với mức ý nghĩa thống kê 5%.

Mô hình 1

Dependent Variable: Y
 Method: Least Squares
 Date: 01/05/11 Time: 10:23
 Sample: 1958 1972
 Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-28067.170	9432.066	-2.976	0.012
X2	147.936	36.443	4.059	0.002
X3	0.404	0.074	5.486	0.000
R-squared	0.910	Mean dependent var	24735.330	
Adjusted R-squared	0.894	S.D. dependent var	4874.173	
S.E. of regression	1583.279	Akaike info criterion	17.749	
Sum squared resid	30081287.000	Schwarz criterion	17.891	
Log likelihood	-130.119	Hannan-Quinn criter.	17.748	
F-statistic	60.341	Durbin-Watson stat	1.039	
Prob(F-statistic)	0.000			

c. $\hat{\alpha}_2$ và $\hat{\alpha}_3$ có ý nghĩa thống kê riêng biệt không?

Mô hình 2

Dependent Variable: LOG(Y)

Method: Least Squares

Date: 01/05/11 Time: 10:27

Sample: 1958 1972

Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.338	2.450	-1.363	0.198
LOG(X2)	1.499	0.540	2.777	0.017
LOG(X3)	0.490	0.102	4.800	0.000
R-squared	0.889	Mean dependent var		10.097
Adjusted R-squared	0.871	S.D. dependent var		0.208
S.E. of regression	0.075	Akaike info criterion		-2.171
Sum squared resid	0.067	Schwarz criterion		-2.029
Log likelihood	19.282	Hannan-Quinn criter.		-2.172
F-statistic	48.069	Durbin-Watson stat		0.891
Prob(F-statistic)	0.000			

Do P-value($\hat{\alpha}_2$) = 0.017 (<0.05) và P-value($\hat{\alpha}_3$) = 0.000 (<0.05) nên $\hat{\alpha}_2$ và $\hat{\alpha}_3$ có ý nghĩa thống kê riêng biệt

d) Kết quả ước lượng $\hat{\alpha}_2$ và $\hat{\alpha}_3$ cho phép chúng ta kết luận gì về sản xuất nông nghiệp của Đài Loan giai đoạn 1958 – 1972:

- Trong khu vực sản xuất nông nghiệp của Đài Loan giai đoạn 1958 – 1972, yếu tố lao động và vốn thực có mối quan hệ đồng biến với tổng sản lượng thực, cụ thể:

+ Trong điều kiện các yếu tố khác không đổi, khi số ngày công lao động tăng lên 1% thì tổng sản lượng thực trong khu vực sản xuất nông nghiệp của Đài Loan giai đoạn 1958 – 1972 tăng lên 1.499% và ngược lại.

+ Trong điều kiện các yếu tố khác không đổi, khi vốn thực tăng lên 1% thì tổng sản lượng thực trong khu vực sản xuất nông nghiệp của Đài Loan giai đoạn 1958 – 1972 tăng lên 0.49% và ngược lại.

- Đồng thời, $\hat{\alpha}_2 + \hat{\alpha}_3 = 1.499 + 0.490 = 1.989 (>1)$ cho thấy khu vực sản xuất nông nghiệp của Đài Loan giai đoạn 1958 – 1972 có năng suất tăng dần theo quy mô. (Bạn có thể thực hiện thêm kiểm định Wald để kết luận được cẩn trọng hơn)

e) Dữ liệu có ủng hộ giả thuyết cho rằng $\beta_2 = \beta_3 = 0$ hay không?

Để biết được dữ liệu có ủng hộ giả thuyết cho rằng $\beta_2 = \beta_3 = 0$ hay không ta sử dụng phương pháp kiểm định F (Wald):

Kiểm định giả thuyết: $H_0: \beta_2 = \beta_3 = 0$

H_1 : Ít nhất có một tham số β_2 hoặc $\beta_3 \neq 0$

Wald Test:
 Equation: EQ01

Test Statistic	Value	df	Probability
F-statistic	60.34143	(2, 12)	0.0000
Chi-square	120.6829	2	0.0000

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(2)	147.9362	36.44344
C(3)	0.403563	0.073561

Restrictions are linear in coefficients.

Từ bảng kết quả ta có : $P_{\text{value}}(F) = 0.000 (< 0.05)$ nên Bác bỏ H_0 .

