

Chương 3

HỒI QUY TUYẾN TÍNH BỘI

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

1. Hàm hồi quy tổng thể (PRF)

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$$

Trong đó

- Y là biến phụ thuộc
- X_2, X_3 là các biến độc lập
- X_{2i}, X_{3i} là giá trị thực tế của X_2, X_3
- U_i là các sai số ngẫu nhiên

Vậy ý nghĩa của $\beta_1, \beta_2, \beta_3$ là gì ?

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

2. Các giả thiết của mô hình

- Các X_{2i}, X_{3i} cho trước và không ngẫu nhiên
- Giá trị trung bình của đại lượng ngẫu nhiên U_i bằng 0, Phương sai của U_i không thay đổi
- Không có sự tương quan giữa các U_i
- Không có sự tương quan (cộng tuyến) giữa X_2 và X_3
- Không có sự tương quan giữa các U_i và X_2, X_3

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

3. Ước lượng các tham số

Chúng ta sử dụng phương pháp bình phương nhỏ nhất OLS

$$PRF : Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$$

Hàm hồi quy mẫu tương ứng sẽ là :

$$SRF : Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + e_i$$

Hay:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$$

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

$$e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}$$

Theo nguyên lý của phương pháp OLS thì các tham số $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ được chọn sao cho

$$\sum e_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2 \rightarrow \min$$

Như vậy, công thức tính của các tham số như sau :

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

$$\text{Ký hiệu: } y_i = Y_i - \bar{Y} \quad x_{2i} = X_{2i} - \bar{X}_2 \quad x_{3i} = X_{3i} - \bar{X}_3$$

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})(\sum y_i x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum x_{2i} x_{3i})(\sum y_i x_{2i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$$

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

Người ta chứng minh được

$$\sum x_{2i}^2 = \sum X_{2i}^2 - n(\bar{X}_2)^2$$

$$\sum x_{3i}^2 = \sum X_{3i}^2 - n(\bar{X}_3)^2$$

$$\sum y_i^2 = \sum Y_i^2 - n(\bar{Y})^2$$

$$\sum x_{2i}x_{3i} = \sum X_{2i}X_{3i} - n\bar{X}_2\bar{X}_3$$

$$\sum y_ix_{2i} = \sum Y_iX_{2i} - n\bar{Y}\bar{X}_2$$

$$\sum y_ix_{3i} = \sum Y_iX_{3i} - n\bar{Y}\bar{X}_3$$

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

Ví dụ minh họa

Bảng dưới đây cho các số liệu về doanh số bán (Y), chi phí chào hàng (X₂) và chi phí quảng cáo (X₃) của một công ty

Hãy ước lượng hàm hồi quy tuyến tính của doanh số bán theo chi phí chào hàng và chi phí quảng cáo

Doanh số bán Y _i (trđ)	Chi phí chào hàng X ₂	Chi phí quảng cáo X ₃
1270	100	180
1490	106	248
1060	60	190
1626	160	240
1020	70	150
1800	170	260
1610	140	250
1280	120	160
1390	116	170
1440	120	230
1590	140	220
1380	150	150

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

Giải Từ số liệu trên, ta tính được các tổng như sau :

$$\begin{aligned} \sum Y_i &= 16956 & \sum X_{2i}^2 &= 188192 \\ \sum X_{2i} &= 1452 & \sum X_{2i}X_{3i} &= 303608 \\ \sum X_{3i} &= 2448 & \sum X_{3i}^2 &= 518504 \\ \sum Y_i^2 &= 24549576 & \bar{Y} &= 1413 \\ \sum Y_iX_{3i} &= 3542360 & \bar{X}_2 &= 121 \\ \sum Y_iX_{2i} &= 2128740 & \bar{X}_3 &= 204 \end{aligned}$$

Có thể dùng Excel để tính toán các số liệu này, như sau

Y _i	X _{2i}	X _{3i}
1270	100	180
1490	106	248
1060	60	190
1626	160	240
1020	70	150
1800	170	260
1610	140	250
1280	120	160
1390	116	170
1440	120	230
1590	140	220
1380	150	150
16956	1452	2448
1413	121	204

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

$$\begin{aligned} \sum y_i^2 &= \sum Y_i^2 - n(\bar{Y})^2 = \\ \sum x_{2i}^2 &= \sum X_{2i}^2 - n(\bar{X}_2)^2 = \\ \sum x_{3i}^2 &= \sum X_{3i}^2 - n(\bar{X}_3)^2 = \\ \sum y_ix_{2i} &= \sum Y_iX_{2i} - n\bar{Y}\bar{X}_2 = \\ \sum y_ix_{3i} &= \sum Y_iX_{3i} - n\bar{Y}\bar{X}_3 = \\ \sum x_{2i}x_{3i} &= \sum X_{2i}X_{3i} - n\bar{X}_2\bar{X}_3 = \end{aligned}$$

