

Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2013

Các Phương Pháp Phân Tích Định Lượng

Gợi ý giải bài tập 10

LỰA CHỌN DẠNG HÀM & KIỂM ĐỊNH ĐẶC TRƯNG MÔ HÌNH

Ngày Phát: Thứ Ba, 17/12/2013

Ngày Nộp: Thứ Năm, 26/12/2013

Bản in nộp lúc 8h20 sáng, tại Hộp nộp bài tập trong phòng Lab

Bản điện tử gửi theo hướng dẫn của trợ giảng

Bài 1

Dữ liệu ở file **data 4-8.wf1** của Ramanathan về các hệ thống cáp của một mẫu gồm 40 thị trường truyền hình, có các biến sau:

Sub: số người đăng ký thuê bao trên mỗi hệ thống (ngàn người)

Home: số lượng nhà mà hệ thống đi qua

Inst: chi phí lắp đặt (USD)

Svc: chi phí dịch vụ hàng tháng của mỗi hệ thống (USD)

Tv: số lượng tín hiệu truyền tải bởi mỗi hệ thống cáp

Age: tuổi của mỗi hệ thống (năm)

Air: số lượng tín hiệu truyền hình miễn phí nhận được

Y: thu nhập bình quân đầu người của mỗi thị trường cáp truyền hình (USD)

a. Bạn hãy ước lượng mô hình không giới hạn sau và giải thích kết quả

$$\text{Sub} = \beta_1 + \beta_2 \text{age} + \beta_3 \text{age}^2 + \beta_4 \text{air} + \beta_5 \text{air}^2 + \beta_6 \text{y} + \beta_7 \text{y}^2 + \beta_8 \text{home} + \beta_9 \text{svc} + \beta_{10} \text{tv} + \beta_{11} \text{inst} + u \quad (1)$$

Gõ lệnh sau đây vào cửa sổ lệnh của Eviews:

ls sub c age age^2 air air^2 y y^2 home svc tv inst

Có được kết quả như hình dưới:

Dependent Variable: SUB
Method: Least Squares
Date: 12/19/13 Time: 09:07
Sample: 1 40
Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-474,0426	194,8894	-2,432367	0,0214
AGE	-1,139888	1,221231	-0,933393	0,3583
AGE^2	0,117436	0,057656	2,036834	0,0509
AIR	16,60937	4,470360	3,715444	0,0009
AIR^2	-1,428268	0,298763	-4,780611	0,0000
Y	0,090681	0,043775	2,071518	0,0473
Y^2	-4,84E-06	2,36E-06	-2,046183	0,0499

HOME	0,492678	0,030764	16,01483	0,0000
SVC	0,353465	1,948934	0,181363	0,8573
TV	0,576447	0,506629	1,137809	0,2645
INST	-0,211522	0,392045	-0,539536	0,5936
R-squared	0,946099	Mean dependent var	24,50850	
Adjusted R-squared	0,927513	S,D, dependent var	33,53720	
S.E. of regression	9,029360	Akaike info criterion	7,467256	
Sum squared resid	2364,351	Schwarz criterion	7,931698	
Log likelihood	-138,3451	Hannan-Quinn criter,	7,635184	
F-statistic	50,90272	Durbin-Watson stat	2,035560	
Prob(F-statistic)	0,000000			

Từ kết quả trên, bạn hãy viết ra phương trình hồi quy mẫu. Sau đó có một số nhận xét.

$$\text{Sub}_i = -474,0426 - 1,139888\text{AGE}_i + 0,117436\text{AGE}_i^2 + 16,60937\text{AIR} - 1,428268\text{AIR}_i^2 + 0,090681\text{Y}_i - 4,84\text{E} - 06\text{Y}_i^2 + 0,492678\text{HOME}_i + 0,353465\text{SVC}_i + 0,576447\text{TV}_i - 0,211522\text{INST}_i + \hat{u}_i$$

Mô hình hồi quy có biến age, biến svc, biến tv, biến inst không có ý nghĩa thống kê ở mức ý nghĩa 10%; Các biến giải thích khác đều có ý nghĩa thống kê ở mức ý nghĩa 10%.

Mức độ giải thích của mô hình rất cao với $R^2 = 0,946$, Adjusted $R^2 = 0,9275$, như vậy trong điều kiện các yếu tố khác không đổi, các biến độc lập trong mô hình giải thích được 92,75% sự biến thiên về mặt trung bình của Sub. Khi mô hình có một số biến không có ý nghĩa thống kê, và mức độ giải thích cao, sẽ hàm chứa dấu hiệu của hiện tượng đa cộng tuyến.

Bạn cũng có thể nhận xét về dấu, dạng của mối quan hệ giữa SUB và từng biến giải thích. Ví dụ, kết quả cho thấy có mối liên hệ phi tuyến (dạng bậc hai, hay dạng hình chữ U ngược) giữa biến SUB và AIR...; biến HOME ảnh hưởng thuận lên SUB.

b. Theo bạn, Mô hình (1) có bị vi phạm sai số đặc trưng: thừa biến không cần thiết hay không?

Nếu có, hãy khắc phục sai số đặc trưng này (giả sử bạn quan tâm đến độ tin cậy ít nhất là 90%) bằng cách loại lần lượt các biến không có ý nghĩa thống kê ra khỏi mô hình và chỉ ra mô hình mà bạn chọn. Nếu không khắc phục, theo bạn, mô hình sẽ có những trục trặc gì?

Do có một số biến không có ý nghĩa thống kê ở mức ý nghĩa 10%, vì vậy mô hình có dấu hiệu của sự thừa biến không cần thiết.

Lần lượt loại các biến không có ý nghĩa thống kê ra khỏi mô hình (biến nào ít có ý nghĩa nhất có thể bị loại ra trước): svc, inst, tv. Mô hình mới (Mô hình giới hạn, R) như sau:

ls sub c age age^2 air air^2 y y^2 home

Dependent Variable: SUB
Method: Least Squares
Date: 12/19/13 Time: 09:12
Sample: 1 40
Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-562,6761	158,0817	-3,559402	0,0012
AGE	-1,557534	0,903735	-1,723441	0,0945
AGE^2	0,139212	0,043758	3,181423	0,0033
AIR	17,30471	4,341038	3,986308	0,0004
AIR^2	-1,417655	0,291945	-4,855898	0,0000
Y	0,110830	0,034791	3,185600	0,0032
Y^2	-5,95E-06	1,88E-06	-3,164360	0,0034

HOME	0,495955	0,028300	17,52477	0,0000
R-squared	0,942507	Mean dependent var	24,50850	
Adjusted R-squared	0,929930	S,D, dependent var	33,53720	
S.E. of regression	8,877524	Akaike info criterion	7,381779	
Sum squared resid	2521,934	Schwarz criterion	7,719555	
Log likelihood	-139,6356	Hannan-Quinn criter,	7,503908	
F-statistic	74,94125	Durbin-Watson stat	2,068971	
Prob(F-statistic)	0,000000			