Như vậy, dữ liệu nghiên cứu không ủng hộ giả thuyết cho rằng $\beta_2 = \beta_3 = 0$, mà có ít nhất một trong 2 tham số này khác 0 ($\beta_2 \neq 0$ hoặc $\beta_3 \neq 0$).

f) Hãy kiểm định giả thuyết cho rằng $\alpha_2 = \alpha_3 = 0$

Kiểm định giả thuyết: $H_0: \alpha_2 = \alpha_3 = 0$

H_1 : Ít nhất có một tham số α_2 hoặc $\alpha_3 \neq 0$

Wald Test:
 Equation: EQ02

Test Statistic	Value	df	Probability
F-statistic	48.06885	(2, 12)	0.0000
Chi-square	96.13770	2	0.0000

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(2)	1.498767	0.539803
C(3)	0.489858	0.102043

Restrictions are linear in coefficients.

Từ bảng kết quả ta có : $P_{\text{value}}(F) = 0.000 (< 0.05) \Rightarrow$ Bác bỏ H_0 .

Như vậy, ở độ tin cậy 95%, bác bỏ giả thuyết cho rằng $\alpha_2 = \alpha_3 = 0$. Hay nói cách khác: có ít nhất một trong 2 tham số α_2 hoặc $\alpha_3 \neq 0$

g) Bạn thích chọn mô hình nào? Vì sao?

Mô hình (1) có các tham số ước lượng $\hat{\beta}_2, \hat{\beta}_3$ phản ánh tác động biên của biến lao động (X_2) và biến vốn thực (X_3) lên tổng sản lượng thực Y bằng những con số mang tính tuyệt đối.

Còn mô hình (2) có dạng hàm Log -Log, các tham số ước lượng $\hat{\alpha}_2$ và $\hat{\alpha}_3$ phản ánh độ co giãn của tổng sản lượng thực Y theo lao động (X_2) và theo vốn thực (X_3). Mô hình này được hình thành từ việc lấy ln hai vế của hàm sản xuất dạng Cobb-Douglas (thường sử dụng khi ước lượng hàm sản xuất).

Vì vậy, chúng ta thường thích chọn mô hình (2) hơn.

h) Hãy so sánh các giá trị R^2 của hai mô hình

R^2 của mô hình 1 = 0.910

Do mô hình 2 có biến phụ thuộc là lnY, nên ta cần tính R^2 giả để có thể so sánh được với R^2 của mô hình 1.

R^2 giả của mô hình 2 = $0.955114^2 = 0.9124$ ($> R^2$ của mô hình 1), nên mô hình 2 tốt hơn mô hình 1.

Bài 2

Bảng 8.32 trong bộ số liệu của Gujarati cho dữ liệu được sử dụng bởi các nhà sản xuất cáp điện thoại để dự báo mức bán cho khách hàng chính trong thời kỳ 1968-1983.

Các biến trong bảng được xác định như sau:

Y = mức bán hàng năm trong MPF, tính bằng triệu feet đôi

X_2 = tổng sản phẩm quốc gia (GNP), tỷ đô la

X_3 = số lượng nhà xây mới, ngàn căn nhà

X_4 = tỉ lệ thất nghiệp, phần trăm

X_5 = lãi suất gốc² của 6 tháng trước

X_6 = khả năng truy cập internet nhanh hơn, %

Câu a: Hãy ước lượng hồi qui $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + \beta_6 X_{6t} + u_t$?

Mô hình U

Dependent Variable: Y

Method: Least Squares

Date: 01/05/11 Time: 12:09

Sample (adjusted): 1 16

Included observations: 16 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5962.656	2507.724	2.378	0.039
X2	4.884	2.513	1.944	0.081
X3	2.364	0.844	2.802	0.019
X4	-819.129	187.707	-4.364	0.001
X5	12.010	147.050	0.082	0.937
X6	-851.393	292.145	-2.914	0.016
R-squared	0.823	Mean dependent var	7543.125	
Adjusted R-squared	0.734	S.D. dependent var	1217.152	
S.E. of regression	627.601	Akaike info criterion	16.002	

Sum squared resid	3938824.000	Schwarz criterion	16.291
Log likelihood	-122.013	Hannan-Quinn criter.	16.017
F-statistic	9.284	Durbin-Watson stat	2.484
Prob(F-statistic)	0.002		

Mô hình hồi quy ước lượng là:

$$\text{SRF: } Y_t = b_1 + b_2 X_{2t} + b_3 X_{3t} + b_4 X_{4t} + b_5 X_{5t} + b_6 X_{6t} + \hat{u}_t$$

$$\text{SRF: } Y_t = 5962.656 + 4.884 X_{2t} + 2.364 X_{3t} - 819.129 X_{4t} + 12.010 X_{5t} - 851.393 X_{6t} + \hat{u}_t$$

Câu b: Dấu kỳ vọng của các hệ số của mô hình này như thế nào?

