

## Chương trình Giảng dạy Kinh tế Fulbright

Học kỳ Thu năm 2014

### Các Phương Pháp Phân Tích Định Lượng

#### GỢI Ý GIẢI BÀI TẬP 6

#### HỒI QUY ĐƠN BIẾN

Ngày Phát: Thứ Ba 04/11/2014

Ngày Nộp: Thứ Ba 11/11/2014

#### **Đáp án bài 1: (35 điểm)**

a. Trong thuật ngữ “hồi quy tuyến tính”, cụm từ “tuyến tính” được hiểu theo nghĩa nào?

Hồi quy tuyến tính có nghĩa là hồi quy tuyến tính theo các thông số (các  $\beta$ ), nghĩa là các thông số chỉ có số bậc bằng 1. Nó có thể tuyến tính hoặc không tuyến tính theo các biến giải thích.

Các mô hình liệt kê dưới đây có phải là mô hình tuyến tính không

Đối với các mô hình “không tuyến tính”, hãy thử dùng các phép biến đổi toán học thích hợp để chuyển chúng về dạng tuyến tính.

$$a1) Y_i = \beta_1 + \beta_2 \left( \frac{X_i}{\ln X_i} + X_i^2 \right) + u_i$$

Đây là mô hình tuyến tính theo tham số, phi tuyến theo biến số.

$$\text{Để thấy: } Y_i = \beta_1 + \beta_2 X_i' + u_i, X_i' = \left( \frac{X_i}{\ln X_i} + X_i^2 \right)$$

$$a2) Y_i = \beta_1 + e^{\beta_2 / (e^{X_i})} + u_i$$

Đây không là mô hình tuyến tính theo tham số. Ta biến đổi về mô hình tuyến tính như sau:<sup>1</sup>

$$Y_i = \beta_1 + e^{\beta_2 / (e^{X_i})} + u_i$$

$$\Leftrightarrow Y_i - \beta_1 = e^{\beta_2 / (e^{X_i})}$$

$$\Leftrightarrow \ln(Y_i - \beta_1) = \ln(e^{\beta_2 / (e^{X_i})})$$

$$\Leftrightarrow \ln(Y_i - \beta_1) = \beta_2 / (e^{X_i}) + v_i$$

$$\Leftrightarrow Y_i' = \beta_2 X_i' + v_i \text{ với } Y_i' = \ln(Y_i - \beta_1 - u_i) \text{ và } X_i' = 1 / (e^{X_i})$$

$$a3) Y_i = \frac{1}{1 + e^{\beta_1 \beta_2 + \beta_2 X_i}} + u_i$$

Đây không phải là mô hình tuyến tính theo tham số. Ta biến đổi về mô hình tuyến tính như sau:

<sup>1</sup> Vì  $u_i$  là sai số nên khi biến đổi có thể tạm loại bỏ ra khỏi mô hình và cộng thêm trở lại ( $v_i$ ) sau khi đã biến đổi xong. Học viên cũng có thể thêm vào các điều kiện để các phép biến đổi có ý nghĩa về mặt toán học.

$$Y_i = \frac{1}{1 + e^{\beta_1 \beta_2 + \beta_2 X_i}} + u_i$$

$$\Leftrightarrow \frac{1}{Y_i} = 1 + e^{\beta_1 \beta_2 + \beta_2 X_i}$$

$$\Leftrightarrow \ln\left(\frac{1}{Y_i} - 1\right) = \ln(e^{\beta_1 \beta_2 + \beta_2 X_i})$$

$$\Leftrightarrow \ln\left(\frac{1}{Y_i} - 1\right) = \beta_1 \beta_2 + \beta_2 X_i + v_i$$

