

Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2012

Các Phương Pháp Phân Tích Định Lượng

LỜI GIẢI ĐỀ NGHỊ BÀI TẬP 8

HỎI QUY ĐA BIẾN

Ngày Phát: Thứ hai 03/12/2012

Ngày Nộp: Thứ hai 10/12/2012

Bản in nộp lúc **8h20 sáng**, tại Hộp nộp bài tập trong phòng Lab

Bản điện tử gửi lên <http://intranet.fetp.edu.vn:81>

Câu 1

Dependent Variable: Y
Method: Least Squares
Date: 11/10/05 Time: 22:56
Sample: 1 200
Included observations: 200

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1689.783	518.2858	-3.260331	A
X2	0.007695	0.003576	2.152021	B
X3	79.76524	49.25838	1.619323	C
X2*X3	74.16483	123.7128	0.599492	D
X3*X3	37.87703	8.552994	4.428511	E
D1	746.9610	297.3025	2.512462	F
R-squared	0.172804		Mean dependent var	1253.115
Adjusted R-squared	G		S.D. dependent var	1764.184
S.E. of regression	1625.076		Akaike info criterion	17.65404
Sum squared resid	H		Schwarz criterion	17.75299
Log likelihood	-1759.404		F-statistic	8.105433
Durbin-Watson stat	1.855648		Prob(F-statistic)	0.000001

Dựa vào bảng trên trả lời các câu hỏi sau:

a. Mô hình được trình bày ở trên có bao nhiêu quan sát, bao nhiêu biến?

Mô hình có:

+ 200 quan sát.

+ 5 biến độc lập và 1 biến phụ thuộc.

b. Mô hình này có vi phạm giả thiết về tính đa cộng tuyến của các biến độc lập hay không? Vì sao?

Một trong những phương pháp để nhận biết mô hình có thể vi phạm giả thiết về đa cộng tuyến là dựa vào kết quả hồi qui với giá trị R-squared, theo đó nếu R-squared lớn (khoảng trên 90%) thì ta có thể cho rằng mô hình có khả năng vi phạm giả thiết về hiện tượng đa cộng tuyến. Với kết quả hồi qui này thì R-squared xấp xỉ 17.28% ta có thể cho rằng mô hình trên không vi phạm giả thiết về đa cộng tuyến.

c. Hãy điền các số liệu còn thiếu (A,B,C,D, E, F, G, H) vào bảng trên đây, trình bày cách tính toán của bạn?

Ta có bảng kết quả sau khi tính toán các giá trị còn thiếu như sau:

Dependent Variable: Y

Method: Least Squares

Date: 11/10/05 Time: 22:56

Sample: 1 200

Included observations: 200

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1689.783	518.2858	-3.260331	A = 0.0013
X2	0.007695	0.003576	2.152021	B = 0.032631
X3	79.76524	49.25838	1.619323	C = 0.107
X2*X3	74.16483	123.7128	0.599492	D = 0.5495
X3*X3	37.87703	8.552994	4.428511	E = 0.0000158
D1	746.9610	297.3025	2.512462	F = 0.0128
R-squared	0.172804		Mean dependent var	1253.115
Adjusted R-squared	G = 0.151485		S.D. dependent var	1764.184
S.E. of regression	1625.076		Akaike info criterion	17.65404
Sum squared resid	H = 512329169.12		Schwarz criterion	17.75299
Log likelihood	-1759.404		F-statistic	8.105433
Durbin-Watson stat	1.855648		Prob(F-statistic)	0.000001

Trình bày các tính toán:

Sử dụng hàm tính excel TDIST(ABS(t-statistic(i)), 194, 2), ta có:

- $A = 0.001314365$
- $B = 0.032631233$
- $C = 0.107002265$
- $D = 0.549544334$
- $E = 1.58301E-05$
- $F = 0.012803976$

- Tính G: Giá trị R^2 hiệu chỉnh theo bậc tự do (\bar{R}^2) được tính theo công thức sau

$$\text{○ } G = \bar{R}^2 = 1 - \frac{n-1}{n-k} (1 - R^2) = 1 - \frac{200-1}{200-6} (1 - 0.172804) = 0.151484515$$

- Tính H: Ta có công thức tính ước lượng phương sai sai số (σ^2) được tính bằng

$$s^2 = \hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-k}, \text{ do đó ta có tổng bình phương sai số (sum squared resid):}$$

$$\text{○ } H = \sum \hat{u}_i^2 = \hat{\sigma}^2 (n-k) = 1625.067^2 * (200-6) = 512,329,169.1205 \text{ (đơn vị bình phương)}$$

d. Viết phương trình hồi quy mẫu?