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

$$\hat{\beta}_2 = \text{_____} =$$

$$\hat{\beta}_3 = \text{_____} =$$

$$\hat{\beta}_1 =$$

Vậy

$$\hat{Y}_i = ? + ? X_{2i} + ? X_{3i}$$

Kết quả của ví dụ trên chạy bằng Eviews như sau :

Equation: EQ01 Workfile: VIDUBABIEN::Untitled1				
View Proc Object Print Name Freeze Estimate Forecast Stats Resids				
Dependent Variable: Y				
Method: Least Squares				
Date: 03/21/09 Time: 09:41				
Sample: 1 12				
Included observations: 12				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	328.1383	71.99136	4.558023	0.0014
X2	4.649510	0.469146	9.910592	0.0000
X3	2.560152	0.379411	6.747707	0.0001
R-squared	0.967693	Mean dependent var	1413.000	
Adjusted R-squared	0.960514	S.D. dependent var	231.7420	
S.E. of regression	46.04989	Akaike info criterion	10.70965	
Sum squared resid	19085.33	Schwarz criterion	10.83087	
Log likelihood	-61.25787	F-statistic	134.7884	
Durbin-Watson stat	2.493309	Prob(F-statistic)	0.000000	

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

4. Hệ số xác định của mô hình

$$TSS = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - n\bar{Y}^2$$

$$ESS = \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}$$

$$RSS = TSS - ESS$$

$$R^2 = \frac{ESS}{TSS}$$

Vì sao khi thêm biến vào mô hình thì R^2 sẽ tăng lên? => [Bài tập](#)

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

4. Hệ số xác định của mô hình

Đối với mô hình hồi quy bội, người ta tính R^2 có hiệu chỉnh như sau :

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k}$$

k là số tham số trong mô hình

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

4. Hệ số xác định của mô hình

\bar{R}^2 có các đặc điểm sau :

❖ Khi $k > 1$ thì $\bar{R}^2 \leq R^2 \leq 1$

❖ \bar{R}^2 có thể âm, và khi nó âm, coi như bằng 0

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

4. Hệ số xác định của mô hình

Ví dụ : Tính hệ số xác định của mô hình hồi quy theo số liệu của ví dụ trước

$$TSS = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - n\bar{Y}^2 \Rightarrow TSS =$$

$$ESS = \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} \Rightarrow ESS =$$

$$RSS = TSS - ESS$$

$$\Rightarrow RSS =$$

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

4. Hệ số xác định của mô hình

$$R^2 = \frac{ESS}{TSS} = \frac{\sum \hat{y}_i^2}{\sum y_i^2} =$$

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k} =$$

Kết quả của ví dụ trên chạy bằng Eviews như sau :

Equation: EQ01 Workfile: VIDUBABIEN::Untitled1				
View Proc Object Print Name Freeze Estimate Forecast Stats Resids				
Dependent Variable: Y				
Method: Least Squares				
Date: 03/21/09 Time: 09:41				
Sample: 1 12				
Included observations: 12				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	328.1383	71.99136	4.558023	0.0014
X2	4.649510	0.469146	9.910592	0.0000
X3	2.560152	0.379411	6.747707	0.0001
R-squared	0.967693	Mean dependent var	1413.000	
Adjusted R-squared	0.960514	S.D. dependent var	231.7420	
S.E. of regression	46.04989	Akaike info criterion	10.70965	
Sum squared resid	19085.33	Schwarz criterion	10.83087	
Log likelihood	-61.25787	F-statistic	134.7884	
Durbin-Watson stat	2.493309	Prob(F-statistic)	0.000000	

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

5. Phương sai của hệ số hồi quy

Phương sai của các tham số hồi quy được tính theo các công thức sau:

$$\sigma_{\hat{\beta}_1}^2 = \hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{X}_2^2 \sum x_{3i}^2 + \bar{X}_3^2 \sum x_{2i}^2 - 2\bar{X}_2\bar{X}_3 \sum x_{2i}x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i}x_{3i})^2} \right]$$

$$se(\hat{\beta}_1) = \sqrt{\sigma_{\hat{\beta}_1}^2}$$

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

4. Phương sai của hệ số hồi quy

$$\sigma_{\hat{\beta}_2}^2 = \hat{\sigma}^2 \left[\frac{\sum x_{3i}^2}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i}x_{3i})^2} \right]$$

$$se(\hat{\beta}_2) = \sqrt{\sigma_{\hat{\beta}_2}^2}$$

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

5. Phương sai của hệ số hồi quy

$$\sigma_{\hat{\beta}_3}^2 = \hat{\sigma}^2 \left[\frac{\sum x_{2i}^2}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i}x_{3i})^2} \right]$$

$$se(\hat{\beta}_3) = \sqrt{\sigma_{\hat{\beta}_3}^2}$$

$$\text{Với } \hat{\sigma}^2 = \frac{RSS}{n-3}$$

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

6. Khoảng tin cậy của các hệ số hồi quy

Khoảng tin cậy của β_1 Với độ tin cậy là $1-\alpha$

$$\left(\hat{\beta}_1 - t_{\frac{\alpha}{2}} \times se(\hat{\beta}_1); \hat{\beta}_1 + t_{\frac{\alpha}{2}} \times se(\hat{\beta}_1) \right)$$