Đặt  $Y'_i = \ln\left(\frac{1}{Y_i} - 1\right)$ ,  $\beta'_1 = \beta_1 \beta_2 \Rightarrow$  ta chuyển được mô hình ban đầu về mô hình tuyến

tính theo tham số dưới đây:  $Y'_i = \beta'_1 + \beta_2 X_i + v_i$

$$a4) \ln Y_i = \frac{X_i}{\beta_1 + \beta_2 \ln X_i} + u_i$$

Đây không phải là mô hình tuyến tính theo tham số. Ta biến đổi về mô hình tuyến tính như sau:

$$\ln Y_i = \frac{X_i}{\beta_1 + \beta_2 \ln X_i}$$

$$\Leftrightarrow \ln Y_i = \frac{X_i}{\beta_1 + \beta_2 \ln X_i}$$

$$\Leftrightarrow \frac{X_i}{\ln Y_i} = \beta_1 + \beta_2 \ln X_i + v_i$$

$$\Leftrightarrow Y'_i = \beta_2 + \beta_1 X'_i + v_i \text{ với } Y'_i = \frac{X_i}{\ln Y_i} \text{ và } X'_i = \ln X_i$$

$$a5) \ln Y_i = \beta_1 + \ln(\beta_2 X_i) + u_i$$

Đây chưa là mô hình tuyến tính theo tham số. Để chuyển về mô hình tuyến tính theo tham số ta thực hiện như sau:

$$\ln Y_i = \beta_1 + \ln(\beta_2 X_i) + u_i$$

$$\Leftrightarrow \ln Y_i = \beta_1 + \ln(\beta_2) + \ln X_i + \frac{u_i}{i}$$

$$\Leftrightarrow \ln Y_i = \beta'_1 + \ln X_i + u_i \text{ với } \beta'_1 = \beta_1 + \ln(\beta_2)$$

- b. Các khẳng định sau đây có chính xác không? Nếu chúng chỉ đúng một phần thì hãy chỉ ra phần đó. Hãy cẩn thận suy xét các câu trả lời của Anh/Chị.

b1) Các ước lượng bình phương nhỏ nhất thông thường (OLS) cho hệ số góc được ước tính chính xác hơn nếu như các giá trị của X gần với các giá trị trung bình mẫu của chúng hơn.

Trước khi trả lời câu này, nhắc lại sự khác nhau giữa “đúng” (accurate) và “chính xác” (precise) là hữu ích. “Đúng” nghĩa là không chệch; “chính xác” nghĩa là phương sai thấp. Do đó, câu hỏi này là về phương sai của các hàm ước lượng bình phương thông thường nhỏ nhất (OLS).

Phương sai của các hàm ước lượng độ dốc OLS trong mô hình hồi quy đơn giản này là :

$$\text{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}$$

Từ biểu thức này chúng ta thấy rằng phương sai là nhỏ hơn (hàm ước lượng này chính xác hơn) nếu các giá trị của X cách xa giá trị trung bình mẫu hơn.

b2) Các ước lượng không thể là ước lượng không chệch tuyến tính tốt nhất (BLUE) trừ khi các  $u_i$  đều có phân phối chuẩn.

BLUE nghĩa là "Hàm ước lượng không chệch tuyến tính tốt nhất."

Trong bối cảnh này, "tuyến tính" chỉ một hàm ước lượng là một hàm tuyến tính của số hạng sai số ngẫu nhiên trong mô hình này, hoặc là một hàm tuyến tính của biến phụ thuộc của mô hình này. Kiểm tra các hàm ước lượng OLS cho độ dốc và tung độ gốc là đủ để xác lập rằng chúng là tuyến tính. Không yêu cầu tính chuẩn.

Không chệch được thiết lập bằng cách lấy kỳ vọng của hàm ước lượng OLS, là điều mà chúng ta đã làm nhiều lần. Không cần tới tính chuẩn khi chứng minh rằng kỳ vọng này bằng với giá trị thực (nhưng chưa biết) của thông số.

Tốt nhất là dùng Định lý Gauss-Markov. Phép chứng minh định lý này không cần tới tính chuẩn.

Chúng ta thấy rằng phát biểu này là sai.

### **Đáp án bài 2: (20 điểm)**

Gọi  $\hat{\beta}_{YX}$  và  $\hat{\beta}_{XY}$  là các hệ số độ dốc trong hồi quy tương ứng của biến Y theo biến X và của biến X theo biến Y.

a. Hãy chứng minh rằng:  $\hat{\beta}_{YX}\hat{\beta}_{XY} = r^2$  (trong đó r là hệ số tương quan giữa X và Y).