Từ kết quả hồi quy đã cho, ta viết được phương trình hồi quy mẫu như sau:

$$\hat{Y}_i = -1689.783 + 0.007695X_{2i} + 79.76524X_{3i} + 74.16483(X_{2i} * X_{3i}) + 37.87703X_{3i}^2 + 746.9610D_{1i}$$

e. Bao nhiêu phần trăm sự biến thiên của Y được giải thích bởi các biến độc lập nói trên?

Với $R^2 = 17,28\%$ ta có thể cho rằng với các điều kiện các yếu tố khác không đổi, các biến độc lập có trong mô hình hồi quy trên giải thích được khoảng 17,28% sự biến thiên về mặt trung bình của biến Y.

f. Tính toán tác động biên tế của X2 lên Y? Tính toán tác động biên tế của X3 lên Y? Tính toán tác động biên tế của D1 lên Y? Bạn có nhận xét gì về đặc trưng của các tác động biên của 3 biến nêu trên?

Tác động biên tế của các biến X2, X3 và D1 lên Y chính là các hệ số bê ta ước lượng của các biến tương ứng hay chính là đạo hàm bậc nhất của Y theo các biến đó. Như vậy, ta có các tính toán sau:

$$\text{- } \hat{\beta}_2 = \frac{\partial \hat{Y}_i}{\partial X_2} = 0.007695 + 74.16483X_{3i}$$

$$- \hat{\beta}_3 = \frac{\partial \hat{Y}_i}{\partial X_3} = 79.76524 + 74.16483X_{2i} + 75.75406X_{3i}$$

$$- \hat{\beta}_6 = \frac{\partial \hat{Y}_i}{\partial D_{1i}} = 746,96$$

Nhận xét đặc trưng: Tác động biên tế của các biến X2, X3, D1:

- + *Biên tế của D1 là một hằng số không đổi*
- + *Biên tế của X2 là một hàm tuyến tính theo X3*
- + *Biên tế của X3 là một hàm tuyến tính đa biến theo X2 và X3*

Tổng quát: Nếu trong mô hình gồm biến tương tác được tạo bởi các biến độc lập với nhau thì tác động biên tế của chính biến độc lập đó lên biến phụ thuộc sẽ là một hàm số theo chính các biến độc lập đã tạo nên biến tương tác. Tác động biên tế của biến độc lập tại một giá trị cụ thể của X_i là khác nhau khi X_i thay đổi và thay đổi theo dạng hàm số đã được xác định.

g. Nếu được phép thay đổi mô hình nói trên thì bạn đề nghị mô hình mới bao gồm các biến nào? Giải thích vì sao bạn lại đề nghị như vậy?

Ta thấy, nếu lấy mức ý nghĩa 5% để đánh giá khả năng bác bỏ giả thuyết các hệ số hồi quy bằng không, ta sẽ loại bỏ các biến có giá trị P_{value} lớn hơn 5% (0.05). Do đó có thể loại bỏ các biến:

- o *X3, có $P\text{-value} = 0.107 > 0.05$*
- o *$X2*X3$ có $P\text{-value} = 0.5495 > 0.05$*

Như vậy mô hình mới sẽ bao gồm các biến:

- *Biến phụ thuộc: Y*
- *Biến độc lập: X2, X3*X3, D1*

Câu 2:

File đính kèm bài tập này (chi_tieu_y_te_thu_nhap.xls) là dữ liệu từ 65 quan sát tại Tp.Hồ Chí Minh. Sử dụng Eview để làm các công việc sau:

a. Lập bảng thống kê mô tả cho tất cả các biến có trong tập dữ liệu nói trên. Phân loại các biến.

