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 9

HỒI QUY ĐA BIẾN (TIẾP THEO)

Ngày Phát: Thứ hai 10/12/2012

Ngày Nộp: Thứ hai 17/12/2012

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: Dưới đây là bảng hồi quy của một mô hình kinh tế lượng

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	0.0013
X2	0.007695	0.003576	2.152021	0.0326
X3	79.76524	49.25838	1.619323	0.1070
X2*X3	74.16483	123.7128	0.599492	0.5495
X3*X3	37.87703	8.552994	4.428511	0.0000
D1	746.9610	297.3025	2.512462	0.0128

R-squared	0.172804	Mean dependent var	1253.115
Adjusted R-squared	0.151484	S.D. dependent var	1764.184
S.E. of regression	1625.076	Akaike info criterion	17.65404
Sum squared resid	5.12E+08	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 biến nào có ý nghĩa thống kê ở mức ý nghĩa 1%, 5%?

Các biến có ý nghĩa thống kê ở mức ý nghĩa α% là những biến có giá trị P-value (β_{Xk}) < α. Theo định nghĩa trên thì:

+ Với mức ý nghĩa 1% thì chỉ có biến X_3*X_3 có ý nghĩa thống kê

+ Với mức ý nghĩa 5%, các biến có ý nghĩa thống kê gồm: X_2, X_3*X_3 và biến D_1

b. Ở mức ý nghĩa bao nhiêu thì biến X_2 và X_3 cùng có ý nghĩa thống kê?

Ta có: P-value(β_{X2}) = 3,26% và P-value(β_{X3}) = 10,7%. Như vậy để biến X_2 và X_3 cùng có ý nghĩa thống kê thì mức ý nghĩa α phải thỏa mãn điều kiện:

α_{min} = max{ P-value(β_{X2}); P-value(β_{X3})} = 10,7%

Như vậy tại mức ý nghĩa tối thiểu là 10,7% thì các biến X_2 và X_3 cùng có ý nghĩa thống kê **(lưu ý: không làm tròn 11%)**

c. Ước lượng khoảng tin cậy 95% của biến X_2?

Ước lượng khoảng tin cậy của biến X_2

KTC = [$\hat{\beta}_2$ - $t_{\alpha/2}$* $s_{\widehat{\beta_2}}$; $\hat{\beta}_2$ - $t_{\alpha/2}$* $s_{\widehat{\beta_2}}$]

Với $\hat{\beta}_2$ = 0,007695

$t_{\alpha/2;(n-k)}$ = $t_{194}^{0,025}$ = 1,9723

$s_{\widehat{\beta_2}}$ = 0,003576

Như vậy KTC của biến X_2 nằm trong khoảng [0,00064;0,01475]

Nhận xét: vì KTC này không chứa giá trị 0 nên ta có thể thấy biến X_2 thực sự có tác động lên sự biến thiên về mặt trung bình của biến Y hay biến X_2 có ý nghĩa thống kê ở mức ý nghĩa 5%.

d. Có người cho rằng mô hình này có tính giải thích rất thấp, các biến được đưa vào mô hình này không giải thích được sự biến thiên của Y (có nghĩa là $R^2 = 0$), hãy kiểm định phát biểu này?

Để kiểm định nhận định trên ta dùng kiểm định F cho cặp giả thuyết sau:

H_o: $\beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$ ($R^2 = 0$)

H_a: có ít nhất 1 β_k khác 0 (k=$\overline{2;6}$) (R^2 khác 0)

Ta có giá trị quan sát của F = $\frac{R^2/(k-1)}{(1-R^2)/(n-k)}$ = $\frac{0.1728/(6-1)}{(1-0.1728)/(200-6)}$ = 8.1054 (chính là giá trị F-statistic trong mô hình hồi quy)

Giá trị tới hạn: $F_{(\alpha;k-1;n-k)}$ = $F_{(5\%,5,194)}$ = 2,2606

So sánh ta thấy F > $F_{(\alpha;k-1;n-k)}$: Ta đủ cơ sở để bác bỏ giả thuyết H_o hay nói cách khác nhận định trên là không đúng, các biến đưa vào trong mô hình có thể giải thích được sự biến thiên về mặt trung bình của biến Y ở mức ý nghĩa 5%.