Công thức tính hệ số tương quan r là:

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}} \Leftrightarrow r^2 = \frac{[\sum (X_i - \bar{X})(Y_i - \bar{Y})]^2}{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}$$

$$\hat{\beta}_{YX} \times \hat{\beta}_{XY} = \frac{[\sum (X_i - \bar{X})(Y_i - \bar{Y})]}{\sum (X_i - \bar{X})^2} \cdot \frac{[\sum (X_i - \bar{X})(Y_i - \bar{Y})]}{\sum (Y_i - \bar{Y})^2} = \frac{[\sum (X_i - \bar{X})(Y_i - \bar{Y})]^2}{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}$$

Vì vậy:  $\hat{\beta}_{YX}\hat{\beta}_{XY} = r^2$

b. Giả sử rằng  $\hat{\beta}_{YX}\hat{\beta}_{XY} = 1$ . Điều gì sẽ xảy ra nếu ta hồi quy biến Y theo biến X và của biến X theo biến Y? Hãy giải thích chi tiết.

Ta có một số công thức sau:

$$\hat{\sigma} = \sqrt{\frac{\sum \hat{u}_i^2}{n-2}}$$

$$SE(\hat{\beta}) = \frac{\hat{\sigma}}{\sqrt{\sum (X_i - \bar{X})^2}}$$

$$T - Stat = \frac{\hat{\beta}}{SE(\hat{\beta})}$$

Trong hồi quy đơn, hệ số xác định ( $R^2$ ) bằng với bình phương của hệ số tương quan giữa hai biến  $Y$  &  $X$  ( $r^2$ ). Nếu  $\hat{\beta}_{YX}\hat{\beta}_{XY} = 1$  thì  $R^2$  của hàm hồi quy  $Y$  theo  $X$ , hay  $X$  theo  $Y$  đều bằng 1. Các biến có tương quan tuyến tính hoàn hảo.

Điều này cho thấy trên đồ thị phân tán, tất cả các giá trị thực tế và giá trị dự báo đều nằm trên một đường thẳng. Nói cách khác, đối với hàm hồi quy  $Y$  theo  $X$  (hoặc  $X$  theo  $Y$ ),  $\sum \hat{u}_i^2$  của hàm hồi quy sẽ bằng 0; vì vậy, sai số chuẩn của phần dư cũng bằng không ( $\hat{\sigma} = 0$ ), dẫn đến sai số chuẩn của hệ số độ dốc  $SE(\hat{\beta})$  bằng không; một số phần mềm máy tính tính toán gần đúng những giá trị này, thường rất nhỏ, xấp xỉ 0 để thuận lợi cho việc tính toán T-stat (T-stat =  $\hat{\beta} / SE(\hat{\beta})$ ); khi đó,  $|T\text{-stat}|$  sẽ rất lớn so với t tra bảng để chỉ ra rằng hệ số hồi quy có ý nghĩa thống kê. Lập luận tương tự đối với hệ số  $\hat{\alpha} \dots$

### **Đáp án bài 3: (20 điểm)**

- Giải thích ý nghĩa kinh tế của hệ số  $\beta_2$ , trong mô hình này khuynh hướng tiết kiệm cận biên MPS là bao nhiêu? Giải thích ý nghĩa kinh tế lượng của hệ số này
  - Ý nghĩa kinh tế: hệ số  $\beta_2$  chính là khuynh hướng tiết kiệm cận biên (MPS). Nó cho biết tác động biên của *Thu nhập của hộ gia đình (INC)* lên *Tiết kiệm của hộ gia đình (SAV)*, khi *Thu nhập của hộ gia đình* tăng thêm 1 đơn vị thì về mặt trung bình *Tiết kiệm của hộ gia đình* sẽ tăng  $\beta_2$  đơn vị.
  - Trong mô hình này, Khuynh hướng tiết kiệm cận biên MPS =  $\beta_2 = 0,1504$
  - Giải thích ý nghĩa kinh tế lượng của hệ số này: Trong điều kiện các yếu tố khác không đổi, nếu thu nhập của hộ gia đình tăng lên 1 USD thì trung bình tiết kiệm của hộ gia đình tăng lên 0,1504 USD.
- Tiết kiệm dự đoán của hộ gia đình có thu nhập khả dụng \$26.500 là bao nhiêu

$$\hat{SAV}_i = -176,25 + 0,1504INC_i = -176,25 + 0,1504 * 26.500 = 3809,35 \text{ (USD)}$$