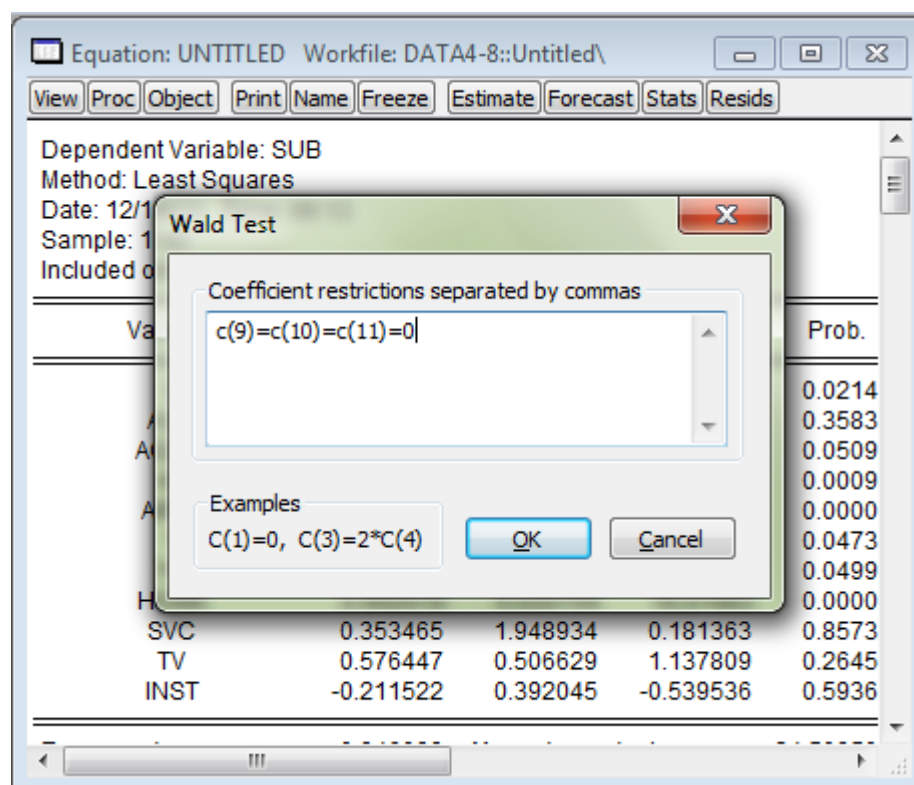
Ta có phương trình mô hình như sau:

$$\text{Sub}_i = -562,6761 - 1,557534\text{AGE}_i + 0,139212\text{AGE}_i^2 + 17,30471\text{AIR} - 1,417655\text{AIR}_i^2 + 0,110830\text{Y}_i - 5,95\text{E} - 06\text{Y}_i^2 + 0,495955\text{HOME}_i + \hat{u}_i$$

Khi mô hình bị thừa biến không cần thiết, nếu không khắc phục, phương sai ước lượng của các hệ số sẽ có giá trị lớn hơn, do vậy các ước lượng sẽ không hiệu quả; những suy diễn xác suất của chúng ta về các thông số sẽ kém chính xác (Xem thêm Gujarati, chương 13)

c. Về mặt thống kê, từ mô hình (1), bằng kiểm định Wald bạn hãy đưa ra kết luận xem có nên loại đồng thời cả ba biến inst, tv, svc ra khỏi mô hình (1) hay không?

Tại cửa sổ Equation của mô hình 1, chọn View\Coefficient Tests\Wald-Coefficient Restrictions



Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	0,644280	(3, 29)	0.5928
Chi-square	1,932839	3	0.5865

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std, Err,
C(9)	0,353465	1,948934
C(10)	0,576447	0,506629
C(11)	-0,211522	0,392045

Restrictions are linear in coefficients.

Do $P\text{-value}(F)=0,5928 (>0,05)$ nên xét ở độ tin cậy 95%, chưa đủ cơ sở để bác bỏ giả thuyết cho rằng cả ba biến svc, tv, inst đều không có ý nghĩa thống kê. Hay nói cách khác, có thể loại đồng thời cả ba biến này ra khỏi mô hình.

d. Về mặt thống kê, bằng kiểm định nhân tử Lagrange, bạn hãy đưa ra kết luận xem mô hình mà bạn chọn ở câu b có bị thiếu ba biến quan trọng sau đây hay không: inst, tv, svc.

Sau khi ước lượng mô hình R (mô hình 2) ở câu b bằng dòng lệnh:

LS sub c age age^2 air air^2 y y^2 home

Ta lưu lại phần dư của mô hình R:

Genr uhat=resid

Ước lượng mô hình hồi quy phụ:

LS uhat c age age^2 air air^2 y y^2 home svc tv inst

Dependent Variable: UHAT
Method: Least Squares
Date: 12/19/13 Time: 09:16
Sample: 1 40
Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	88,63348	194,8894	0,454789	0,6526
AGE	0,417646	1,221231	0,341988	0,7348
AGE^2	-0,021776	0,057656	-0,377691	0,7084
AIR	-0,695342	4,470360	-0,155545	0,8775
AIR^2	-0,010612	0,298763	-0,035520	0,9719
Y	-0,020150	0,043775	-0,460297	0,6487
Y^2	1,11E-06	2,36E-06	0,471069	0,6411
HOME	-0,003277	0,030764	-0,106530	0,9159
SVC	0,353465	1,948934	0,181363	0,8573
TV	0,576447	0,506629	1,137809	0,2645
INST	-0,211522	0,392045	-0,539536	0,5936
R-squared	0,062485	Mean dependent var		7,81E-14
Adjusted R-squared	-0,260796	S,D, dependent var		8,041454
S.E. of regression	9,029360	Akaike info criterion		7,467256
Sum squared resid	2364,351	Schwarz criterion		7,931698
Log likelihood	-138,3451	Hannan-Quinn criter,		7,635184
F-statistic	0,193284	Durbin-Watson stat		2,035560
Prob(F-statistic)	0,995335			

Gán giá trị cho trị kiểm định:

$$\text{scalar } nR^2_{aux} = 0,062485 * 40$$

Hàm hồi quy phụ có $R^2 = 0,062485$ và $n = 40 \Rightarrow nR^2 = 0,062485 * 40 = 2,4994$

$$\chi^2_{0,05,3} = \text{CHIINV}(0.05, 3) = 7,8147^1$$

Do $nR^2 < \chi^2_{0,05,3}$ nên chưa đủ cơ sở để bác bỏ giả thuyết cho rằng cả ba biến svc , tv , $inst$ đều không có ý nghĩa thống kê với mức ý nghĩa 5%. Nói cách khác, không nên đưa biến nào trong ba biến này vào mô hình.

Nhận định này càng rõ ràng hơn khi ta thấy P-value của các biến SVC , TV , $INST$ ở hàm hồi quy phụ không có ý nghĩa thống kê.

[Ghi chú: khi so sánh các hệ số hồi quy, Sai số chuẩn, P-value của các biến svc , tv , $inst$ ở mô hình 1 với mô hình hồi quy phụ bạn sẽ thấy chúng tương tự như ở mô hình hồi quy phụ]

e. Hãy xem mô hình mà bạn lựa chọn ở câu b có bị sai số đặc trưng không bằng thống kê Durbin-Watson.

Cách 1:

Hệ số DW được sử dụng để kiểm định tự tương quan bậc 1 (trường hợp dữ liệu chuỗi thời gian)... Bên cạnh đó, theo Gujarati (2003), hệ số này còn được sử dụng để kiểm định xem mô hình có bị sai số đặc trưng hay không (sai dạng hàm, hay thiếu biến quan trọng...) dù là dữ liệu chéo hay dữ liệu chuỗi thời gian. Trong mô hình ở câu b, $DW = 2,068$ rất gần 2 nên mô hình không có dấu hiệu bị sai số đặc trưng (nếu không có bảng tra DW, ta có thể sử dụng quy tắc kinh nghiệm $1,5 < DW < 2,5$ cho nhanh)

Cách 2:

Ta có thể tra bảng DW ở mức ý nghĩa 5%.