(có thể dùng P-value của F-statistic để kiểm định)

e. Các lý thuyết và nghiên cứu thực tế trước đó đều kết luận X3 tác động lên Y một cách rõ rệt. Theo chiến lược xây dựng mô hình từ tổng quát tới đơn giản bạn sẽ loại những biến nào ra khỏi mô hình? Tại sao?

Theo chiến lược xây dựng mô hình từ tổng quát đến đơn giản thì ta sẽ loại những biến không có ý nghĩa thống kê.

Với kết quả hồi quy trên thì ở mức ý nghĩa $\alpha = 5\%$ thì những biến X_3 và biến X_2*X_3 sẽ không có ý nghĩa thống kê ở mức ý nghĩa này (do các giá trị P-value $>\alpha$).

Tuy nhiên các lý thuyết và các nghiên cứu thực tế đều kết luận X_3 có tác động lên Y một cách rõ rệt do đó ta đủ cơ sở lý thuyết để giữ lại biến X_3 trong mô hình R và chỉ loại biến X_2*X_3 ra khỏi mô hình.

Ngoài ra, để có cơ sở chắc chắc cho việc loại biến X_2*X_3 ra khỏi mô hình R ta cũng cần tiến hành kiểm định Wald-test cho việc loại biến để có thể kết luận chắc chắn hơn.

Câu 2: File đính kèm theo 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. Xây dựng mô hình hồi quy về mối quan hệ tuyến tính với biến phụ thuộc là "Chi tiêu cho y tế của chủ hộ" và các biến giải thích lần lượt là: "Thu nhập", "Tuổi của chủ hộ", "Trẻ em dưới 15 tuổi", "Số phụ nữ trong độ tuổi sinh đẻ", "Quy mô hộ" . Bạn có nhận xét gì về kết quả ước lượng nói trên?

Xây dựng mô hình hồi quy về mối quan hệ với biến phụ thuộc là "Chi tiêu cho y tế của chủ hộ" và các biến giải thích lần lượt là: "Thu nhập", "Tuổi của chủ hộ", "Trẻ em dưới 15 tuổi", "Số phụ nữ trong độ tuổi sinh đẻ", "Quy mô hộ"

Dependent Variable: TONG_CHI_TIEU_CHO_Y_TE_C
Method: Least Squares
Date:
Sample: 1 65
Included observations: 63

Variable	Coefficient	Std. Error	t-Statistic	Prob.
THU_NHAP_X1	0.019487	0.005236	3.721762	0.0005
TUOI_CUA_CHU_HO_X3	56.15605	13.98608	4.015139	0.0002
TRE_EM_DUOI_15_TUOI_X5	227.6936	207.3142	1.098302	0.2767
PHU_NU_TRONG_DO_TUOI_SIN	-307.7094	175.6201	-1.752132	0.0851
SO_NHAN_KHAU_X7	28.72945	63.47708	0.452596	0.6526
C	-2574.010	854.5214	-3.012224	0.0039

R-squared	0.369727	Mean dependent var	1171.921
Adjusted R-squared	0.314440	S.D. dependent var	1925.728
S.E. of regression	1594.475	Akaike info criterion	17.67687
Sum squared resid	1.45E+08	Schwarz criterion	17.88098
Log likelihood	-550.8214	Hannan-Quinn criter.	17.75715
F-statistic	6.687407	Durbin-Watson stat	1.640800
Prob(F-statistic)	0.000058		

Hàm hồi quy tuyến tính sẽ là

$\hat{Y}_i = 2.574{,}010 + 0{,}19487X_{1i} + 56{,}15605X_{3i} + 227{,}6936X_{5i} - 307.70494X_{6i} + 28{,}72845X_{7i}$

Nhận xét:

+ Dựa vào P-value của các ước lượng hệ số bê ta, ta thấy với mức ý nghĩa 5% thì các biến X_1, X_3 có ý nghĩa thống kê và các biến còn lại trong mô hình (X_5, X_6, X_7) không có ý nghĩa thống kê ở mức ý nghĩa này.