Sử dụng Eview, Group Views/Descriptive Statistic ta có bảng thống kê mô tả như sau:

	CHI_TIEU_Y TE	TUOI_CUA_C HU_HO	TRE_EM_DU OI_15_TUOI	THU_NHAP	SO_NHAN_K HAU	PHU_NU_TR ONG_TUOI_S INH	GIOL_TINH_C HU_HO	BANG_CAP_ CAO_NHAT
Mean	1171.921	51.60317	0.936508	45361.98	5.269841	1.301587	1.571429	1.777778
Median	500.0000	48.00000	1.000000	36001.00	4.000000	1.000000	2.000000	1.000000

Maximum	10530.00	87.00000	4.000000	165875.0	19.00000	5.000000	2.000000	7.000000
Minimum	0.000000	18.00000	0.000000	101.0000	1.000000	0.000000	1.000000	0.000000
Std. Dev.	1925.728	14.77087	1.075730	39142.42	3.755453	1.239601	0.498847	2.003581
Skewness	3.166478	0.233275	0.988476	1.095942	1.605422	1.409027	-0.288675	1.230287
Kurtosis	13.91185	2.283607	3.298211	3.665004	5.337275	4.868147	1.083333	3.855619
Jarque-Bera	417.8342	1.918583	10.49284	13.77229	41.40249	30.00743	10.51823	17.81458
Probability	0.000000	0.383164	0.005266	0.001022	0.000000	0.000000	0.005200	0.000135
Sum	73831.00	3251.000	59.00000	2857805.	332.0000	82.00000	99.00000	112.0000
Sum Sq. Dev.	2.30E+08	13527.08	71.74603	9.50E+10	874.4127	95.26984	15.42857	248.8889
Observations	63	63	63	63	63	63	63	63

Mô hình trên với biến Y (tổng chi tiêu cho y tế của hộ) là biến phụ thuộc và các biến còn lại là biến giải thích, ta có kết quả phân loại như sau:

Định lượng gồm biến: Y, X1, X3, X5, X6 và X7

Định tính gồm biến: X2 và X4

b. Lập bảng hệ số tương quan giữa các biến trong tập dữ liệu nói trên. Bạn có nhận xét gì về mối tương quan giữa các biến này?

Bảng hệ số tương quan:

Chi tiêu	Tổng chi tiêu cho y tế của hộ Y (1000đ/năm)	Thu nhập X1 (1000đ/năm)	Bằng cấp cao nhất đạt được X2	Tuổi của chủ hộ X3	Giới tính của chủ hộ X4	Trẻ em dưới 15 tuổi X5	Phụ nữ trong độ tuổi đẻ X6	Số nhân khẩu X7
Y (1000đ/năm)	1.0000							
X1 (1000đ/năm)	0.3919	1.0000						
X2	0.0929	0.3978	1.0000					
X3	0.3886	-0.0197	-0.3156	1.0000				
X4	-0.0022	0.0679	-0.0583	0.0682	1.0000			
X5	0.1143	-0.0721	0.1332	0.0155	-0.0351	1.0000		
X6	-0.1387	-0.0287	-0.1008	0.0861	-0.0435	0.1235	1.0000	
X7	0.0233	0.0794	0.1931	-0.1271	-0.1961	0.4004	0.3392	1.0000

Nhận xét: Nhìn vào bảng trên, ta cũng thấy mối quan hệ giữa các biến trong tập dữ liệu hầu như là mối quan hệ tương quan trung bình và thấp, có một số mối quan hệ tương quan quá thấp, không đáng kể.

Có những mối quan hệ là đồng biến, một số là nghịch biến giữa các biến độc lập và biến phụ thuộc. Cụ thể:

+ Các biến X1, X2, X3, X5 và X7 có tương quan đồng biến với Y.

+ Các biến X4, X6 tương quan nghịch biến với Y.

Ngoài ra ta cũng có thể tham chiếu để xem xét mức độ tương quan giữa các biến theo bảng sau:

Hệ số tương quan	Ý nghĩa
±0.01 đến ±0.1	Mối tương quan quá thấp, không đáng kể
±0.2 đến ±0.3	Mối tương quan thấp
±0.4 đến ±0.5	Mối tương quan trung bình
±0.6 đến ±0.7	Mối tương quan cao
±0.8 – 1.0	Mối tương quan rất cao

c. Trong tập dữ liệu nói trên, nếu dùng các biến còn lại để giải thích cho sự biến động của “tổng chi tiêu y tế của hộ” thì các bạn có kỳ vọng gì về dấu của các hệ số hồi quy?