Với $k' = 7$, $n = 40 \Rightarrow d_L = 1,120$, $d_U = 1,924$ và các cận $4 - d_U$, $4 - d_L$ như sau:

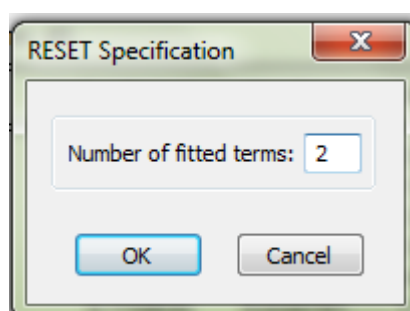
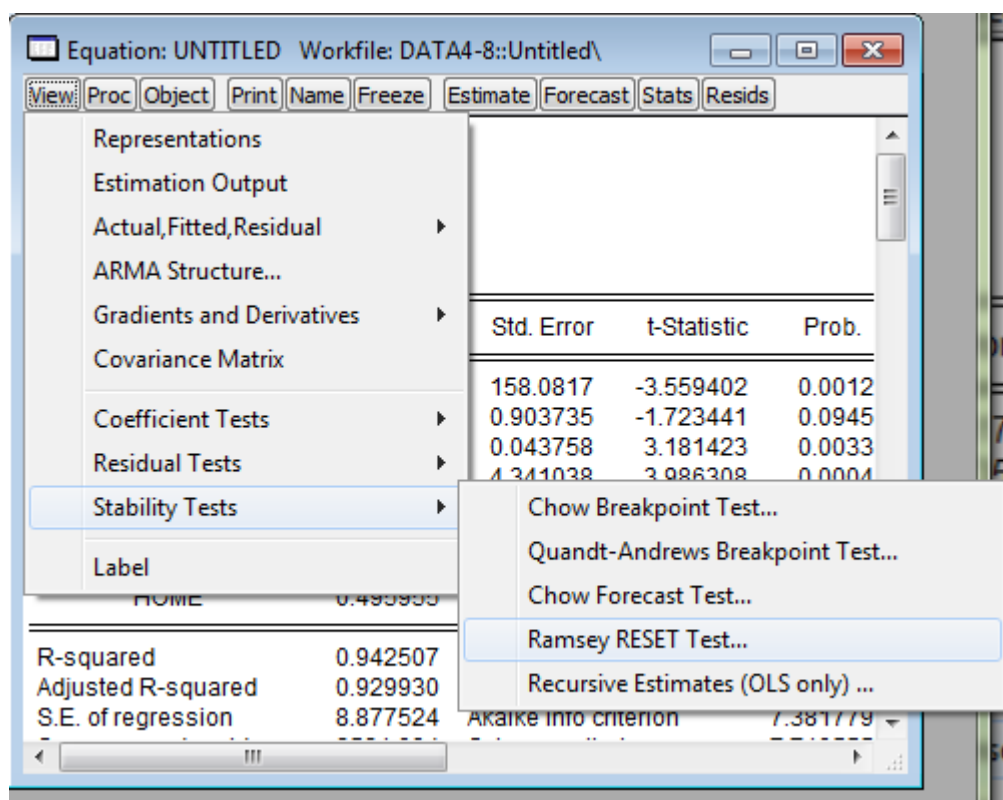
d_L	d_U	2	$4 - d_U$	$4 - d_L$
1.12	1.924	2	2.076	2.88

Do DW nằm trong khoảng $(d_U, 4 - d_U)$ nên mô hình R không có dấu hiệu bị sai số đặc trưng ở mức ý nghĩa 5%.

f. Hãy sử dụng kiểm định RESET của Ramsey để xem xét mô hình mà bạn đã chọn ở câu b có bị sai số đặc trưng hay không, và cho biết nhận xét của bạn từ kết quả tính toán.

Ngoài cách tính thủ công, từ mô hình ở câu b, ta thực hiện kiểm định RESET của Ramsey như sau: View\Stability Tests\Ramsey RESET Test, nhập số 2 để đưa 2 biến $\hat{Y}^2$ và $\hat{Y}^3$ vào hàm hồi quy phụ của kiểm định Ramsey RESET:

¹ Học viên cũng có thể so sánh với mức ý nghĩa 10%



Ramsey RESET Test:

F-statistic	12,55575	Prob, F(2,30)	0,0001
Log likelihood ratio	24,32644	Prob, Chi-Square(2)	0,0000

Test Equation:

Dependent Variable: SUB

Method: Least Squares

Date: 12/19/13 Time: 09:26

Sample: 1 40

Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob,
C	-72,55425	163,5784	-0,443544	0,6606
AGE	0,042819	0,763460	0,056085	0,9556
AGE^2	0,033500	0,040563	0,825884	0,4154
AIR	5,973589	4,306765	1,387025	0,1757
AIR^2	-0,517724	0,307263	-1,684954	0,1024
Y	0,013654	0,034290	0,398191	0,6933
Y^2	-7,76E-07	1,84E-06	-0,421254	0,6766
HOME	0,117054	0,106433	1,099792	0,2802

FITTED^2	0,012407	0,005020	2,471520	0,0194
FITTED^3	-4,69E-05	2,56E-05	-1,829000	0,0774
R-squared	0,968704	Mean dependent var	24,50850	
Adjusted R-squared	0,959315	S.D, dependent var	33,53720	
S.E. of regression	6,764656	Akaike info criterion	6,873618	
Sum squared resid	1372,817	Schwarz criterion	7,295838	
Log likelihood	-127,4724	Hannan-Quinn criter,	7,026279	
F-statistic	103,1751	Durbin-Watson stat	1,634277	
Prob(F-statistic)	0,000000			

P-value(F) trong kiểm định Ramsey RESET là 0,0001 (<0,05) nên mô hình có dấu hiệu bị sai số đặc trưng.

Nhận xét: Tuy kiểm định Reset của Ramsey cho biết mô hình có dấu hiệu của sai số đặc trưng nhưng kiểm định này không cho ta biết mô hình bị sai số đặc trưng nào: có thể là thừa biến không cần thiết, thiếu biến quan trọng, sai dạng hàm, hay có sai số trong đo lường. Để sửa chữa mô hình, ta cần phải thực hiện thêm các kiểm định khác nữa.

g. Với mô hình mà bạn đã chọn ở câu b, hãy cho biết tác động biên của Y theo các biến có ở trong mô hình (nếu tác động biên phụ thuộc vào một biến giải thích khác, hãy tính tác động biên này ở giá trị trung bình của các biến giải thích)?