+ Dựa vào dấu của các ước lượng hệ số bê ta thì các biến X_1, X_3 X_5 và X_7 có quan hệ đồng biến với biến phụ thuộc Y ;trong khi đó biến X_6 có quan hệ nghịch biến với biến Y.

+ Với $R^2 = 36{,}97\%$ điều này cho biết với các điều kiện khác không đổi thì các biến độc lập có trong mô hình hồi quy trên có khả năng giải thích được khoảng 36,97% sự biến thiên về mặt trung bình của biến phụ thuộc Y.

Ngoài ra, bạn nhận xét thêm tác động biên của từng biến lên biến độc lập cũng như giá trị P-value(F-statistic)…

b. Dùng chiến lược xây dựng mô hình từ tổng quát đến đơn giản. Nếu mô hình dùng trong câu (a) là mô hình không giới hạn (mô hình U) thì mô hình giới hạn (mô hình R) của bạn là gì?

Mô hình U: $\hat{Y}_i = 2.574{,}010 + 0{,}19487X_{1i} + 56{,}15605X_{3i} + 227{,}6936X_{5i} - 307.70494X_{6i} + 28{,}72845X_{7i}$

Vì các biến X_5, X_6, X_7 có trong mô hình U không có ý nghĩa thống kê ở mức ý nghĩa 5%, do đó ta sẽ đề xuất loại nhóm biến này ra khỏi mô hình và khi đó mô hình R sẽ gồm hai biến X_1 và X_3 có dạng như sau:

Mô hình R: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{1i} + \hat{\beta}_3 X_{3i}$

Để có cơ sở loại các biến X_5, X_6, X_7 ra khỏi mô hình U, ta sẽ thực hiện kiểm định Wald cho cặp giải thiết sau:

H_o: $\beta_4 = \beta_5 = \beta_6 = 0$

Ha: có ít nhất một hệ số bê ta trong các hệ số β_4 , β_5 , β_6 khác 0

Kết quả Wald- test như sau:

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	1.509768	(3, 57)	0.2218
Chi-square	4.529305	3	0.2097

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(3)	227.6936	207.3142
C(4)	-307.7094	175.6201
C(5)	28.72945	63.47708

Restrictions are linear in coefficients.

Ta có P-value (F-statistic) = 22,18% > mức ý nghĩa α (5%)

Kết luận: Chưa đủ cơ sở để bác bỏ giả thiết Ho ở mức ý nghĩa 5%, như vậy ta có thể loại bỏ nhóm biến này khỏi mô hình U và mô hình R đề xuất sẽ có dạng tuyến tính như sau: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{1i} + \hat{\beta}_3 X_{3i}$ với độ tin cậy 95%

c. Hồi quy mô hình giới hạn mà bạn đề nghị trong câu (b), bạn lựa chọn mô hình nào? Tại sao?

Hồi quy mô hình R:

Dependent Variable: TONG_CHI_TIEU_CHO_Y_TE_C
Method: Least Squares
Date:
Sample: 1 65
Included observations: 65

Variable	Coefficient	Std. Error	t-Statistic	Prob.
THU_NHAP_X1	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		

Mô hình R: $\hat{Y}_i$ = -2.359,85 + 0,019489X_{1i} + 51,00887X_{3i}

So sánh kết quả hồi quy của 2 mô hình U và R ta thấy R^2_{adj} của hai mô hình này tương đương nhau tuy nhiên để lựa chọn mô hình thì ta sẽ lựa chọn mô hình R vì:

+ Mô hình U có chứa các biến không có ý nghĩa thống kê ở mức ý nghĩa 5%

+ Kiểm định F để loại biến ra khỏi mô hình U (như kết quả câu b) ta thấy đủ cơ sở để loại bỏ các biến X_5, X_6, và X_7.

+ Mô hình R là mô hình rút gọn của mô hình U do đó khi tiến hành thu thập số liệu và xử lý số liệu sẽ giúp ta tiến hành dễ dàng hơn và tiết kiệm được chi phí điều tra.

d. Nếu bạn xây dựng mô hình theo chiến lược từ đơn giản đến tổng quát thì bạn sẽ bắt đầu bằng biến giải thích nào? Lý giải sự lựa chọn của bạn?