Bạn có thể kỳ vọng dấu của các hệ số b trong mô hình là:

b₂: dương; khi GNP tăng, mức bán hàng năm trong MPF tăng.

b₃: dương; khi số lượng nhà dân tăng lên, mức bán hàng năm trong MPF tăng.

b₄: âm; khi tỷ lệ thất nghiệp tăng, mức bán hàng năm trong MPF giảm.

b₅: âm; khi lãi suất gốc của 6 tháng trước (lãi suất cho vay thấp nhất dành cho khách) càng cao, mức bán hàng năm trong MPF càng thấp.

b₆: dương. Kỳ vọng khi khả năng truy cập Internet nhanh hơn, người tiêu dùng sẽ thích dùng internet hơn nên nhu cầu mua cáp tăng.

(Tùy theo lập luận của bạn, những kiến thức, nghiên cứu thực nghiệm mà bạn đã đọc ... kỳ vọng về dấu có thể khác)

Câu c: Các kết quả thực nghiệm có tuân theo các kỳ vọng tiên nghiệm không?

Theo kết quả thực nghiệm ta thấy các hệ số b₂, b₃, b₄ đều tuân theo các kỳ vọng tiên nghiệm, b₅, b₆ là trái với kỳ vọng

Câu d: Các hệ số hồi qui riêng ước lượng được có ý nghĩa thống kê riêng biệt tại mức ý nghĩa 5% không?

Ta thấy các hệ số ước lượng b₃, b₄, b₆ đều có P_{value} nhỏ hơn 0.05 nên mỗi quan hệ giữa biến Y và các biến X₃, X₄, X₆ có ý nghĩa thống kê tại mức ý nghĩa 5%.

Trong khi đó, các hệ số ước lượng b₂, b₅ có P_{value} lớn hơn 0.05 nên mỗi quan hệ giữa biến Y và các biến X₂, X₅ không có ý nghĩa thống kê tại mức ý nghĩa 5%.

Câu e: Giả sử rằng: đầu tiên bạn chỉ hồi qui Y trên X₂, X₃ và X₄, sau đó bạn quyết định thêm các biến X₅ và X₆. Bạn làm thế nào để tìm ra rằng việc thêm các biến X₅ và X₆ có đáng hay không? Bạn sử dụng phép kiểm định nào? Trình bày các tính toán cần thiết.

Mô hình ở câu a là mô hình không giới hạn (Mô hình U)

Bạn ước lượng mô hình giới hạn (Mô hình R) như sau:

Mô hình R

Dependent Variable: Y

Method: Least Squares

Date: 01/05/11 Time: 12:34

Sample (adjusted): 1 16

Included observations: 16 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	195.604	2308.317	0.085	0.934
X2	6.208	1.906	3.258	0.007
X3	1.495	0.625	2.391	0.034
X4	-469.718	198.570	-2.366	0.036
R-squared	0.601	Mean dependent var		7543.125
Adjusted R-squared	0.502	S.D. dependent var		1217.152
S.E. of regression	859.306	Akaike info criterion		16.562
Sum squared resid	8860877.000	Schwarz criterion		16.756
Log likelihood	-128.500	Hannan-Quinn criter.		16.572
F-statistic	6.031	Durbin-Watson stat		1.581
Prob(F-statistic)	0.010			

Mô hình U: $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + \beta_6 X_{6t} + u_t$ Mô hình U

Mô hình R: $Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \alpha_4 X_{4t} + v_t$ Mô hình R

$H_0: \beta_5 = \beta_6 = 0$

H_1 : Ít nhất một trong hai hệ số β_5 hoặc β_6 khác 0

$$F = \frac{(R_U^2 - R_R^2) / (k - m)}{(1 - R_U^2) / (n - k)} = \frac{(0.8227 - 0.6012) / 2}{(1 - 0.8227) / 10} = 6.25$$

$$F_{\alpha, k-1, n-k} = F_{0.05, 2, 10} = 4.1028$$

Do $F > F_{0.05, 2, 10}$ nên bác bỏ giả thuyết H_0 cho rằng cả 2 biến X_5, X_6 đều không ảnh hưởng đến Y. Như vậy, trong mô hình R, nên bổ sung thêm ít nhất một trong hai biến X_5 và X_6