Thống kê mô tả cho các biến giải thích như sau:

	AGE	AIR	Y	HOME	INST	SUB	SVC	TV
Mean	8,566750	6,375000	9209,675	58,18805	15,47750	24,50850	7,339000	10,60000
Median	7,500000	6,000000	9040,000	32,68950	15,00000	13,70000	7,000000	9,500000
Maximum	26,00000	13,00000	11741,00	350,0000	25,00000	170,0000	10,00000	22,00000
Minimum	0,170000	4,000000	7683,000	1,700000	5,950000	1,000000	5,600000	6,000000
Std. Dev.	5,423340	2,238217	879,9279	76,69182	4,791953	33,53720	1,180799	4,005125
Skewness	0,747991	1,243875	0,549449	2,384819	0,353951	2,698333	0,785988	1,093445
Kurtosis	3,927810	4,147316	3,183012	8,346872	2,627315	10,82550	3,070614	3,585187
Jarque-Bera	5,164656	12,50872	2,068449	85,56415	1,066702	150,6040	4,126830	8,541559
Probability	0,075598	0,001922	0,355502	0,000000	0,586636	0,000000	0,127019	0,013971
Sum	342,6700	255,0000	368387,0	2327,522	619,1000	980,3400	293,5600	424,0000
Sum Sq. Dev.	1147,092	195,3750	30196653	229383,8	895,5498	43865,00	54,37716	625,6000
Observations	40	40	40	40	40	40	40	40

$$Sub_i = -562,6761 - 1,557534AGE_i + 0,139212AGE_i^2 + 17,30471AIR - 1,417655AIR_i^2 + 0,110830Y_i - 5,95E - 06Y_i^2 + 0,495955HOME_i + \hat{u}_i$$

Tác động biên của SUB theo home là 0,49595 ; Trong điều kiện các yếu tố khác không đổi, Nếu home tăng 1 đơn vị thì trung bình SUB tăng 0,49595 đơn vị, và ngược lại.

Tác động biên của SUB theo AGE là $-1,55753 + 2 \cdot 0,139211780597 \cdot AGE = -1,55753 + 2 \cdot 0,139211780597 \cdot 8,57 = 0,82765$

Tương tự, ta tính được tác động biên của SUB theo AIR là $17,30471408 - 2 \cdot 1,417655395 \cdot 6,375000 = -0,77039125$.

Tác động biên của SUB theo Y là $0,110830437 + 2 \cdot -5,95E - 06 \cdot 9,210 = 0,00126297$

h. Anh Tèo đã ước lượng mô hình sau:

$$\ln(\text{sub}) = \alpha_1 + \alpha_2 \text{age} + \alpha_3 \text{age}^2 + \alpha_4 \text{home} + v \quad (2)$$

Trong mô hình trên, nếu biến home có ý nghĩa thống kê ở mức ý nghĩa 10%, và ít nhất một trong 2 biến age, age^2 có ý nghĩa thống kê ở mức ý nghĩa 10% thì Tèo có thể sử dụng. Theo bạn, từ thông tin của mẫu, hệ số ước lượng của α_4 cho Tèo biết điều gì?

ls log(sub) c age age^2 home

Dependent Variable: LOG(SUB)

Method: Least Squares

Date: 12/19/13 Time: 09:29

Sample: 1 40

Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1,125250	0,336367	3,345309	0,0019
AGE	0,125856	0,067117	1,875169	0,0689
AGE^2	-0,003802	0,002925	-1,299761	0,2019
HOME	0,011432	0,001860	6,146409	0,0000
R-squared	0,625836	Mean dependent var		2,480569
Adjusted R-squared	0,594655	S.D. dependent var		1,282355
S.E. of regression	0,816433	Akaike info criterion		2,526895
Sum squared resid	23,99626	Schwarz criterion		2,695783
Log likelihood	-46,53791	Hannan-Quinn criter,		2,587960
F-statistic	20,07146	Durbin-Watson stat		1,353227
Prob(F-statistic)	0,000000			

Hệ số hồi quy đứng trước biến home=0,01143 cho biết: trong điều kiện các yếu tố khác không đổi, nếu home tăng một nhà thì về mặt trung bình, sub sẽ tăng 1,1432 %

i. Với mô hình 2, từ thông tin của mẫu, xét tại giá trị trung bình của home, age, theo bạn tác động biên, và hệ số co giãn của sub theo age sẽ là bao nhiêu?

$$\ln(\text{sub}_i) = \hat{\alpha}_1 + \hat{\alpha}_2 \text{age}_i + \hat{\alpha}_3 \text{age}_i^2 + \hat{\alpha}_4 \text{home}_i + \hat{v}_i$$

$$\ln(\text{sub}_i) = 1,125250 + 0,125856 \text{age}_i - 0,003802 \text{age}_i^2 + 0,011432 \text{home}_i + \hat{v}_i$$

Từ các công thức trên, tại giá trị trung bình của home=58,18805, age=8,56675 ta tính được $\ln(\text{sub})=2,589606916 \Rightarrow \text{sub}=13,32453291$. Thế các giá trị tương ứng vào công thức trên, sẽ tính được các giá trị cụ thể:

Theo Ramanathan (2002, chương 6) với dạng hàm như ở mô hình 2.

Tác động biên của Sub theo home là $\hat{\alpha}_4 * \text{sub} = 0,15232606$

Tác động biên của Sub theo age là $(\hat{\alpha}_2 + 2\hat{\alpha}_3 \text{age}) * \text{sub} = 0,808991461$

Hệ số co giãn của Sub theo home là $\hat{\alpha}_4 * \text{home} = 0,665205788$

Hệ số co giãn của Sub theo age là $(\hat{\alpha}_2 + 2\hat{\alpha}_3 \text{age}) * \text{age} = 0,520125369$

j. Với Stata, bạn hãy cho biết những câu lệnh cần thiết trong do-file để tính toán được các kết quả cần thiết nhằm trả lời cho câu a, b, c, h. Các kết quả ấy là gì?

Làm lại câu a bằng Stata


```
. gen agesq=age^2
. gen airsq=air^2
. gen ysq=y^2
. reg sub age agesq air airsq y ysq home svc tv inst
```

Source	SS	df	MS	Number of obs = 40		
Model	41500.6508	10	4150.06508	F(10, 29) = 50.90		
Residual	2364.35072	29	81.5293351	Prob > F = 0.0000		
				R-squared = 0.9461		
				Adj R-squared = 0.9275		
Total	43865.0015	39	1124.74363	Root MSE = 9.0294		

sub	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	-1.13989	1.221231	-0.93	0.358	-3.637588	1.357808
agesq	.1174357	.057656	2.04	0.051	-.000484	.2353554
air	16.60937	4.47036	3.72	0.001	7.466456	25.75228
airsq	-1.428267	.2987626	-4.78	0.000	-2.039305	-.8172293
y	.0906809	.0437751	2.07	0.047	.0011508	.1802109
ysq	-4.84e-06	2.36e-06	-2.05	0.050	-9.67e-06	-2.26e-09
home	.4926775	.0307638	16.01	0.000	.4297584	.5555966
svc	.3534657	1.948934	0.18	0.857	-3.632552	4.339483
tv	.5764471	.506629	1.14	0.265	-.4597255	1.61262
inst	-.211522	.3920449	-0.54	0.594	-1.013344	.5903
_cons	-474.0426	194.8893	-2.43	0.021	-872.6359	-75.4493

Làm lại câu b bằng Stata

```
. test svc tv inst
```

```
( 1)  svc = 0
( 2)  tv = 0
( 3)  inst = 0
```

```
F( 3, 29) = 0.64
Prob > F = 0.5928
```

```
reg sub age agesq air airsq y ysq home
```

Source	SS	df	MS	Number of obs = 40		
Model	41343.0677	7	5906.15253	F(7, 32) = 74.94		
Residual	2521.9338	32	78.8104313	Prob > F = 0.0000		
				R-squared = 0.9425		
				Adj R-squared = 0.9299		
Total	43865.0015	39	1124.74363	Root MSE = 8.8775		

sub	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	-1.557536	.9037355	-1.72	0.094	-3.398385	.2833129
agesq	.1392119	.0437577	3.18	0.003	.0500803	.2283435
air	17.30471	4.341038	3.99	0.000	8.462305	26.14712
airsq	-1.417655	.2919451	-4.86	0.000	-2.012328	-.8229826
y	.1108304	.034791	3.19	0.003	.0399634	.1816974
ysq	-5.95e-06	1.88e-06	-3.16	0.003	-9.78e-06	-2.12e-06
home	.4959548	.0283002	17.52	0.000	.4383091	.5536005
_cons	-562.6759	158.0816	-3.56	0.001	-884.6775	-240.6743