- Trong mô hình hồi quy tuyến tính tiết kiệm theo thu nhập khả dụng, ui có thể chứa những loại yếu tố nào? Hãy liệt kê các yếu tố đó, có kèm theo giải thích.

$u_i$  có thể chứa các loại yếu tố :

- Sự mơ hồ của lý thuyết: lý thuyết tiết kiệm phụ thuộc thu nhập khả dụng là một suy luận hợp lý, tuy nhiên có thể có những biến khác như các yếu tố kì vọng (lạc quan/bi quan) về tình hình phát triển của nền kinh tế, số thành viên phụ thuộc trong gia đình, tình trạng sức khỏe của các thành viên gia đình, tuổi tác của các thành viên trong gia đình.....hay một số biến

chưa biết khác cũng có khả năng ảnh hưởng đến tiết kiệm mà không được đưa vào mô hình,  $u_i$  được sử dụng làm 1 biến thay thế cho tất cả các biến bị bỏ khỏi mô hình đó.

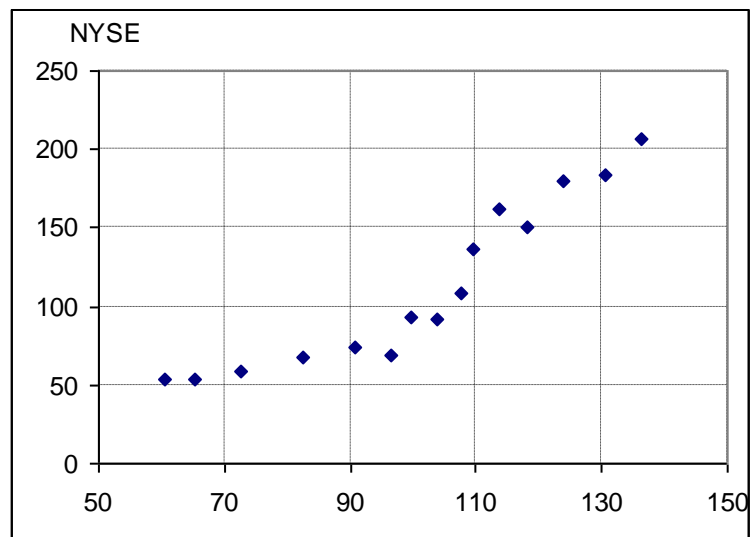
- Dữ liệu không có sẵn: có những biến bị loại bỏ ta có thể xác định được như hiểu biết về đầu tư tài chính, khả năng sinh lợi hấp dẫn từ các kênh tiết kiệm khác nhau... nhưng đôi khi không thể đưa vào mô hình vì khó thu thập được dữ liệu (ví dụ thông tin hiểu biết về đầu tư tài chính là không có) nên phải loại ra khỏi mô hình và đưa vào  $u_i$ .
- Các biến cốt lõi và những biến ngoại vi: có thể các yếu tố liệt kê khác ngoài thu nhập khả dụng cũng ảnh hưởng đến số tiền tiết kiệm, nhưng ảnh hưởng chung của tất cả hay một vài biến này có thể rất nhỏ, thậm chí là không hệ thống và ngẫu nhiên. Nên khi xét về phương diện thực tế và chi phí thì việc đưa chúng vào mô hình là không có ích lợi. Chúng ta cho rằng ảnh hưởng kết hợp chung của chúng có thể xử lý như là biến ngẫu nhiên  $u_i$ .
- Bản chất ngẫu nhiên của con người: dù ta cố gắng đưa tất cả các biến liên quan vào mô hình thì chắc chắn vẫn còn một số ngẫu nhiên thuộc về bản chất của việc quyết định số tiền tiết kiệm của hộ gia đình mà ta không thể giải thích được dù cố gắng đến thế nào đi. Đó là lí do biến nhiễu  $u_i$  xuất hiện để giải thích bản chất ngẫu nhiên này.
- Các biến thay thế kém: 2 biến nghiên cứu là thu nhập khả dụng thường xuyên của hộ gia đình và tiết kiệm thường xuyên của hộ gia đình đó. Tuy nhiên chúng ta chưa có dữ liệu chính xác của 2 biến này mà phải dùng các biến thay thế là thu nhập khả dụng hiện tại của hộ gia đình và tiết kiệm hiện tại của hộ gia đình đó. Vì không biết được thời gian sau đó thu nhập và tiết kiệm của hộ gia đình sẽ thay đổi thế nào nên ta dùng tạm những biến thay thế hiện thời có thể quan sát được. Tuy nhiên những biến thay thế này có thể không tương đương với biến nghiên cứu dẫn đến những sai sót trong tính toán. Số hạng nhiễu  $u_i$  tượng trưng cho những sai sót trong tính toán.
- Nguyên tắc chi li: dù chúng ta muốn đưa thêm 2, 3 biến để giải thích mô hình, tuy nhiên lại muốn giữ cho mô hình càng đơn giản càng tốt, nên nếu lý thuyết không đủ mạnh để đưa thêm biến vào thì sẽ không đưa vào mô hình.  $u_i$  sẽ biểu diễn tất cả các biến đó.
- Dạng hàm sai: có thể dạng hàm  $SAV_i = \hat{\beta}_1 + \hat{\beta}_2 INC_i + u_i$  để biểu diễn quan hệ giữa biến phụ thuộc SAV và biến độc lập INC là chưa chính xác. Có thể tiết kiệm của hộ gia đình không phải là hàm tuyến tính theo biến thu nhập khả dụng của hộ gia đình đó.