Khoảng tin cậy của β_2 Với độ tin cậy là $1-\alpha$

$$\left(\hat{\beta}_2 - t_{\frac{\alpha}{2}} \times se(\hat{\beta}_2); \hat{\beta}_2 + t_{\frac{\alpha}{2}} \times se(\hat{\beta}_2) \right)$$

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

6. Khoảng tin cậy của các hệ số hồi quy

Khoảng tin cậy của β_3 Với độ tin cậy là $1-\alpha$

$$\left(\hat{\beta}_3 - t_{\frac{\alpha}{2}} \times se(\hat{\beta}_3); \hat{\beta}_3 + t_{\frac{\alpha}{2}} \times se(\hat{\beta}_3) \right)$$

Lưu ý khi tra bảng T-Student, trong trường hợp hàm hồi quy 3 biến thì bậc tự do là **(n-3)**

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

6. Khoảng tin cậy của các hệ số hồi quy

Ví dụ : Tính khoảng tin cậy của β_2 và β_3 mô hình hồi quy theo số liệu của ví dụ trước với độ tin cậy 95%

Giải: tra bảng T-Student bậc tự do (n-3)=12-3=9

$$t_{0,025} =$$

$$\hat{\sigma}^2 = \frac{RSS}{n-3} =$$

$$\sigma_{\hat{\beta}_2}^2 = \Rightarrow se(\hat{\beta}_2) = \sqrt{\sigma_{\hat{\beta}_2}^2} =$$

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

6. Khoảng tin cậy của các hệ số hồi quy

Khoảng tin cậy của β_2 là $(? < \beta_2 < ?)$

$$\sigma_{\hat{\beta}_3}^2 = \Rightarrow se(\hat{\beta}_3) = \sqrt{\sigma_{\hat{\beta}_3}^2} =$$

Khoảng tin cậy của β_3 là $(? < \beta_3 < ?)$

Kết quả của ví dụ trên chạy bằng Eviews như sau :

Equation: EQ01 Workfile: VDIUBABIEN::Untitled1				
View Proc Object Print Name Freeze Estimate Forecast Stats Resids				
Dependent Variable: Y				
Method: Least Squares				
Date: 03/21/09 Time: 09:41				
Sample: 1 12				
Included observations: 12				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	329.1383	71.99136	4.558023	0.0014
X2	4.649510	0.469146	9.910592	0.0000
X3	2.560152	0.379411	6.747707	0.0001
R-squared	0.967693	Mean dependent var	1413.000	
Adjusted R-squared	0.960514	S.D. dependent var	231.7420	
S.E. of regression	46.04989	Akaike info criterion	10.70965	
Sum squared resid	19085.33	Schwarz criterion	10.83087	
Log likelihood	-61.25787	F-statistic	134.7894	
Durbin-Watson stat	2.493309	Prob(F-statistic)	0.000000	

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

7. Kiểm định giả thiết

a) Kiểm định giả thiết về $\beta_1, \beta_2, \beta_3$

$$H_0: \beta_i = \beta_0 \quad \text{Độ tin cậy là } 1-\alpha$$

$$H_1: \beta_i \neq \beta_0$$

Bước 1: Lập khoảng tin cậy

Bước 2: Nếu β_0 thuộc khoảng tin cậy thì chấp nhận H_0 . Nếu β_0 không thuộc khoảng tin cậy thì bác bỏ H_0 .

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

7. Kiểm định giả thiết

a) Kiểm định giả thiết về $\beta_1, \beta_2, \beta_3$

Ví dụ : (theo số liệu trước), yêu cầu kiểm định các giả thiết

$$H_0: \beta_2 = 0 \quad H_0: \beta_3 = 0$$

$$H_1: \beta_2 \neq 0 \quad H_1: \beta_3 \neq 0$$

Với độ tin cậy 95%

Kết quả của ví dụ trên chạy bằng Eviews như sau :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	328.1383	71.99136	4.558023	0.0014
X2	4.649510	0.469146	9.910592	0.0000
X3	2.560152	0.379411	6.747707	0.0001

R-squared	0.967693	Mean dependent var	1413.000
Adjusted R-squared	0.960514	S.D. dependent var	231.7420
S.E. of regression	46.04989	Akaike info criterion	10.70965
Sum squared resid	19085.33	Schwarz criterion	10.83087
Log likelihood	-61.25787	F-statistic	134.7884
Durbin-Watson stat	2.493309	Prob(F-statistic)	0.000000