Dựa trên các lý thuyết và nghiên cứu trước cũng như những quan sát thực tế ta có thể kỳ vọng về dấu của các hệ số hồi quy của các biến độc lập như sau:

Biến	Kỳ vọng về dấu của hệ số hồi quy
BANGCAONHAT	+
GIOITINHCHUHO	-
PHUNUTRONGTUOISINHSAN	-
QUIMOHOTREEMDUOI15	+
SONHANKHAU	+
THUNHAP	+
TUOICUACHUHO	+

Nhận xét:

+ Các biến “Tuổi của chủ hộ”, “Trẻ em dưới 15 tuổi”, “Thu nhập”, “Số nhân khẩu” và “Bằng cấp cao nhất” sẽ có hệ số hồi quy mang dấu dương, cùng với mối quan hệ đồng biến của các biến này với biến “Tổng chi tiêu y tế của hộ”.

+ Các biến “Phụ nữ trong tuổi sinh đẻ”, “Giới tính của chủ hộ” sẽ có hệ số hồi quy mang dấu âm, cùng với mối quan hệ nghịch biến của các biến này với biến “Tổng chi tiêu y tế của hộ”.

d. Xây dựng hàm hồi quy tuyến tính của mỗi quan hệ trong câu (c), dấu của các hệ số hồi quy có đúng như kỳ vọng của bạn không?

Sử dụng Eview ta có kết quả hồi quy như sau:

Dependent Variable: TONG_CHI_TIEU_Y_TE_CUA_H
 Method: Least Squares
 Date: 01/03/13 Time: 15:19
 Sample: 1 65
 Included observations: 63

Variable	Coefficient	Std. Error	t-Statistic	Prob.
THU_NHAP_X1__1000D_NAM_	0.018986	0.005865	3.237059	0.0020
BANG_CAP_CAO_NHAT_DAT_DU	35.27848	123.1935	0.286366	0.7757
TUOI_CUA_CHU_HO_X3	58.04288	15.05469	3.855469	0.0003
GIOI_TINH_CUA_CHU_HO_X4	-246.4273	423.5130	-0.581865	0.5630
TRE_EM_DUOI_15_TUOI_X5	222.8695	213.1087	1.045802	0.3002
PHU_NU_TRONG_DO_TUOI_SIN	-298.4548	179.7214	-1.660652	0.1025
SO_NHAN_KHAU_X7	18.87725	66.08648	0.285645	0.7762
C	-2279.714	1148.312	-1.985274	0.0521
R-squared	0.374655	Mean dependent var		1171.921
Adjusted R-squared	0.295065	S.D. dependent var		1925.728
S.E. of regression	1616.849	Akaike info criterion		17.73251
Sum squared resid	1.44E+08	Schwarz criterion		18.00466
Log likelihood	-550.5742	Hannan-Quinn criter.		17.83955
F-statistic	4.707340	Durbin-Watson stat		1.688747
Prob(F-statistic)	0.000343			

Ta có hàm hồi quy mẫu như sau:

$$\hat{Y}_i = -2279.714 + 0.018986X_{1i} + 35.27848X_{2i} + 58.04288X_{3i} - 246.4273X_{4i} + 222.8695X_{5i} - 298.4548X_{6i} + 18.87725X_{7i}$$

Dấu của các hệ số hồi quy đúng như kỳ vọng ở câu c.

e. Nêu ý nghĩa của các hệ số hồi quy tìm được trong câu d?

Các hệ số hồi quy trong câu (d) chính là tác động biên tế hay tác động riêng phần của các biến X1, X2, X3, X4, X5, X6, X7 lên sự biến thiên của Y về mặt trung bình trong điều kiện các yếu tố khác không đổi. Từ đó, ta có thể rút ra được ý nghĩa của các hệ số hồi quy như sau:

+ Giá trị $\hat{\beta}_1 = -2279.714$ cho biết giá trị tung độ gốc của biến Y khi các giá trị X_i đều bằng 0

+ $\hat{\beta}_2 = 0.018986 > 0$: Mỗi quan hệ giữa biến thu nhập và biến chi tiêu cho y tế là quan hệ đồng biến, ngoài ra hệ số này còn cho biết trong điều kiện các yếu tố khác không đổi, khi thu nhập tăng lên (giảm xuống) 1 nghìn đồng/năm thì chi tiêu cho y tế tăng lên (giảm xuống) về mặt trung bình khoảng 0,0189 nghìn đồng/năm.

