

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 8
HỘI QUI TUYỂN TÍNH ĐA BIẾN
Ngày Phát: Thứ Hai, 6/12/2010
Ngày Nộp: 8:20 sáng, Thứ Hai, 13/12/2010

Bản in nộp tại Phòng Giáo Vụ
Bản điện tử gửi đến thầy Nguyễn Khánh Duy theo địa
chỉ duykn@fep.vnn.vn

Bài 1: (Bài 7.20 về Nhu cầu về hoa hồng trong Gujarati 3).

* Bảng dưới đây cho dữ liệu hàng quý về những biến sau:

Y = số lượng hoa hồng bán được, theo tá

X₂ = giá bán sỉ trung bình của hoa hồng, \$/tá

X₃ = giá bán sỉ trung bình của hoa cẩm chướng, \$/tá

X₄ = thu nhập khả dụng trung bình hàng tuần của gia đình, \$/tuần

X₅ = biến xu hướng có các giá trị 1, 2 và tiếp tục, trong giai đoạn 1971-III cho đến 1975-II tại khu vực trung tâm Detroit.

Năm và quý	Y	X ₂	X ₃	X ₄	X ₅
1971 - III	11,484	2.26	3.49	158.11	1
-IV	9,348	2.54	2.85	173.36	2
1972 - I	8,429	3.07	4.06	165.26	3
- II	10,079	2.91	3.64	172.92	4
- III	9,240	2.73	3.21	178.46	5
- IV	8,862	2.77	3.66	198.62	6
1973 - I	6,216	3.59	3.76	186.28	7
- II	8,253	3.23	3.49	188.98	8
- III	8,038	2.60	3.13	180.49	9
- IV	7,476	2.89	3.20	183.33	10
1974 - I	5,911	3.77	3.65	181.87	11
- II	7,950	3.64	3.60	185.00	12
- III	6,134	2.82	2.94	184.00	13
- IV	5,868	2.96	3.12	188.20	14
1975 - I	3,160	4.24	3.58	175.67	15
- II	5,872	3.69	3.53	188.00	16

* Tôi biết ơn Joe Walsh vì đã thu thập những dữ liệu này từ một nhà buôn sỉ lớn ở khu vực trung tâm Detroit và tiếp đó đã xử lý những dữ liệu này.

Bạn được yêu cầu xem xét các hàm nhu cầu sau:

$$Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \alpha_4 X_{4t} + \alpha_5 X_{5t} + u_t \quad (1)$$

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_4 \ln X_{4t} + \beta_5 \ln X_{5t} + u_t \quad (2)$$

a) Ước lượng các tham số của mô hình tuyến tính và giải thích các kết quả?

Kết quả ước lượng mô hình (1) từ Eviews như sau

Dependent Variable: Y
 Method: Least Squares
 Date: 01/05/11 Time: 04:30
 Sample: 1971Q3 1975Q2
 Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10816.040	5988.348	1.806	0.098
X2	-2227.704	920.466	-2.420	0.034
X3	1251.141	1157.021	1.081	0.303
X4	6.283	30.622	0.205	0.841
X5	-197.400	101.561	-1.944	0.078

R-squared	0.835	Mean dependent var	7645.000
Adjusted R-squared	0.775	S.D. dependent var	2042.814
S.E. of regression	969.874	Akaike info criterion	16.843
Sum squared resid	10347220.000	Schwarz criterion	17.084
Log likelihood	-129.740	Hannan-Quinn criter.	16.855
F-statistic	13.886	Durbin-Watson stat	2.334
Prob(F-statistic)	0.000		

Hay hàm hồi quy mẫu là:

$$\hat{Y} = 10816.040 - 2227.704 * X_2 + 1251.141 * X_3 + 6.283 * X_4 - 197.400 * X_5$$

Giải thích kết quả:

- Các hệ số ước lượng $\hat{\alpha}_2$, $\hat{\alpha}_5$ đều có P_{value} nhỏ hơn 0.1 nên mỗi quan hệ giữa biến Y và các biến X_2 , X_5 có ý nghĩa thống kê tại mức ý nghĩa 10%.
- Hệ số ước lượng $\hat{\alpha}_3$, $\hat{\alpha}_4$, có P_{value} tương ứng là 0.303, 0.841 (>0.1) nên chúng không có ý nghĩa thống kê tại mức ý nghĩa 10%.
- Trong điều kiện các yếu tố khác không đổi, khi giá bán sỉ trung bình của hoa hồng (X_2) tăng (hoặc giảm) \$1/tá thì lượng hoa hồng (Y) bán được giảm (hoặc tăng) trung bình 2227.704 tá.