Nếu xây dựng mô hình theo chiến lược từ đơn giản đến tổng quát thì ta thấy kết quả hồi quy ở mô hình R đều cho thấy cả hai biến THUNHAP và TUOICHUHO đều có ý nghĩa thống kê ở mức ý nghĩa 5% và có ảnh hưởng đến sự biến thiên của TONGCHITIEUCHOYTE. Tuy nhiên để xây dựng mô hình này ta nên bắt đầu từ biến giải thích là THUNHAP vì:

Những lý thuyết kinh tế như lý thuyết của Keynes đã chỉ rõ mối quan hệ giữa thu nhập và chi tiêu mà trong đó chi tiêu cho y tế là một cấu phần của chi tiêu hộ.

Ngoài ra, so sánh hệ số tương quan giữa các biến giải thích trong mô hình R với biến phụ thuộc thì nhận thấy biến THUNHAP có tương quan mạnh với biến TONGCHITIEUCHOYTE hơn so với biến TUOICHUHO.

	Tổng chi tiêu cho y tế của hộ Y (1000đ/năm)	Thu nhập X1 (1000đ/năm)	Tuổi của chủ hộ X3
Tổng chi tiêu cho y tế của hộ Y (1000đ/năm)	1		
Thu nhập X1 (1000đ/năm)	***0.391875886***	1	
Tuổi của chủ hộ X3	***0.388580482***	-0.019732636	1

e. Có người cho rằng dạng hàm đúng của mô hình tuyến tính này (mô hình bị giới hạn – R – trong câu c) phải dùng logarit cơ số tự nhiên của "tổng chi tiêu y tế của hộ" (ln(Y)) làm biến phụ thuộc. Bạn có đồng ý với quan điểm này không? Tại sao?

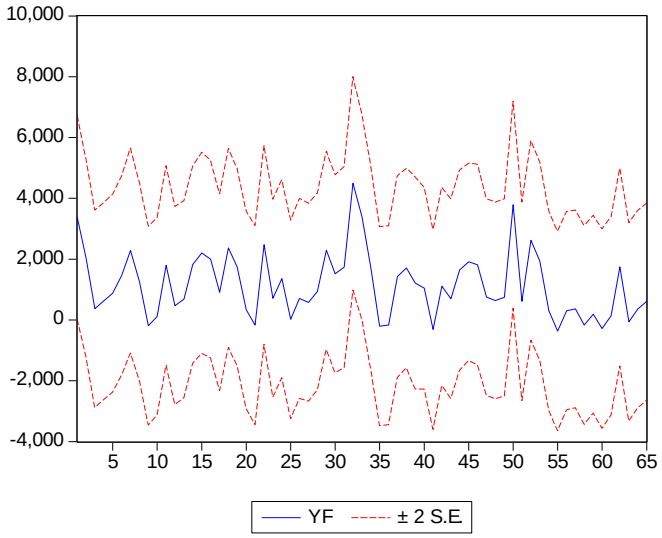
Tiến hành kiểm định cặp giả thuyết sau:

H_o: $Y_i = \beta_1 + \beta_2 X_{1i} + \beta_3 X_{3i} + u_i$ là mô hình đúng (1)

H_a: $\ln Y_i = \beta'_1 + \beta'_2 X_{1i} + \beta'_3 X_{3i} + v_i$ là mô hình đúng (2)

Tiến hành các bước trong Eviews như sau:

+ tính yf



Forecast: YF
Actual: TONG_CHI_TIEU_CHO_Y_TE_C
Forecast sample: 1 65
Included observations: 65