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

7. Kiểm định giả thiết

b) Kiểm định giả thiết về R^2

$$H_0: R^2 = 0 \quad \text{Độ tin cậy là } 1 - \alpha$$

$$H_1: R^2 \neq 0$$

Bước 1: tính $F = \frac{R^2(n-3)}{2(1-R^2)}$

Bước 2: Tra bảng tìm $F(2, n-3)$, mức ý nghĩa là α

Bước 3: Nếu $F > F(2, n-3)$, bác bỏ H_0 . Nếu $F \leq F(2, n-3)$, chấp nhận H_0 .

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

7. Kiểm định giả thiết

b) Kiểm định giả thiết về R^2

Ví dụ : Yêu cầu kiểm định giả thiết

$$H_0: R^2 = 0 \quad \text{Độ tin cậy là 95\%}$$

$$H_1: R^2 \neq 0$$

Giải :

$$F = \frac{R^2(n-3)}{2(1-R^2)}$$

$$F(2,9) = 4,26 \quad (\alpha = 0,05)$$

Vì $F > F(2,9)$ nên

Kết quả của ví dụ trên chạy bằng Eviews như sau :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	328.1383	71.99136	4.558023	0.0014
X2	4.649510	0.469146	9.910592	0.0000
X3	2.560152	0.379411	6.747707	0.0001

R-squared	0.967693	Mean dependent var	1413.000
Adjusted R-squared	0.960514	S.D. dependent var	231.7420
S.E. of regression	46.04989	Akaike info criterion	10.70965
Sum squared resid	19085.33	Schwarz criterion	10.83087
Log likelihood	-61.25787	F-statistic	134.7884
Durbin-Watson stat	2.493309	Prob(F-statistic)	0.000000

II. MỘT SỐ DẠNG HÀM

1. Hàm sản xuất Cobb-Douglas

Hàm sản xuất Cobb-Douglas được biểu diễn như sau:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{U_i}$$

Trong đó : Y_i : sản lượng của doanh nghiệp
 X_{2i} : lượng vốn
 X_{3i} : lượng lao động
 U_i : sai số ngẫu nhiên

Hàm sản xuất Cobb-Douglas có thể đưa được về dạng tuyến tính bằng cách lấy logarit hai vế

II. MỘT SỐ DẠNG HÀM

1. Hàm sản xuất Cobb-Douglas

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + U_i$$

Đặt $Y_i^* = \ln Y_i$

$$\beta_1^* = \ln \beta_1$$

$$X_{2i}^* = \ln X_{2i}$$

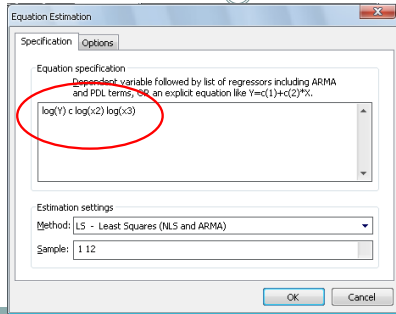
$$X_{3i}^* = \ln X_{3i}$$

Dạng tuyến tính sẽ là :

$$Y_i^* = \beta_1^* + \beta_2 X_{2i}^* + \beta_3 X_{3i}^* + U_i$$

Để hồi quy dạng tuyến tính logarit trong Eviews, ta nhập phương trình hồi quy như sau :

$$\ln Y_i = \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + U_i$$



Kết quả hồi quy

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.512994	0.221228	15.87952	0.0000
LOG(X2)	0.385488	0.030092	12.81033	0.0000
LOG(X3)	0.357654	0.045624	7.839189	0.0000

R-squared	0.977179	Mean dependent var	7.240475
Adjusted R-squared	0.972108	S.D. dependent var	0.170894
S.E. of regression	0.028541	Akaike info criterion	-4.062630
Sum squared resid	0.007331	Schwarz criterion	-3.941404
Log likelihood	27.37578	Hannan-Quinn criter.	-4.107513
F-statistic	192.6867	Durbin-Watson stat	2.878392
Prob(F-statistic)	0.000000		

II. MỘT SỐ DẠNG HÀM

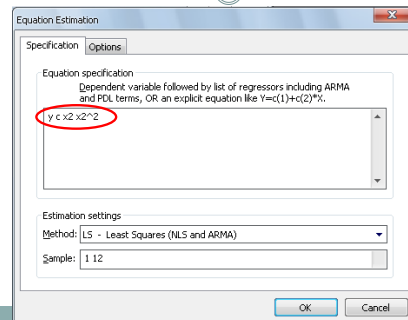
2. Hàm hồi quy đa thức bậc 2

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + U_i$$

Mặc dù chỉ có một biến độc lập X_i nhưng nó xuất hiện với các lũy thừa khác nhau khiến cho mô hình trở thành hồi quy ba biến