- Bạn có thể giải thích ý nghĩa của các hệ số $\hat{\alpha}_3, \hat{\alpha}_4$ tương tự cho dữ liệu của mẫu
- Trong điều kiện các yếu tố khác không đổi, cứ sau mỗi quý lượng hoa hồng bán được giảm trung bình 197.4 tá.

Hệ số Adjusted R^2 cho biết 77.5% biến thiên của biến Y được giải thích bởi mô hình.

b) Ước lượng các tham số của mô hình log - log và giải thích các kết quả.

Dependent Variable: LOG(Y)
 Method: Least Squares
 Date: 01/05/11 Time: 05:35
 Sample: 1971Q3 1975Q2
 Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.627	6.148	0.102	0.921
LOG(X2)	-1.274	0.527	-2.418	0.034
LOG(X3)	0.937	0.659	1.422	0.183
LOG(X4)	1.713	1.201	1.426	0.182
LOG(X5)	-0.182	0.128	-1.420	0.183
R-squared	0.778	Mean dependent var		8.902
Adjusted R-squared	0.697	S.D. dependent var		0.307
S.E. of regression	0.169	Akaike info criterion		-0.469
Sum squared resid	0.314	Schwarz criterion		-0.228
Log likelihood	8.753	Hannan-Quinn criter.		-0.457
F-statistic	9.635	Durbin-Watson stat		1.783
Prob(F-statistic)	0.001			

$$\ln(Y) = 0.627 - 1.274 \ln(X_2) + 0.937 \ln(X_3) + 1.713 \ln(X_4) - 0.182 \ln(X_5) + \hat{u}$$

- Chỉ có hệ số ước lượng $\hat{\beta}_2$ có P_{value} nhỏ hơn 0.1 nên mối quan hệ giữa Y và X_2 có ý nghĩa về mặt thống kê tại mức ý nghĩa 10%.
- Nếu các yếu tố khác không đổi, khi giá bán sỉ trung bình của hoa hồng (X_2) tăng (giảm) 1% thì lượng hoa hồng trung bình (Y) bán được giảm (tăng) tương ứng khoảng 1.274 %.

- Các hệ số đứng trước các biến giải thích khác không có ý nghĩa thống kê, tuy nhiên bạn có thể giải thích ý nghĩa kinh tế của các hệ số ấy cho dữ liệu mẫu tương tự.

Hệ số Adjusted R^2 cho biết 69.7% biến thiên của biến $\ln(Y)$ được giải thích bởi mô hình.

c) β_2 , β_3 và β_4 lần lượt cho biết độ co giãn của cầu theo giá riêng, giá chéo, và theo thu nhập. Dấu tiên nghiệm của các hệ số là gì? Các kết quả có giống như các kỳ vọng tiên nghiệm không?

- β_2 được kỳ vọng có dấu âm, vì theo lý thuyết kinh tế vi mô, có mối quan hệ nghịch biến giữa giá bán và lượng cầu của hoa hồng.
- β_3 được kỳ vọng có dấu dương, vì theo lý thuyết kinh tế vi mô, có mối quan hệ đồng biến giữa giá bán của hoa cẩm chướng, một hàng hóa thay thế cho hoa hồng, và lượng cầu hoa hồng.
- β_4 được kỳ vọng có dấu dương, vì theo lý thuyết kinh tế vi mô, thu nhập của hộ là một trong những nhân tố ảnh hưởng tới lượng cầu ngoài giá bán. Khi thu nhập tăng thì lượng cầu của hoa hồng được kỳ vọng sẽ tăng (nếu hoa hồng là hàng hóa bình thường).

d) Dựa trên các phân tích của bạn, bạn sẽ chọn mô hình nào và cho biết tại sao?