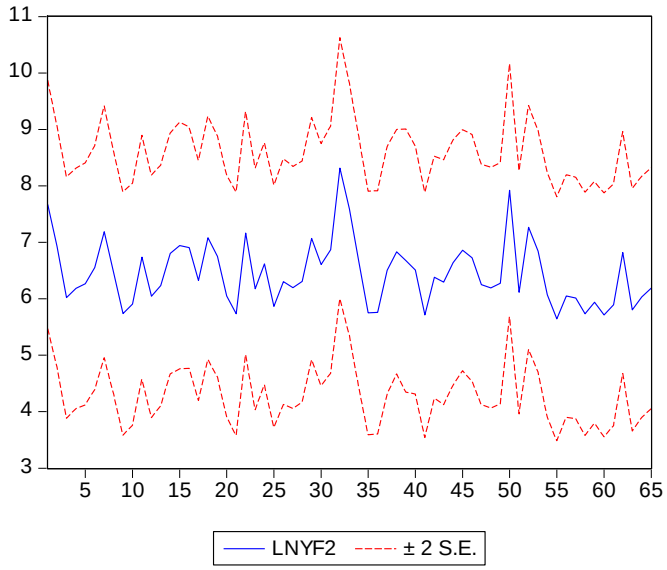
Root Mean Squared Error	1566.290
Mean Absolute Error	1059.415
Mean Abs. Percent Error	182.5686
Theil Inequality Coefficient	0.416281
Bias Proportion	0.000000
Variance Proportion	0.284194
Covariance Proportion	0.715806

+ Chạy hồi quy mô hình 2

Dependent Variable: LNY
Method: Least Squares
Date: 01/09/13 Time: 23:05
Sample: 1 65
Included observations: 56

Variable	Coefficient	Std. Error	t-Statistic	Prob.
THU_NHAP_X1	1.14E-05	3.50E-06	3.258668	0.0020
TUOI_CUA_CHU_HO_X3	0.024577	0.009848	2.495600	0.0157
C	4.677854	0.548282	8.531835	0.0000

R-squared	0.251455	Mean dependent var	6.474261
Adjusted R-squared	0.223208	S.D. dependent var	1.195229
S.E. of regression	1.053425	Akaike info criterion	2.994053
Sum squared resid	58.81430	Schwarz criterion	3.102554
Log likelihood	-80.83349	Hannan-Quinn criter.	3.036119
F-statistic	8.902001	Durbin-Watson stat	1.609544
Prob(F-statistic)	0.000464		



Forecast: LNYF2
Actual: LNY
Forecast sample: 1 65
Included observations: 56

Root Mean Squared Error	1.024820
Mean Absolute Error	0.860039
Mean Abs. Percent Error	14.02893
Theil Inequality Coefficient	0.078331
Bias Proportion	0.000000
Variance Proportion	0.332043
Covariance Proportion	0.667957

+ Đặt biến Z = $\ln\hat{Y}$ - $\widehat{lnY}$ và sau đó chạy mô hình gồm Y, X_1, X_3 và Z:

Dependent Variable: Y
Method: Least Squares
Date: 01/09/13 Time: 23:11
Sample: 1 65
Included observations: 56

Variable	Coefficient	Std. Error	t-Statistic	Prob.
THU_NHAP_X1	0.026935	0.006009	4.482799	0.0000
TUOI_CUA_CHU_HO_X3	80.39745	17.67308	4.549147	0.0000
Z	-990.8401	424.7608	-2.332702	0.0236
C	-4107.620	1039.917	-3.949951	0.0002

R-squared	0.380104	Mean dependent var	1268.714
Adjusted R-squared	0.344341	S.D. dependent var	2020.810
S.E. of regression	1636.306	Akaike info criterion	17.70702
Sum squared resid	1.39E+08	Schwarz criterion	17.85169
Log likelihood	-491.7965	Hannan-Quinn criter.	17.76311
F-statistic	10.62835	Durbin-Watson stat	1.905137
Prob(F-statistic)	0.000015		

Từ kết quả trên ta có P-value($\hat{\beta}_Z$) = 0,0236 < α (5%), như vậy trong mô hình trên biến Z có ý nghĩa thống kê. Hay, ta có cơ sở để bác bỏ giả thuyết H_o.

Kết luận: Mô hình giới hạn (R) có dạng hàm đúng là dạng hàm logarit cơ số tự nhiên của biến TONGCHITIEUCHOYTE và đồng ý với quan điểm trên.