Làm lại câu c bằng Stata

```
. reg sub age agesq air airsq y ysq home svc tv inst
```

. test svc=tv= inst=0

```
( 1)  svc - tv = 0
( 2)  svc - inst = 0
( 3)  svc = 0
```

```
F( 3, 29) = 0.64
Prob > F = 0.5928
```

Làm lại câu h bằng Stata

. gen lnsub=ln(sub)

. reg lnsub age agesq home

Source	SS	df	MS	Number of obs =	40
Model	40.1366617	3	13.3788872	F(3, 36) =	20.07
Residual	23.996258	36	.666562722	Prob > F =	0.0000
				R-squared =	0.6258
				Adj R-squared =	0.5947
Total	64.1329197	39	1.64443384	Root MSE =	.81643

lnsub	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	.1258564	.0671174	1.88	0.069	-.0102639	.2619767
agesq	-.003802	.0029251	-1.30	0.202	-.0097344	.0021305
home	.0114318	.0018599	6.15	0.000	.0076597	.0152039
_cons	1.12525	.3363667	3.35	0.002	.4430672	1.807434

Bài 2.

Bạn hãy thu thập dữ liệu về GDP theo giá cố định của Việt Nam từ năm 1995 đến 2010.

<http://world-economic-outlook.findthedata.org/l/6087/Vietnam> (Giá so sánh 1994)

Đơn vị: tỷ VND

1995	1996	1997	1998	1999	2000	2001	2002
195,567	213,833	231,264	244,596	256,272	273,666	292,535	313,247

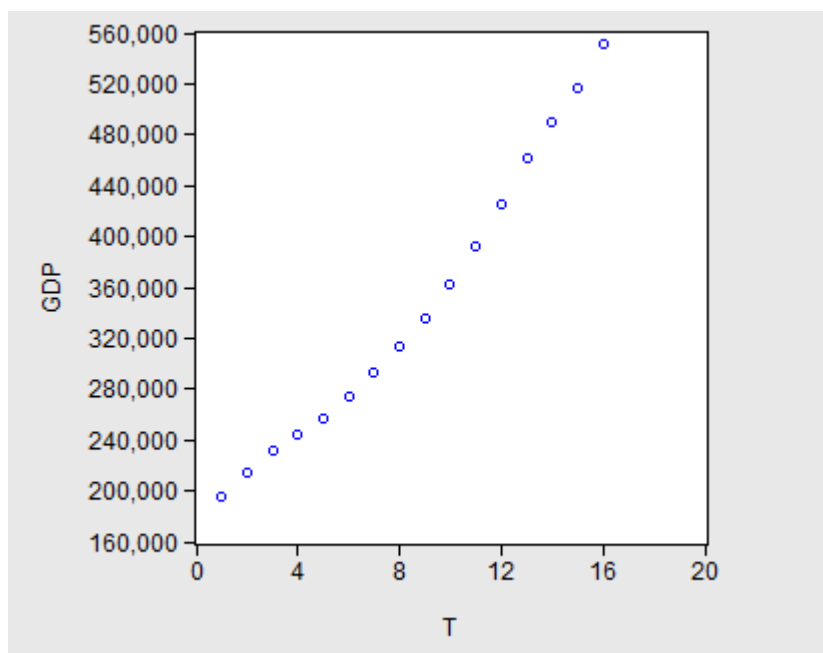
2003	2004	2005	2006	2007	2008	2009	2010
336,243	362,435	393,031	425,373	461,343	490,459	516,566	551,609

a. Tạo ra biến T là thứ tự thời gian, với T=1 cho năm 1995, T=2 cho năm 1996 ... và T=16 cho năm 2010 và vẽ biểu đồ biến động của GDP theo T. Dựa vào đồ thị, theo bạn dạng hàm GDP=f(T) nào có thể phù hợp cho chuỗi dữ liệu GDP trên?

Trong cửa sổ lệnh Eview gõ:

genr t=@trend(1994)

Về biểu đồ Scatter giữa GDP và T, ta được biểu đồ như sau:



Đồ thị phân tán của GDP theo T cho thấy có thể $GDP=f(T)$ theo dạng hàm bậc nhất, bậc hai, bậc 3, hoặc cũng có thể theo dạng Logarithmic, tăng trưởng mũ...

b. Giả sử bạn đang phân vân một trong các mô hình sau:

$$\widehat{GDP} = a_0 + a_1 T$$

$$\widehat{GDP} = b_0 + b_1 T + b_2 T^2$$

$$\widehat{GDP} = c_0 + c_1 \ln(T)$$

$$\widehat{GDP} = e^{\lambda_0 + \lambda_1 T}$$

Hãy ước lượng các mô hình trên và chọn ra 1 mô hình mà bạn cho là phù hợp nhất.

Mô hình 1:

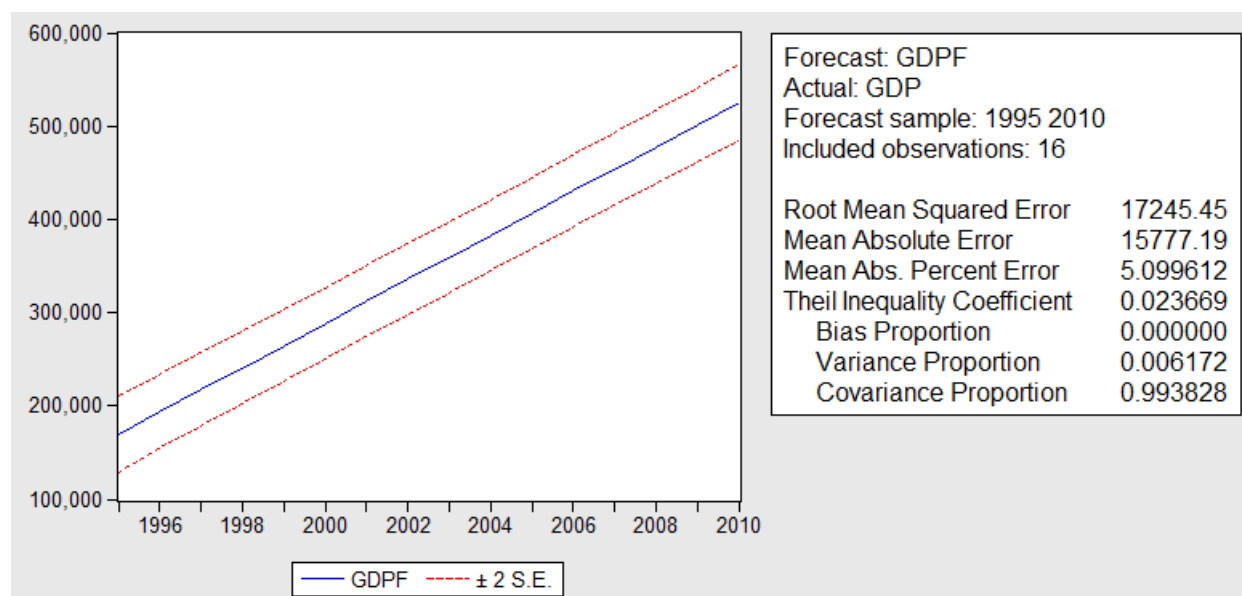
LS GDP C T

Dependent Variable: GDP
Method: Least Squares
Date: 12/20/13 Time: 10:22
Sample: 1995 2010
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	146236,7	9668,004	15,12585	0,0000
T	23663,61	999,8412	23,66737	0,0000
R-squared	0,975616	Mean dependent var		347377,4
Adjusted R-squared	0,973874	S,D, dependent var		114060,5
S.E. of regression	18436,16	Akaike info criterion		22,59848

