

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

Y _i	X _{2i}	X _{3i}	X _{2i} ²	X _{3i} ²	Y _i ²	X _{2i} X _{3i}	X _{2i} Y _i	X _{3i} Y _i
1270	100	180	10000	32400	1612900	18000	127000	228600
1490	106	248	11236	61504	2220100	26288	157940	369520
1060	60	190	3600	36100	1123600	11400	63600	201400
1626	160	240	25600	57600	2643876	38400	260160	390240
1020	70	150	4900	22500	1040400	10500	71400	153000
1800	170	260	28900	67600	3240000	44200	306000	468000
1610	140	250	19600	62500	2592100	35000	225400	402500
1280	120	160	14400	25600	1638400	19200	153600	204800
1390	116	170	13456	28900	1932100	19720	161240	236300
1440	120	230	14400	52900	2073600	27600	172800	331200
1590	140	220	19600	48400	2528100	30800	222600	349800
1380	150	150	22500	22500	1904400	22500	207000	207000
16956	1452	2448	188192	518504	24549576	303608	2128740	3542360
1413	121	204						

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

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

$$\begin{aligned} \sum y_i^2 &= \sum Y_i^2 - n(\bar{Y})^2 = 590748 \\ \sum x_{2i}^2 &= \sum X_{2i}^2 - n(\bar{X}_2)^2 = 12500 \\ \sum x_{3i}^2 &= \sum X_{3i}^2 - n(\bar{X}_3)^2 = 19112 \\ \sum y_ix_{2i} &= \sum Y_iX_{2i} - n\bar{Y}\bar{X}_2 = 77064 \\ \sum y_ix_{3i} &= \sum Y_iX_{3i} - n\bar{Y}\bar{X}_3 = 83336 \\ \sum x_{2i}x_{3i} &= \sum X_{2i}X_{3i} - n\bar{X}_2\bar{X}_3 = 7400 \end{aligned}$$

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

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	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?

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 \bar{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

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

4. Hệ số xác định của mô hình

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

Equation: EQ01 Workfile: VDUABAIEN::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	2.568152	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:

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

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

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

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

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

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

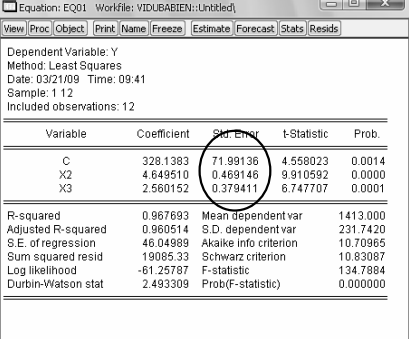
6. Khoảng tin cậy của các hệ số hồi quy

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%

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

6. Khoảng tin cậy của các hệ số hồi quy

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	Adjusted R-squared	S.E. of regression	Sum squared resid	Log likelihood	Durbin-Watson stat
	0.967693	0.960514	46.04989	19085.33	-61.25787	2.493309

	Mean dependent var	S.D. dependent var	Akaike info criterion	Schwarz criterion	F-statistic	Prob(F-statistic)
	1413.000	231.7420	10.70965	10.83087	134.7894	0.000000

I. MÔ HÌNH HỒI QUY TUYẾN TÍNH 3 BIẾN

7. Kiểm định giả thiết

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

$$\begin{aligned} H_0: \beta_i &= \beta_0 \\ H_1: \beta_i &\neq \beta_0 \end{aligned} \quad \text{Độ tin cậy là } 1-\alpha$$

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

$$\begin{aligned} H_0: \beta_2 &= 0 & H_0: \beta_3 &= 0 \\ H_1: \beta_2 &\neq 0 & H_1: \beta_3 &\neq 0 \end{aligned}$$

Với độ tin cậy 95%

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

Equation: EQ01 Workfile: VIDUBABIEN::Unltded				
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

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 :

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

Equation: EQ01 Workfile: VIDUBABIEN::Unltded				
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	

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 $Y_i^* = \ln Y_i$

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

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

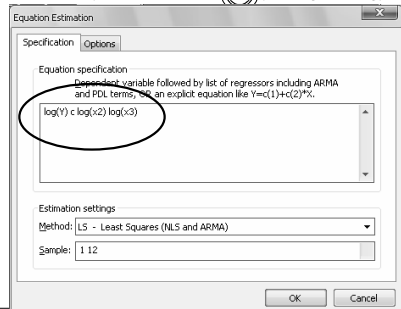
$$X_{3i}^* = \ln X_{3i}$$

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 :

$$\ln Y_i = \beta_1 + \beta_2 \ln X_{1i} + \beta_3 \ln X_{3i} + U_i$$



Kết quả hồi quy

Equation: UNTITLED Workfile: VDUABABEN:Untitled				
View Proc Object Print Name Freeze Estimate Forecast Stats Resids				
Dependent Variable: LOG(Y)				
Method: Least Squares				
Date: 04/01/09 Time: 09:57				
Sample: 1 12				
Included observations: 12				
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	2.876392	
Prob(F-statistic)	0.000000			

II. MỘT SỐ DẠNG HÀM

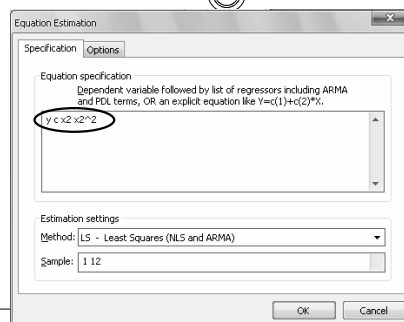
2. Hàm hồi quy đa thức bậc 2

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

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

Equation: UNTITLED Workfile: VDUABABEN:Untitled				
View Proc Object Print Name Freeze Estimate Forecast Stats Resids				
Dependent Variable: Y				
Method: Least Squares				
Date: 04/01/09 Time: 09:59				
Sample: 1 12				
Included observations: 12				
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