Hai mô hình đều có nhiều biến không có ý nghĩa thống kê, và dấu của các hệ số hồi quy đều phù hợp với kỳ vọng. Lựa chọn mô hình có thể phụ thuộc vào mục tiêu của bạn ví dụ như ước lượng tác động biên, hay ước lượng hệ số co giãn ... Tuy nhiên bạn có thể căn cứ vào độ chính xác của mô hình trong việc dự báo (thông qua các chỉ tiêu MAPE, RMSE...), R^2 giả, hay thực hiện kiểm định MWD để lựa chọn.

Nếu sử dụng R^2 giả = $r(Y, Y_f \text{ từ mô hình 2})^2 = 0.8764^2 = 0.768$ thì bạn chọn mô hình 1, vì R^2 giả của hàm log-log nhỏ hơn R^2 của hàm hồi quy tuyến tính (R^2 của hàm tuyến tính bằng 0.835)

Khi đề cập đến dự báo, ta quan tâm đến hệ số xác định của hai mô hình theo nguyên tắc: hệ số xác định càng tiến đến 1 thì mô hình được kỳ vọng dự báo càng chính xác. Tuy nhiên, trong trường hợp này, biến phụ thuộc của hai mô hình là khác nhau nên không thể so sánh hệ số xác định giữa hai mô hình một cách trực tiếp mà phải thông qua một phương pháp khác (đọc thêm Chương 7, sách Kinh tế lượng căn bản của Damodar N. Gujarati, trang 19).

Phương pháp này được tiến hành như sau:

- Ước lượng \hat{Y}_t từ $\ln Y_t$ trong mô hình (2).
- Tính hệ số r^2 giữa \hat{Y}_t và Y_t theo công thức:

$$r^2 = \frac{[\sum (Y_t - \bar{Y})(\hat{Y}_t - \bar{Y})]^2}{\sum (Y_t - \bar{Y})^2 \sum (\hat{Y}_t - \bar{Y})^2}$$

So sánh giá trị trên với R^2 của mô hình (1)

Bài 2: (Bài 7.23 Nhu cầu về thịt gà ở Mỹ, 1960-1982 trong Gujarati 3).

Để nghiên cứu về tiêu thụ thịt gà bình quân đầu người ở Mỹ, bạn có được những dữ liệu sau:

trong đó Y = tiêu thụ thịt gà bình quân đầu người, cân Anh

X₂ = thu nhập khả dụng thực bình quân đầu người, đôla

X₃ = giá bán lẻ thực thịt gà tính trên mỗi cân Anh, xu

X₄ = giá bán lẻ thực thịt heo tính trên mỗi cân Anh, xu

X₅ = giá bán lẻ thực thịt bò tính trên mỗi cân Anh, xu

X₆ = giá thực đa hợp của các sản phẩm thay thế cho thịt gà tính trên mỗi cân Anh, xu,

đây là trung bình có trọng số của các giá bán lẻ thực trên một pao của thịt heo và thịt bò, các trọng số là mức tiêu thụ thịt bò và thịt heo tương đối tính trong tổng tiêu thụ thịt heo và thịt bò.

Bây giờ hãy xem xét các hàm nhu cầu sau đây:

$$\ln Y_t = \alpha_1 + \alpha_2 \ln X_{2t} + \alpha_3 \ln X_{3t} + u_t \quad (1)$$

$$\ln Y_t = \gamma_1 + \gamma_2 \ln X_{2t} + \gamma_3 \ln X_{3t} + \gamma_4 \ln X_{4t} + u_t \quad (2)$$

$$\ln Y_t = \lambda_1 + \lambda_2 \ln X_{2t} + \lambda_3 \ln X_{3t} + \lambda_4 \ln X_{5t} + u_t \quad (3)$$

$$\ln Y_t = \theta_1 + \theta_2 \ln X_{2t} + \theta_3 \ln X_{3t} + \theta_4 \ln X_{4t} + \theta_5 \ln X_{5t} + u_t \quad (4)$$