+ $\hat{\beta}_3 = 35.27848 > 0$: Mối quan hệ giữa biến bằng cấp cao nhất đạt được và biến chi tiêu cho y tế là quan hệ đồng biến, ngoài ra hệ số này còn cho biết sự khác biệt theo quan hệ đồng biến về chi tiêu cho y tế về mặt trung bình trong điều kiện khác không đổi là khoảng 35.28 nghìn đồng/năm khi người đó thay đổi một cấp độ của bằng cấp cao nhất mà người đó đạt được

+ Tương tự, các bạn nhận xét ý nghĩa cho hệ số hồi quy còn lại

f. Theo bạn trong các biến nói trên, biến nào có ảnh hưởng mạnh nhất đến tổng chi tiêu cho y tế của chủ hộ? Tại sao?

Để xem xét biến nào có ảnh hưởng mạnh nhất đến biến chi tiêu y tế của hộ ta phải dùng hệ số hồi quy chuẩn hóa nhằm khử sự khác biệt về đơn vị giữa các biến.

$$\text{Hệ số hồi quy chuẩn hóa} = \frac{\widehat{\beta}_i - \bar{\beta}_i}{se(\widehat{\beta}_i)}$$

Với giả định trung bình của các ước lượng hệ số hồi quy của các biến X_i đều bằng 0 thì ta có các hệ số hồi quy chuẩn hóa cũng chính là các giá trị T-statistic. Như vậy, từ bảng hồi quy ở câu d, ta thấy biến Tuổi chủ hộ (X3) có giá trị T-statistic (3,855) là giá trị cao nhất.

Vậy, với giả định trên ta có thể xem biến X3 là biến có ảnh hưởng mạnh nhất đến biến chi tiêu y tế của hộ ở mức ý nghĩa 5%.

g. Hãy loại trừ những biến có Pvalue > 5%, hồi quy biến phụ thuộc nói trên với các biến độc lập còn lại?

Loại trừ các biến có giá trị P-value > 5% theo đó ta sẽ loại nhóm biến X2, X4, X5, X6, X7. Chạy hồi quy Y theo X1 và X3 ta có kết quả sau:

Dependent Variable: TONG_CHI_TIEU_Y_TE_CUA_H
 Method: Least Squares
 Date: 01/03/13 Time: 15:06
 Sample: 1 65
 Included observations: 65

Variable	Coefficient	Std. Error	t-Statistic	Prob.
THU_NHAP_X1__1000D_NAM_	0.019489	0.005142	3.789989	0.0003
TUOI_CUA_CHU_HO_X3	51.00887	13.56855	3.759346	0.0004
C	-2359.865	769.0325	-3.068615	0.0032
R-squared	0.310692	Mean dependent var		1145.923
Adjusted R-squared	0.288456	S.D. dependent var		1901.220
S.E. of regression	1603.737	Akaike info criterion		17.64312
Sum squared resid	1.59E+08	Schwarz criterion		17.74347
Log likelihood	-570.4012	Hannan-Quinn criter.		17.68271
F-statistic	13.97264	Durbin-Watson stat		1.647208
Prob(F-statistic)	0.000010			

Ta có hàm hồi quy mẫu lúc này sẽ là: $\hat{Y}_i = -2359.865 + 0.019489X_{1i} + 51.00887X_{3i}$

h. Bạn có nhận xét gì về hệ số R^2 của hai mô hình và giải thích các quan sát của bạn?

Ta thấy, giá trị R^2 :

- Mô hình 1 (mô hình đầy đủ): 0.374655
- Mô hình 2 (mô hình loại bỏ một số biến): 0.310692

Như vậy giá trị R^2 của mô hình 1 cao hơn mô hình 2, cho thấy rằng càng bỏ thêm nhiều biến mặc dù chưa xác định biến đưa vào có ý nghĩa hay không thì giá trị R^2 sẽ tăng.

Điều này được giải thích như sau:
$$R^2 = 1 - \frac{ESS}{TSS} = 1 - \frac{\sum \hat{u}_i^2}{\sum (Y_i - \bar{Y})^2}$$

Khi càng đưa thêm biến giải thích vào mô hình thì sẽ càng khiến phần dư giảm xuống (vì bản chất những gì không giải thích được đều nằm ở phần dư), do vậy tăng thêm biến sẽ khiến tổng bình phương phần dư giảm, trong khi TSS không đổi, dẫn tới R^2 luôn luôn tăng.

Giá trị R^2 tăng khả năng giải thích của mô hình, nhưng bản chất thì lại không làm rõ được tầm quan trọng của biến đưa vào, do đó nếu dựa vào giá trị R^2 để đánh giá tính hiệu quả của mô hình sẽ dẫn đến tình huống không chính xác vì sẽ đưa quá nhiều biến không cần thiết, làm phức tạp mô hình.

i. Bạn có nhận xét gì về R^2 điều chỉnh giữa hai mô hình? Giải thích các quan sát của bạn?