Để hồi quy dạng đa thức trong Eviews

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + U_i$$



Kết quả hồi quy dạng đa thức

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	625.0366	376.3580	1.660750	0.1311
X2	6.962680	6.806957	1.022877	0.3331
X2^2	-0.003477	0.029341	-0.118488	0.9083

R-squared	0.804554	Mean dependent var	1413.000
Adjusted R-squared	0.761122	S.D. dependent var	231.7420
S.E. of regression	113.2642	Akaike info criterion	12.50964
Sum squared resid	115459.1	Schwarz criterion	12.63087
Log likelihood	-72.05785	Hannan-Quinn criter.	12.46476
F-statistic	18.52431	Durbin-Watson stat	1.917414
Prob(F-statistic)	0.000645		

Để chuẩn bị tốt cho buổi học sau, đề nghị sinh viên tự ôn tập lại kiến thức về **ma trận** gồm : các phép toán ma trận (**cộng, chuyển vị, nhân 2 ma trận**) ; tính **định thức** ; tìm **ma trận nghịch đảo**. Giảng viên sẽ **hỏi** phần này trên lớp trước khi vào bài mới

III. HỒI QUY TUYẾN TÍNH K BIẾN

1. Hàm hồi quy tổng thể (PRF)

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + U_i$$

Trong đó

- Y là biến phụ thuộc
- X_2, X_3, \dots, X_k là các biến độc lập
- U_i là các sai số ngẫu nhiên

• β_1 : Hệ số tự do

$\beta_2, \beta_3, \dots, \beta_k$ là các hệ số hồi quy riêng

III. HỒI QUY TUYẾN TÍNH K BIẾN

1. Hàm hồi quy tổng thể (PRF)

Quan sát thứ 1 :

$$Y_1 = \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \dots + \beta_k X_{k1} + U_1$$

Quan sát thứ 2 :

$$Y_2 = \beta_1 + \beta_2 X_{22} + \beta_3 X_{32} + \dots + \beta_k X_{k2} + U_2$$

Quan sát thứ n :

$$Y_n = \beta_1 + \beta_2 X_{2n} + \beta_3 X_{3n} + \dots + \beta_k X_{kn} + U_n$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

1. Hàm hồi quy tổng thể (PRF)

Ký hiệu

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{pmatrix} \quad U = \begin{pmatrix} U_1 \\ U_2 \\ \dots \\ U_n \end{pmatrix}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

$$X = \begin{pmatrix} 1 & X_{21} & X_{31} & \dots & X_{k1} \\ 1 & X_{22} & X_{32} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{2n} & X_{3n} & \dots & X_{kn} \end{pmatrix}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

Ta có

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{21} & \dots & X_{k1} \\ 1 & X_{22} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots \\ 1 & X_{2n} & \dots & X_{kn} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{pmatrix} + \begin{pmatrix} U_1 \\ U_2 \\ \dots \\ U_n \end{pmatrix}$$

$$\Rightarrow PRF : Y = X \cdot \beta + U$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

2. Các giả thiết của mô hình

Giả thiết 1 : Các biến độc lập X_2, X_3, \dots, X_k không ngẫu nhiên

Giả thiết 2 : Các sai số ngẫu nhiên U_i có giá trị trung bình bằng 0 và có phương sai không thay đổi

$$E(U_i | X) = 0 \quad \text{Var}(U_i | X) = \sigma^2$$

Giả thiết 3 : Không có sự tương quan giữa các sai số U_i

$$\text{Cov}(U_i, U_j | X) = 0, i \neq j$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

2. Các giả thiết của mô hình

Giả thiết 4 : Không có hiện tượng cộng tuyến giữa các biến độc lập X_2, X_3, \dots, X_k

Giả thiết 5 : Không có tương quan giữa các biến độc lập X_2, X_3, \dots, X_k với các sai số ngẫu nhiên U_i

$$\text{Cov}(U, X) = 0$$

Mô hình hồi quy tuyến tính bội

$$Y = X\beta + U$$

$$\text{VarCov}(U) = \sigma^2 I_n$$

$$\text{rank}(X) = k$$

$$E(U_i | X) = 0$$

Vì sao ? => Bài tập cộng điểm

$$\begin{aligned} \text{Cov}(\varepsilon_i, v_i) &= E[(\varepsilon_i - E(\varepsilon_i))' [v_i - E(v_i)]] \\ \text{Gợi ý : } \text{VarCov}(\varepsilon) &= E[(\varepsilon - E(\varepsilon))' (\varepsilon - E(\varepsilon))] \end{aligned}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

3. Ước lượng các tham số

Hàm hồi quy mẫu :