Sum squared resid	4,76E+09	Schwarz criterion	22,69506
Log likelihood	-178,7879	Hannan-Quinn criter,	22,60343
F-statistic	560,1444	Durbin-Watson stat	0,175745
Prob(F-statistic)	0,000000		

Dùng lệnh Forecast để tính các chỉ tiêu đo lường sai số của mô hình bậc nhất



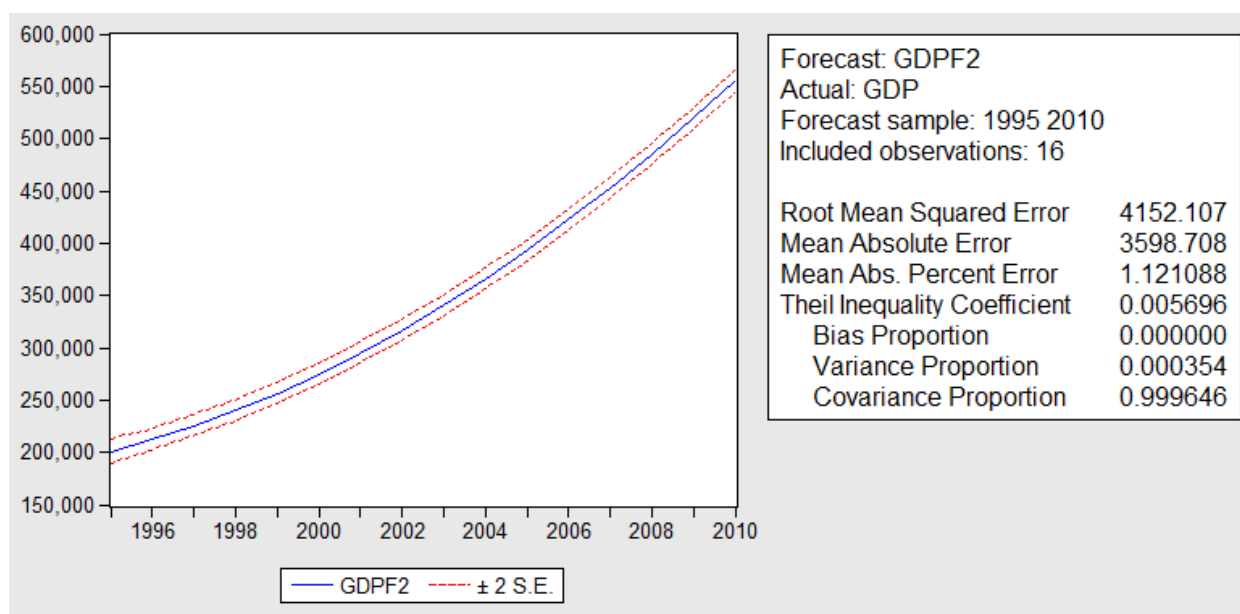
Mô hình 2:

LS GDP C T T²

Dependent Variable: GDP
Method: Least Squares
Date: 12/20/13 Time: 10:24
Sample: 1995 2010
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	191416,5	3936,629	48,62445	0,0000
T	8603,709	1065,814	8,072427	0,0000
T ²	885,8766	60,94847	14,53485	0,0000

R-squared	0,998587	Mean dependent var	347377,4
Adjusted R-squared	0,998369	S,D, dependent var	114060,5
S.E. of regression	4606,350	Akaike info criterion	19,87562
Sum squared resid	2,76E+08	Schwarz criterion	20,02048
Log likelihood	-156,0050	Hannan-Quinn criter,	19,88304
F-statistic	4592,022	Durbin-Watson stat	0,830896
Prob(F-statistic)	0,000000		

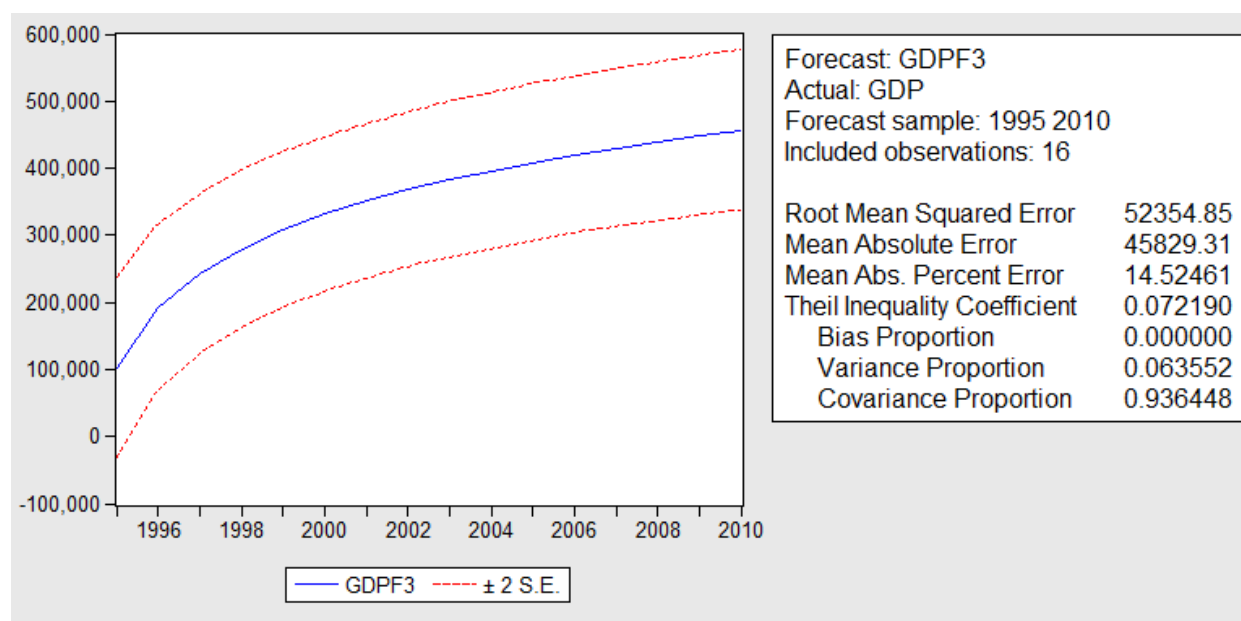


Mô hình 3:

LS GDP C LOG(T)

Dependent Variable: GDP
Method: Least Squares
Date: 12/20/13 Time: 10:26
Sample: 1995 2010
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	103419,1	37790,38	2,736652	0,0161
LOG(T)	127261,1	18312,28	6,949493	0,0000
R-squared	0,775264	Mean dependent var		347377,4
Adjusted R-squared	0,759212	S,D, dependent var		114060,5
S.E. of regression	55969,69	Akaike info criterion		24,81948
Sum squared resid	4,39E+10	Schwarz criterion		24,91605
Log likelihood	-196,5558	Hannan-Quinn criter,		24,82442
F-statistic	48,29545	Durbin-Watson stat		0,228533
Prob(F-statistic)	0,000007			