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_4 \ln X_{6t} + u_t \quad (5)$$

Năm	Y	X ₂	X ₃	X ₄	X ₅	X ₆
1960	27.8	397.5	42.2	50.7	78.3	65.8
1961	29.9	413.3	38.1	52.0	79.2	66.9
1962	29.8	439.2	40.3	54.0	79.2	67.8
1963	30.8	459.7	39.5	55.3	79.2	69.6
1964	31.2	492.9	37.3	54.7	77.4	68.7
1965	33.3	528.6	38.1	63.7	80.2	73.6
1966	35.6	560.3	39.3	69.8	80.4	76.3
1967	36.4	624.6	37.8	65.9	83.9	77.2
1968	36.7	666.4	38.4	64.5	85.5	78.1
1969	38.4	717.8	40.1	70.0	93.7	84.7
1970	40.4	768.2	38.6	73.2	106.1	93.3
1971	40.3	843.3	39.8	67.8	104.8	89.7
1972	41.8	911.6	39.7	79.1	114.0	100.7
1973	40.4	931.1	52.1	95.4	124.1	113.5
1974	40.7	1,021.5	48.9	94.2	127.6	115.3
1975	40.1	1,165.9	58.3	123.5	142.9	136.7
1976	42.7	1,349.6	57.9	129.9	143.6	139.2
1977	44.1	1,449.4	56.5	117.6	139.2	132.0
1978	46.7	1,575.5	63.7	130.9	165.5	132.1
1979	50.6	1,759.1	61.6	129.8	203.3	154.4
1980	50.1	1,994.2	58.9	128.0	219.6	174.9
1981	51.7	2,258.1	66.4	141.0	221.6	180.8
1982	52.9	2,478.7	70.4	168.2	232.6	189.4

Nguồn: Dữ liệu về Y được lấy rút ra từ Citibase và dữ liệu về X₂ cho đến X₆ là từ Bộ nông nghiệp Mỹ. Tôi mang ơn Robert J.Fisher vì đã thu thập những dữ liệu này và những phân tích thống kê.

Từ lý thuyết kinh tế vi mô ta biết được rằng nhu cầu về một hàng hóa thường phụ thuộc vào thu nhập thực của người tiêu dùng, giá cả thực của hàng hóa, và giá cả thực của các hàng hóa cạnh tranh hay hỗ trợ. Chú ý những điều cân nhắc này và trả lời các câu hỏi sau

a) Các bạn sẽ chọn hàm nhu cầu nào trong số những hàm cho ở trên, và cho biết tại sao?

Theo lý thuyết kinh tế học vi mô, lượng tiêu thụ thịt gà phụ thuộc vào thu nhập khả dụng thực của người tiêu dùng, giá cả thực của thịt gà, giá cả thực của các hàng hóa thay thế hay bổ sung. Vì vậy, trong 5 mô hình trên, mô hình (4) và (5) mang đầy đủ các thông tin về các biến giải thích hơn các mô hình còn lại. Tuy nhiên, mô hình (5) sẽ tốt hơn so với mô hình (4) vì trong mô hình 5 có biến X_6 là giá thực đa hợp của các sản phẩm thay thế cho thịt gà nên tránh được hiện tượng đa cộng tuyến có thể xảy ra bởi hai biến X_4 , và X_5 như ở mô hình (4).

b) Các bạn giải thích các hệ số của $\ln X_{2t}$ và $\ln X_{3t}$ trong những mô hình này như thế nào?

Hệ số hồi quy đứng trước biến $\ln(X_2)$ và $\ln(X_3)$ là hệ số của giãn của cầu theo thu nhập, và hệ số co giãn của cầu theo giá

c) Các đặc trưng trong (2) và (4) khác nhau ở chỗ nào?