SRF: $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} + e_i$

hoặc: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki}$

Hay : (Viết dưới dạng ma trận)

$$Y = X\hat{\beta} + e$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

3. Ước lượng các tham số

Với

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \dots \\ \hat{\beta}_k \end{pmatrix} \quad e = \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

SRF: $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} + e_i$

hoặc: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki}$

Khi đó

$$e_i = (Y_i - \hat{Y}_i)$$

$$= Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_k X_{ki}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

Theo nguyên lý của phương pháp OLS thì các tham số $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \dots, \hat{\beta}_k$ được chọn sao cho

$$\begin{aligned} \sum e_i^2 &= \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - X_i' \hat{\beta})^2 \\ &= \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_k X_{ki})^2 \\ &\rightarrow \min \end{aligned}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

Khi đó :

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Vì sao? => Bài tập cộng điểm

III. HỒI QUY TUYẾN TÍNH K BIẾN

$$X^T X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots \\ X_{k1} & X_{k2} & \dots & X_{kn} \end{pmatrix} \begin{pmatrix} 1 & X_{21} & \dots & X_{k1} \\ 1 & X_{22} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots \\ 1 & X_{2n} & \dots & X_{kn} \end{pmatrix}$$

$$= \begin{pmatrix} n & \sum X_{2i} & \dots & \sum X_{ki} \\ \sum X_{2i} & \sum X_{2i}^2 & \dots & \sum X_{2i} X_{ki} \\ \dots & \dots & \dots & \dots \\ \sum X_{ki} & \sum X_{2i} X_{ki} & \dots & \sum X_{ki}^2 \end{pmatrix}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

$$X^T Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots \\ X_{k1} & X_{k2} & \dots & X_{kn} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix} = \begin{pmatrix} \sum Y_i \\ \sum X_{2i} Y_i \\ \dots \\ \sum X_{ki} Y_i \end{pmatrix}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

3. Ví dụ minh họa

Bảng dưới đây cho các số liệu về lượng hàng bán được của một loại hàng hóa (Y), thu nhập của người tiêu dùng (X_2) và giá bán của loại hàng này (X_3)

Tìm hàm hồi quy tuyến tính

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$$

Y_i (tấn/tháng)	X_2 (triệu đồng/năm)	X_3 (ngàn đồng/kg)
20	8	2
18	7	3
19	8	4
18	8	4
17	6	5
17	6	5
16	5	6
15	5	7
13	4	8
12	3	8

III. HỒI QUY TUYẾN TÍNH K BIẾN

Giải Từ số liệu trên, ta tính được các tổng như sau :

$$\begin{aligned} \sum Y_i &= 165 & \sum X_{2i}^2 &= 388 \\ \sum X_{2i} &= 60 & \sum X_{2i} X_{3i} &= 282 \\ \sum X_{3i} &= 52 & \sum X_{3i}^2 &= 308 \\ \sum Y_i^2 &= 2781 & \bar{Y} &= 16,5 \\ \sum Y_i X_{3i} &= 813 & \bar{X}_2 &= 6 \\ \sum Y_i X_{2i} &= 1029 & \bar{X}_3 &= 5,2 \end{aligned}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

$$X^T X = \begin{pmatrix} n & \sum X_{2i} & \sum X_{3i} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i} X_{3i} \\ \sum X_{3i} & \sum X_{3i} X_{2i} & \sum X_{3i}^2 \end{pmatrix} = \begin{pmatrix} 10 & 60 & 52 \\ 60 & 388 & 282 \\ 52 & 282 & 308 \end{pmatrix}$$

$$(X^T X)^{-1} = \begin{pmatrix} 26.165 & -2.497 & -2.131 \\ -2.497 & 0.246 & 0.196 \\ -2.131 & 0.196 & 0.183 \end{pmatrix}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

$$X^T Y = \begin{pmatrix} \sum Y_i \\ \sum Y_i X_{2i} \\ \sum Y_i X_{3i} \end{pmatrix} = \begin{pmatrix} 165 \\ 1028 \\ 813 \end{pmatrix}$$

Các hệ số hồi quy này có ý nghĩa gì ?