Mô hình 4:

LS LOG(GDP) C T

Dependent Variable: LOG(GDP)

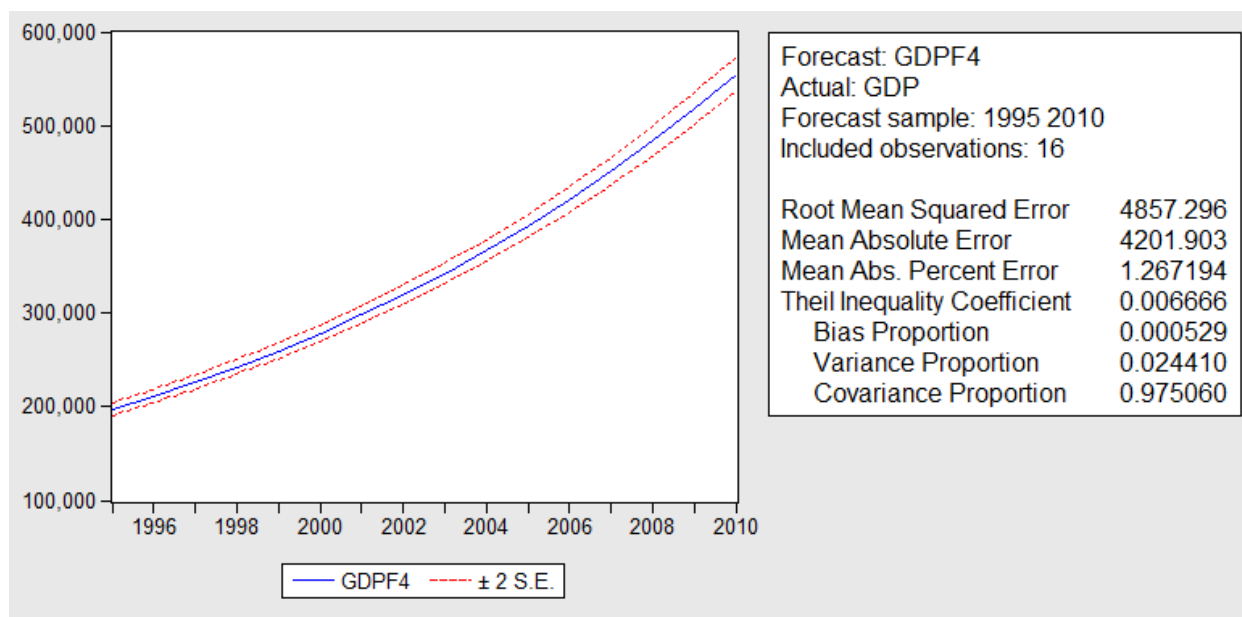
Method: Least Squares

Date: 12/20/13 Time: 10:27

Sample: 1995 2010

Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12,11906	0,008004	1514,089	0,0000
T	0,069222	0,000828	83,62466	0,0000
R-squared	0,998002	Mean dependent var	12,70745	
Adjusted R-squared	0,997859	S,D, dependent var	0,329894	
S.E. of regression	0,015263	Akaike info criterion	-5,410250	
Sum squared resid	0,003262	Schwarz criterion	-5,313676	
Log likelihood	45,28200	Hannan-Quinn criter,	-5,405305	
F-statistic	6993,084	Durbin-Watson stat	0,608465	
Prob(F-statistic)	0,000000			



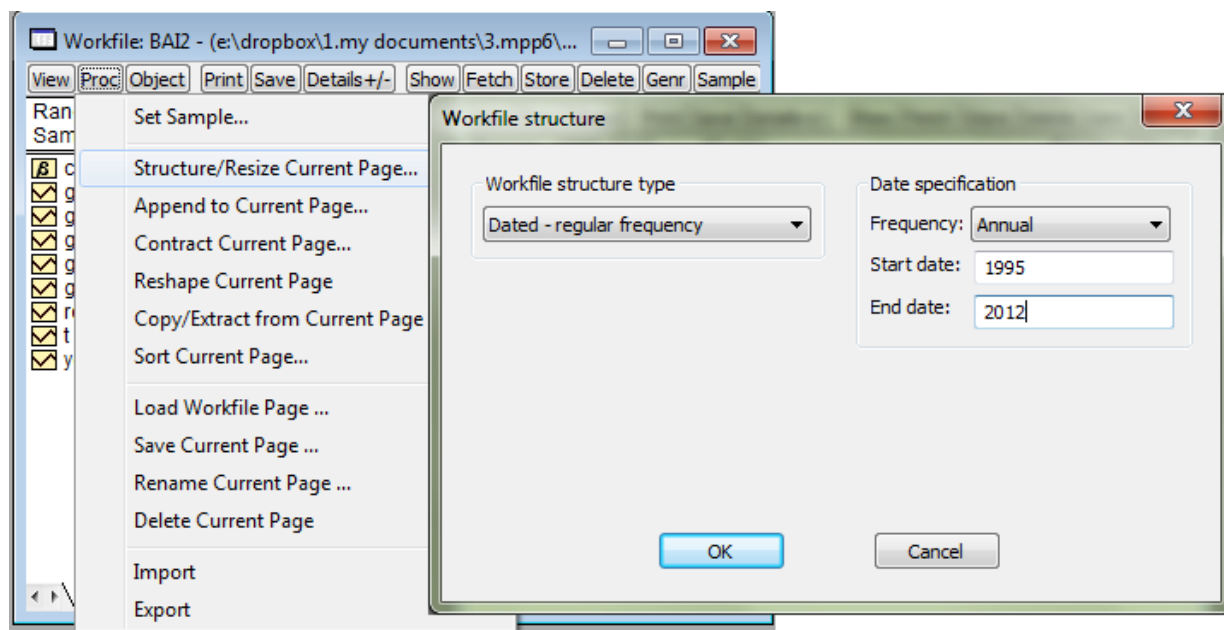
Từ Kết quả của hàm Log-Lin, ta có thể viết được phương trình dự báo dạng tăng trưởng mũ như sau: $\widehat{GDP}_i = e^{12.11906 + 0.069222T}$

Có thể dựa vào chỉ tiêu RMSE, MAPE, Theil's U của các mô hình để lựa chọn, mô hình nào có các chỉ tiêu này càng nhỏ thì càng tốt. Nếu bạn sử dụng MAPE thì thấy rằng MAPE của hàm bậc 2 là 1.12%, tăng trưởng mũ là 1.27%, hàm bậc nhất là 5.10%, Logarithmic là 14.52%. Như vậy, trong 4 mô hình, mô hình hàm bậc 2 có sai số nhỏ nhất, hay độ chính xác tốt nhất.