Về giá cả của sản phẩm thay thế, mô hình 2 chỉ có biến giá thị heo, còn mô hình 4 xem xét cả giá thị heo và giá thị bò

d) Nếu chọn đặc trưng (4) các bạn có dự đoán được trước những vấn đề gì? (Gợi ý: Giá của cả thịt heo và thịt bò được đưa vào mô hình cùng với giá thịt gà)

Giá thị heo, và giá thị bò đưa cùng vào mô hình có thể gây ra hiện tượng đa cộng tuyến

Thật vậy, trị tuyệt đối của hệ số tương quan giữa 2 biến $\ln(X_4)$ và $\ln(X_5)$ rất lớn

	LOG(X4)	LOG(X5)
LOG(X4)	1	0.9543
LOG(X5)	0.9543	1

e) Bởi vì đặc trưng (5) bao gồm giá đa hợp của thịt heo và thịt bò, các bạn có nghĩ rằng hàm nhu cầu (5) tốt hơn (4) không? Tại sao?

Mô hình 5 tốt hơn mô hình 4 vì có thể tránh được hiện tượng đa cộng tuyến

f) Thịt heo và/hay thịt bò có phải là các sản phẩm cạnh tranh hay hỗ trợ của thịt gà không? Làm sao các bạn biết?

Thịt heo và thịt bò nên là sản phẩm thay thế (cạnh tranh) của thịt gà. Bởi vì khi giá của sản phẩm thịt heo, hay thịt bò tăng, người ta thường có khuynh hướng chuyển sang tiêu thụ sản phẩm thay thế với hai sản phẩm này, ví dụ thịt gà (làm cầu thịt gà tăng). Cũng có thể lập luận rằng, khi giá thịt gà tăng, người ta có thể sử dụng thịt bò, hay thịt heo nhiều hơn. Nói cách khác hệ số co giãn của cầu thịt gà theo giá thịt bò, hay giá thịt heo thường là một số dương... Nếu dữ liệu tốt, bạn có thể kiểm định điều này bằng phương pháp hồi quy.

g) Giả thiết rằng hàm (5) là hàm nhu cầu "đúng". Hãy ước lượng các tham số của mô hình này, tính các sai số chuẩn của chúng, R^2 và \bar{R}^2 . Giải thích các kết quả này.

Phương trình hồi quy, sai số chuẩn tương ứng với các hệ số hồi quy, R^2 , \bar{R}^2 như sau:

$$\ln Y = 2.030 + 0.481 \ln X_2 - 0.351 \ln X_3 - 0.061 \ln X_6 + \hat{u}$$

(SE) (0.119) (0.068) (0.079) (0.130)

$R^2 = 0.98$ Adjusted $R^2 = 0.977$

Biến $\ln X_2$, $\ln X_3$ có ý nghĩa thống kê ở độ tin cậy 99%, trong khi $\ln X_6$ không có ý nghĩa thống kê ngay ở độ tin cậy 90%.

Trong điều kiện các yếu tố khác không đổi, khi X_2 tăng (hoặc giảm) 1% thì Y tăng (hoặc giảm) trung bình 0.481%.

Trong điều kiện các yếu tố khác không đổi, khi X_3 giảm (hoặc tăng) 1% thì Y tăng (hoặc giảm) trung bình 0.351%.

Mô hình hồi quy này giải thích được 97.7% biến thiên của biến $\ln(Y)$

Dependent Variable: LOG(Y)

Method: Least Squares

Date: 01/05/11 Time: 08:22

Sample: 1960 1982

Included observations: 23

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.030	0.119	17.103	0.000
LOG(X2)	0.481	0.068	7.058	0.000
LOG(X3)	-0.351	0.079	-4.416	0.000
LOG(X6)	-0.061	0.130	-0.470	0.644
R-squared	0.980	Mean dependent var		3.664
Adjusted R-squared	0.977	S.D. dependent var		0.188
S.E. of regression	0.028	Akaike info criterion		-4.132
Sum squared resid	0.015	Schwarz criterion		-3.935
Log likelihood	51.521	Hannan-Quinn criter.		-4.083
F-statistic	315.206	Durbin-Watson stat		1.911
Prob(F-statistic)	0.000			