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \begin{pmatrix} 14.992 \\ 0.762 \\ -0.589 \end{pmatrix}$$

yây: $\hat{\beta}_1 = 14,992$
 $\hat{\beta}_2 = 0,762$
 $\hat{\beta}_3 = -0,589$

$$\hat{Y}_i = 14,992 + 0,762 X_{2i} - 0,589 X_{3i}$$

Kết quả hồi quy bằng Eviews như sau :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.99215	2.922713	5.129531	0.0014
X2	0.761780	0.283438	2.687639	0.0312
X3	-0.589005	0.244593	-2.408105	0.0469

R-squared	0.960934	Mean dependent var	16.50000
Adjusted R-squared	0.949773	S.D. dependent var	2.549510
S.E. of regression	0.571382	Akaike info criterion	1.981807
Sum squared resid	2.285340	Schwarz criterion	2.052582
Log likelihood	-6.809035	Hannan-Quinn criter.	1.862226
F-statistic	86.09278	Durbin-Watson stat	1.800073
Prob(F-statistic)	0.000012		

III. HỒI QUY TUYẾN TÍNH K BIẾN

4. Hệ số xác định của mô hình

$$TSS = Y^T Y - n(\bar{Y})^2$$

$$ESS = \hat{\beta}^T X^T Y - n(\bar{Y})^2$$

$$RSS = TSS - ESS$$

Hệ số xác định: $R^2 = \frac{ESS}{TSS}$

Hệ số xác định hiệu chỉnh: $\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k}$

Kết quả hồi quy bằng Eviews như sau :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.99215	2.922713	5.129531	0.0014
X2	0.761780	0.283438	2.687639	0.0312
X3	-0.589005	0.244593	-2.408105	0.0469

R-squared	0.960934	Mean dependent var	16.50000
Adjusted R-squared	0.949773	S.D. dependent var	2.549510
S.E. of regression	0.571382	Akaike info criterion	1.981807
Sum squared resid	2.285340	Schwarz criterion	2.052582
Log likelihood	-6.809035	Hannan-Quinn criter.	1.862226
F-statistic	86.09278	Durbin-Watson stat	1.800073
Prob(F-statistic)	0.000012		

III. HỒI QUY TUYẾN TÍNH K BIẾN

4. Khoảng tin cậy và kiểm định giả thiết

$$VarCov(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$$

Vì sao? => Bài tập cộng điểm

III. HỒI QUY TUYẾN TÍNH K BIẾN

4. Khoảng tin cậy và kiểm định giả thiết

Gọi c_{jj} là phần tử nằm ở dòng j cột j của ma trận $(X^T X)^{-1}$

$$\text{Khi đó: } \sigma_{\hat{\beta}_j}^2 = \sigma^2 \cdot c_{jj} \approx \hat{\sigma}^2 \cdot c_{jj}$$

$$se(\hat{\beta}_j) = \sqrt{\sigma_{\hat{\beta}_j}^2}$$

$$\text{Với } \hat{\sigma}^2 = \frac{RSS}{n-k} \quad (k \text{ là số tham số})$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

4. Khoảng tin cậy và kiểm định giả thiết

Khoảng tin cậy của β_j là

$$(\hat{\beta}_j - t_{\frac{\alpha}{2}} se(\hat{\beta}_j); \hat{\beta}_j + t_{\frac{\alpha}{2}} se(\hat{\beta}_j))$$

Hoặc tính giá trị tới hạn của β_j là

$$t = \frac{\hat{\beta}_j - \beta_j^*}{se(\hat{\beta}_j)} \quad \text{Bậc tự do là } (n-k)$$

Kết quả hồi quy bằng Eviews như sau :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.99215	2.922713	5.129531	0.0014
X2	0.761780	0.283438	2.687639	0.0312
X3	-0.589005	0.244593	-2.408105	0.0469

R-squared	0.960934	Mean dependent var	16.50000
Adjusted R-squared	0.949773	S.D. dependent var	2.549510
S.E. of regression	0.571382	Akaike info criterion	1.981807
Sum squared resid	2.285340	Schwarz criterion	2.052582
Log likelihood	-6.809035	Hannan-Quinn criter.	1.862226
F-statistic	86.09278	Durbin-Watson stat	1.800073
Prob(F-statistic)	0.000012		

III. HỒI QUY TUYẾN TÍNH K BIẾN

4. Khoảng tin cậy và kiểm định giả thiết

Kiểm định giả thiết về R^2 $H_0: R^2 = 0$

Với độ tin cậy $1-\alpha$ $H_1: R^2 \neq 0$

Bước 1: tính $F = \frac{R^2(n-k)}{(k-1)(1-R^2)}$

Bước 2: Tra bảng tìm $F(k-1, n-k)$, mức ý nghĩa là α

Bước 3: Nếu $F > F(k-1, n-k)$, bác bỏ H_0 . Nếu $F \leq F(k-1, n-k)$, chấp nhận H_0

Kết quả hồi quy bằng Eviews như sau :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.99215	2.922713	5.129531	0.0014
X2	0.761780	0.283438	2.687639	0.0312
X3	-0.589005	0.244593	-2.408105	0.0469

R-squared	0.960934	Mean dependent var	16.50000
Adjusted R-squared	0.949773	S.D. dependent var	2.549510
S.E. of regression	0.571382	Akaike info criterion	1.981807
Sum squared resid	2.285340	Schwarz criterion	2.052582
Log likelihood	-6.809035	Hannan-Quinn criter.	1.862226
F-statistic	86.09278	Durbin-Watson stat	1.800073
Prob(F-statistic)	0.000012		