Forecast: GDPF	Forecast: GDPF2	Forecast: GDPF3	Forecast: GDPF4
Actual: GDP	Actual: GDP	Actual: GDP	Actual: GDP
Forecast sample: 1995 2010	Forecast sample: 1995 2010	Forecast sample: 1995 2010	Forecast sample: 1995 2010
Included observations: 16	Included observations: 16	Included observations: 16	Included observations: 16
Root Mean Squared Error 17245.45	Root Mean Squared Error 4152.107	Root Mean Squared Error 52354.85	Root Mean Squared Error 4857.296
Mean Absolute Error 15777.19	Mean Absolute Error 3598.708	Mean Absolute Error 45829.31	Mean Absolute Error 4201.903
Mean Abs. Percent Error 5.099612	Mean Abs. Percent Error 1.121088	Mean Abs. Percent Error 14.52461	Mean Abs. Percent Error 1.267194
Theil Inequality Coefficient 0.023669	Theil Inequality Coefficient 0.005696	Theil Inequality Coefficient 0.072190	Theil Inequality Coefficient 0.006666
Bias Proportion 0.000000	Bias Proportion 0.000000	Bias Proportion 0.000000	Bias Proportion 0.000529
Variance Proportion 0.006172	Variance Proportion 0.000354	Variance Proportion 0.063552	Variance Proportion 0.024410
Covariance Proportion 0.993828	Covariance Proportion 0.999646	Covariance Proportion 0.936448	Covariance Proportion 0.975060

c. Với mô hình mà bạn đã chọn ở câu b, nếu quy luật biến động của dữ liệu theo thời gian vẫn còn đúng đến năm 2012; bạn hãy cho biết giá trị dự báo điểm, khoảng tin cậy 95% của giá trị dự báo GDP (giá trị cá biệt) sẽ là bao nhiêu?

Để thêm 2 năm 2012 và 2013 vào cơ sở dữ liệu, trong Eview ta dùng lệnh:



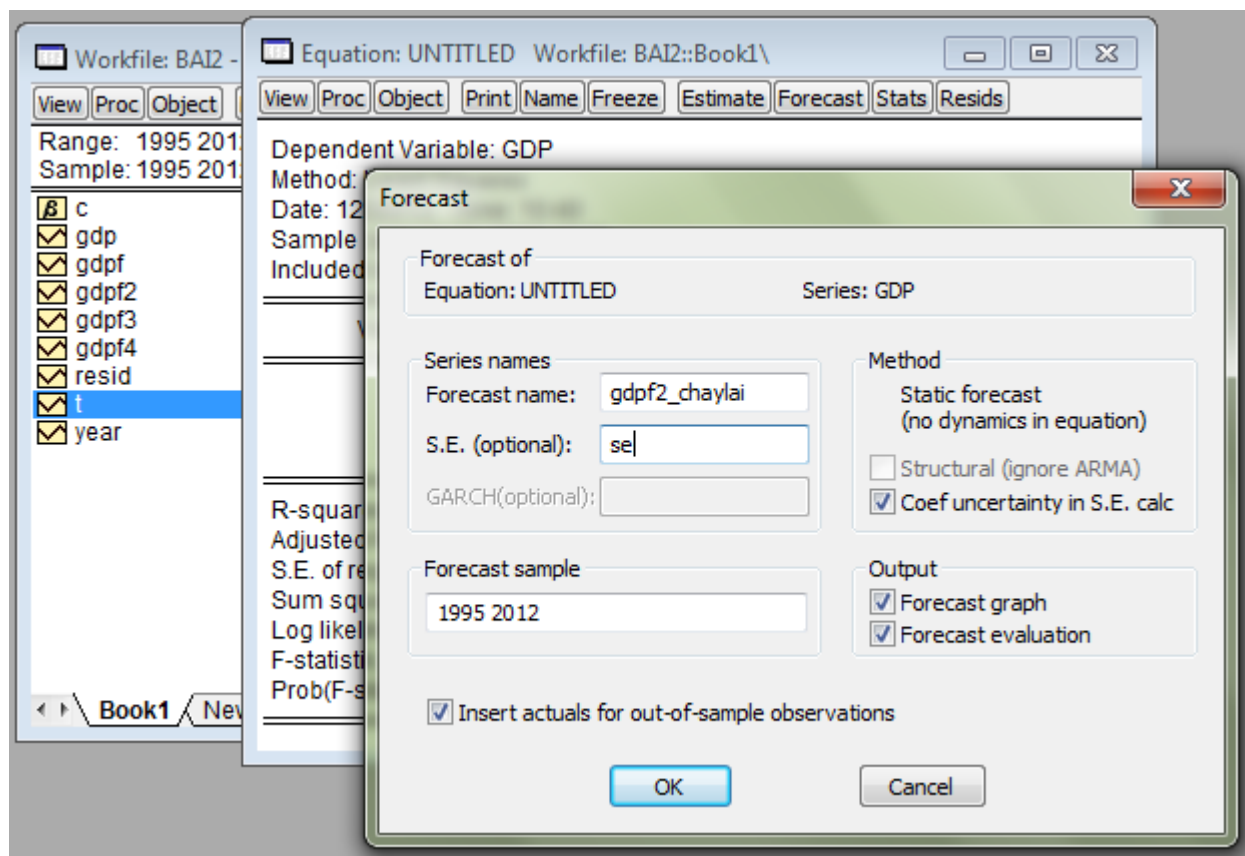
Sau đó, cập nhật lại số năm cho biến T bằng lệnh:

Genr t=@trend(1994)

Ở năm 2012, Nếu sử dụng dạng hàm bậc hai để dự báo POP, kết quả dự báo điểm của GDP sẽ là: $191416,5 + 8603,709 \cdot 18 + 885,8766 \cdot 18^2 = 633307,2804$ (tỷ VNĐ),

Chạy lại mô hình, trong Eview ta dùng lệnh: **LS Gdp C T T^2**

Sau đó tiếp tục dùng lệnh forecast như hình dưới



Ta được $SE_{2012} = 6772,19496$

Với $t_{\alpha/2; (df=n-k)} = t_{0,025 (df = 16-3)} = 2,160368652^2$

Khoảng tin cậy 95% của GDP sẽ là $[633307,2804 \pm 2,160368652 * 6772,19496] = [618676,8; 647937,6754]$ tỷ VNĐ

Có thể thực hiện trên Eview ta cũng được kết quả tương tự:

Cận dưới là: $genr\ low = gdpf2_chaylai - 2,160368652 * se$

Cận trên là: $genr\ high = gdpf2_chaylai + 2,160368652 * se$

Group: UNTITLED Workfile: BAI2::Book1\					
View	Proc	Object	Print	Name	Freeze
Default	Sort	Transpose	Edit+/-	Smpl+/-	Tit
obs	GDP	GDPF2_CH...	SE	LOW	HIGH
1995	195567	200906.0	5532.228	188954.4	212857.7
1996	213833	212167.4	5178.755	200979.4	223355.4
1997	231264	225200.5	4973.261	214456.4	235944.5
1998	244596	240005.3	4879.761	229463.2	250547.4
1999	256272	256581.9	4859.011	246084.7	267079.2
2000	273666	274930.3	4875.497	264397.4	285463.1
2001	292535	295050.4	4901.636	284461.0	305639.7
2002	313247	316942.2	4919.187	306315.0	327569.5
2003	336243	340605.8	4919.187	329978.6	351233.1
2004	362435	366041.2	4901.636	355451.9	376630.5
2005	393031	393248.3	4875.497	382715.4	403781.2
2006	425373	422227.2	4859.011	411729.9	432724.4
2007	461343	452977.8	4879.761	442435.7	463519.9
2008	490459	485500.2	4973.261	474756.1	496244.3
2009	516566	519794.3	5178.755	508606.3	530982.3
2010	551609	555860.2	5532.228	543908.5	567811.8
2011	NA	593697.8	6059.332	580607.4	606788.2
2012	NA	633307.2	6772.195	618676.8	647937.7

² Chú ý bậc tự do của t luôn không đổi