VD:  $SAV_i = \hat{\beta}_1 + \hat{\beta}_2 INC_i + \hat{\beta}_3 INC_i^2 + u_i$ . Nếu vậy ảnh hưởng của  $INC_i^2$  sẽ được bao hàm trong  $u_i$ .

#### **Đáp án bài 4: (25 điểm)**

a. Vẽ đồ thị phân tán của hai chỉ số trên và nhận xét:

Nhận xét: Bằng trực quan, ta thấy giữa CPI và NYSE có mối quan hệ đồng biến.



b. Mô hình hồi quy tuyến tính:  $NYSE_t = \alpha + \beta CPI_t + u_t$ .

❖ **Dựa trên các công thức OLS:**

Năm						
t	Y	X	y=Y-Ytb	x=X-Xtb	xy	x <sup>2</sup>
1977	53.69	60.60	-58.85467	-40.18	2364.78	1614.43
1978	53.7	65.20	-58.84467	-35.58	2093.69	1265.94
1979	58.32	72.60	-54.22467	-28.18	1528.05	794.11
1980	68.1	82.40	-44.44467	-18.38	816.89	337.82
1981	74.02	90.90	-38.52467	-9.88	380.62	97.61
1982	68.93	96.50	-43.61467	-4.28	186.67	18.32
1983	92.63	99.60	-19.91467	-1.18	23.50	1.39
1984	92.46	103.90	-20.08467	3.12	-62.66	9.73
1985	108.9	107.60	-3.644667	6.82	-24.86	46.51
1986	136	109.60	23.45533	8.82	206.88	77.79
1987	161.7	113.60	49.15533	12.82	630.17	164.35
1988	149.91	118.30	37.36533	17.52	654.64	306.95
1989	180.02	124.00	67.47533	23.22	1566.78	539.17
1990	183.46	130.70	70.91533	29.92	2121.79	895.21
1991	206.33	136.20	93.78533	35.42	3321.88	1254.58
<b>Tổng</b>					<b>15808.82</b>	<b>7423.924</b>

$$Ytb = 112.5447$$

$$S_{xy} = 15808.82$$

$$Xtb = 100.78$$

$$S_{xx} = 7423.924$$

$$\text{Vậy: } \hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{15.808,82}{7.423,924} = 2,129$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = -102,061$$