Ta thấy, giá trị R^2 điều chỉnh:

- Mô hình 1: 0.295065
- Mô hình 2: 0.288456

Giá trị R^2 điều chỉnh của mô hình 2 là xấp xỉ giá trị R^2 điều chỉnh của mô hình 1. Sự khác biệt trên chưa thể kết luận là có sự khác biệt đáng kể về mặt thống kê. Việc loại các biến có giá trị P-value >5% khi chưa thực hiện các kiểm định cần thiết hoặc cũng sẽ có những cơ sở lý thuyết ủng hộ các biến vừa loại là thực sự có tác động đến biến chi tiêu y tế của hộ. Do đó, ta cần nghiên cứu cơ sở lý thuyết để củng cố và làm cơ sở loại biến, tránh việc loại biến thực sự có tác động quan trọng đến biến phụ thuộc

j. Bạn dùng chỉ số nào để so sánh về tính phù hợp của hai mô hình nói trên?

Việc sử dụng R^2 hiệu chỉnh theo bậc tự do cho biết mức độ quan trọng (khả năng giải thích được) của biến được thêm vào và đánh giá được mức độ ý nghĩa của biến thêm vào là có đáng kể hay không. Do đó chúng ta sử dụng độ thích hợp được hiệu chỉnh theo bậc tự do (R^2

điều chỉnh) để đánh giá tính thích hợp của mô hình. Lý do R^2 điều chỉnh cho phép đánh đổi giữa việc tăng R^2 và giảm bậc tự do, cũng như đánh giá đúng mức độ ý nghĩa hay độ quan trọng của các biến được thêm vào.

k. Thay vì coi “Y - tổng chi tiêu cho y tế của hộ” là biến phụ thuộc, bạn hãy dùng logarit cơ số tự nhiên của “tổng chi tiêu y tế của hộ” ($\ln(Y)$) làm biến phụ thuộc. Theo bạn thực sự mô hình nào tốt hơn? Tại sao?

Thực hiện các bước tiến hành trên Eview, ta có mô hình hồi quy của biến LNY theo các biến giải thích đã cho ở mô hình hồi quy của biến Y như sau:

Dependent Variable: LNY
 Method: Least Squares
 Date: 01/03/13 Time: 15:26
 Sample: 1 65
 Included observations: 54

Variable	Coefficient	Std. Error	t-Statistic	Prob.
THU_NHAP_X1__1000D_NAM_	1.15E-05	4.07E-06	2.827762	0.0069
BANG_CAP_CAO_NHAT_DAT_DU	-0.010697	0.086072	-0.124284	0.9016
TUOI_CUA_CHU_HO_X3	0.028601	0.011199	2.553833	0.0140
GIOI_TINH_CUA_CHU_HO_X4	-0.003740	0.315134	-0.011869	0.9906
TRE_EM_DUOI_15_TUOI_X5	0.032342	0.161353	0.200443	0.8420
PHU_NU_TRONG_DO_TUOI_SIN	-0.158518	0.131931	-1.201524	0.2357
SO_NHAN_KHAU_X7	0.036062	0.059494	0.606147	0.5474
C	4.500093	0.861073	5.226147	0.0000
R-squared	0.293188	Mean dependent var		6.502430
Adjusted R-squared	0.185630	S.D. dependent var		1.205794
S.E. of regression	1.088138	Akaike info criterion		3.142767
Sum squared resid	54.46604	Schwarz criterion		3.437431
Log likelihood	-76.85470	Hannan-Quinn criter.		3.256407
F-statistic	2.725850	Durbin-Watson stat		1.527577
Prob(F-statistic)	0.018798			

Ta có phương trình hồi quy như sau:

$$LNY_i = 4.500093 + 1.15E - 05X_{1i} - 0.010697X_{2i} + 0.028601X_{3i} - 0.00374X_{4i} + 0.032342X_{5i} - 0.158518X_{6i} + 0.036062X_{7i}$$

Để so sánh mô hình nào tốt hơn ta dựa vào R^2 điều chỉnh với yêu cầu phải cùng số lượng quan sát và biến phụ thuộc của hai mô hình phải giống nhau.

Tuy nhiên trong trường hợp này biến phụ thuộc của hai mô hình là khác nhau nên ta không thể kết luận mô hình nào thực sự tốt hơn.