Một vài kết quả hồi quy khác bằng Eviews

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.477215	2.207579	-1.575126	0.1180
X2_CNMS	0.080708	0.444005	0.181772	0.8561
X3_COM	0.543629	0.382117	1.422675	0.1576
X4_SUA	0.005008	0.001267	3.951200	0.0001
X5_CD	0.086614	0.009635	8.989518	0.0000
X6_RAUI	0.011712	0.015035	0.779036	0.4376
X7_CB	0.143078	0.019623	7.291392	0.0000

R-squared	0.702570	Mean dependent var	17.42500
Adjusted R-squared	0.686778	S.D. dependent var	3.364902
S.E. of regression	1.883210	Akaike info criterion	4.160395
Sum squared resid	400.7523	Schwarz criterion	4.322989
Log likelihood	-242.6237	Hannan-Quinn criter.	4.228430
F-statistic	44.48696	Durbin-Watson stat	1.881293
Prob(F-statistic)	0.000000		

Theo kết quả bài tập của nhóm 13 lớp KK1_05 trường Đại học Hồng Bàng

Các yếu tố ảnh hưởng đến giá bán 1 căn nhà

Dependent Variable: Y				
Method: Least Squares				
Date: 04/01/09 Time: 10:32				
Sample: 1 225				
Included observations: 225				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.485774	0.648155	3.835155	0.0002
X2	0.003751	0.004468	0.839525	0.4021
D1	0.402730	0.294895	1.365721	0.1734
D2	0.480869	0.241677	1.989719	0.0479
D3	-1.285296	0.563680	-2.280187	0.0236
D4	1.122022	0.532487	2.107134	0.0362
D5	2.560528	0.366850	6.979768	0.0000
R-squared	0.689341	Mean dependent var	3.515511	
Adjusted R-squared	0.660240	S.D. dependent var	2.844026	
S.E. of regression	1.657752	Akaike info criterion	3.879419	
Sum squared resid	599.0949	Schwarz criterion	3.985698	
Log likelihood	-429.4346	Hannan-Quinn criter.	3.922313	
F-statistic	73.54820	Durbin-Watson stat	1.764107	
Prob(F-statistic)	0.000000			

X2 : diện tích
D1 : môi trường
D2 : khu vực
kinh doanh
D3 : nhu cầu bán
D4 : an ninh khu
vực
D5 : vị trí nhà
D6 : thị trường
đóng băng

Theo kết quả bài tập của nhóm 4 lớp KK2.05 trường Đại học Hồng Bàng

III. HỒI QUY TUYẾN TÍNH K BIẾN

4. Vấn đề dự báo

$$\text{Cho } X_o = \begin{pmatrix} 1 \\ X_2^0 \\ \dots \\ X_k^0 \end{pmatrix}$$

Yêu cầu dự báo giá trị Y_0 của Y

III. HỒI QUY TUYẾN TÍNH K BIẾN

4. Vấn đề dự báo

Dự báo điểm :

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_2 X_2^0 + \hat{\beta}_3 X_3^0 + \dots + \hat{\beta}_k X_k^0$$

Dự báo khoảng :

$$(\hat{Y}_0 - t_{\frac{\alpha}{2}} se(\hat{Y}_0); \hat{Y}_0 + t_{\frac{\alpha}{2}} se(\hat{Y}_0))$$

Bậc tự do là (n-k)

III. HỒI QUY TUYẾN TÍNH K BIẾN

4. Vấn đề dự báo

$$\sigma_{\hat{Y}_0}^2 = \hat{\sigma}^2 X_0^T (X^T X)^{-1} X_0$$

$$se(\hat{Y}_0) = \sqrt{\sigma_{\hat{Y}_0}^2}$$

III. HỒI QUY TUYẾN TÍNH K BIẾN

5. Ví dụ (số liệu trước)

Ví dụ : Tính khoảng tin cậy của β_2 theo số liệu của ví dụ trước với độ tin cậy 95%

III. HỒI QUY TUYẾN TÍNH K BIẾN

5. Ví dụ (số liệu trước)

yêu cầu kiểm định các giả thiết

$$H_0: \beta_2 = 0$$

$$H_1: \beta_2 \neq 0$$

Với độ tin cậy 95%

III. HỎI QUY TUYẾN TÍNH K BIẾN

5. Ví dụ (số liệu trước)

Yêu cầu kiểm định các giả thiết

$$H_0: R^2 = 0$$

$$H_1: R^2 \neq 0$$

Với độ tin cậy 95%

III. HỎI QUY TUYẾN TÍNH K BIẾN

5. Ví dụ (số liệu trước)

Yêu cầu dự báo giá trị của Y khi $X_2=9$
và $X_3=9$ với độ tin cậy 95%

Hết