❖ **Sử dụng các lệnh SLOPE và INTERCEPT trong EXCEL:**

$$\hat{\alpha} = \text{INTERCEPT}(\text{NYSE}_t, \text{CPI}_t) = -102,061$$

$$\hat{\beta} = \text{SLOPE}(\text{NYSE}_t, \text{CPI}_t) = 2,129$$

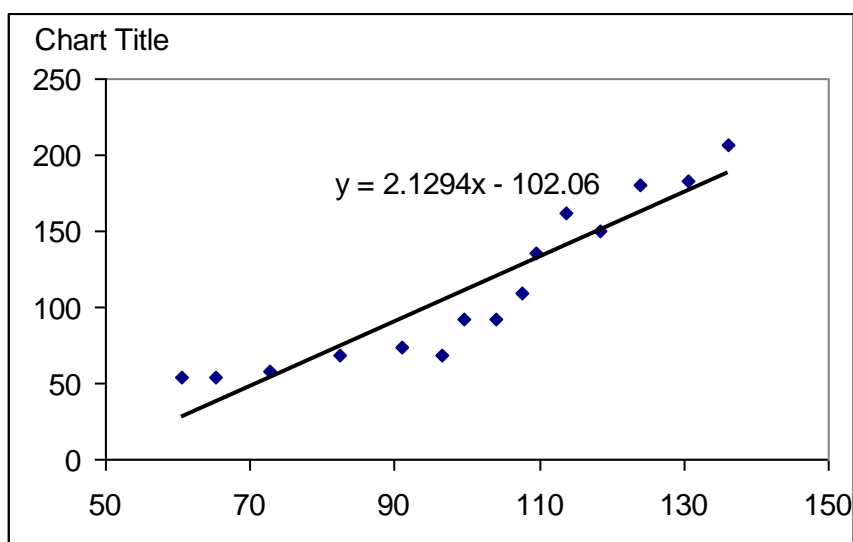
❖ **Sử dụng công cụ DATA ANALYSIS trong EXCEL:**

Dùng lệnh TOOLS/ DATA ANALYSIS/ REGRESSION trong EXCEL ta được:

SUMMARY OUTPUT						
<i>Regression Statistics</i>						
Multiple R	0.931681352					
R Square	0.868030141					
Adjusted R Square	0.857878613					
Standard Error	19.84179586					
Observations	15					
<i>ANOVA</i>						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	1	33663.97216	33663.97216	85.50734163	4.42725E-07	
Residual	13	5118.059218	393.6968629			
Total	14	38782.03137				
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	-102.06055	23.76678031	-4.29425226	0.000872516	-153.405547	-50.7155527
CPI	2.129442516	0.230284241	9.247017986	4.42725E-07	1.631943756	2.626941276

Vậy  $\hat{\alpha} = -102,061$ ,  $\hat{\beta} = 2,129$

❖ **Dùng công cụ ADD TRENDLINE trong EXCEL:**



- c. Viết phương trình hồi quy ước lượng của mô hình trên. Giải thích ý nghĩa của hệ số độ dốc và tung độ gốc:

$$\text{Vậy : } NYSE_t = -102,061 + 2,129CPI_t + \hat{u}_t$$

Ý nghĩa độ dốc:  $\hat{\beta} = 2,129$  có nghĩa: Theo dữ liệu của mẫu, trong thời kỳ 1977-1991 ở Mỹ, trong điều kiện các yếu tố khác không đổi, chỉ số giá tiêu dùng (CPI) tăng lên 1 đơn vị thì chỉ số chứng khoán trên thị trường chứng khoán New York (NYSE) trung bình sẽ tăng 2,129 đơn vị.

- d. Hãy tính hệ số xác định  $R^2$ . Hệ số này nói lên điều gì ?

$$R^2 = 0.8680$$

Ý nghĩa: Trong mô hình, 86,8% sự thay đổi của chỉ số chứng khoán trên thị trường chứng khoán New York (NYSE) được giải thích bởi chỉ số giá tiêu dùng (CPI).