

Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2013

Các Phương Pháp Phân Tích Định Lượng

Gợi ý giải bài tập 9

HỒI QUY ĐA BIẾN (TIẾP THEO)

Ngày Phát: Thứ ba 10/12/2013

Ngày Nộp: Thứ ba 17/12/2013

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 (50 điểm) Một nghiên cứu của tác giả Tô Trung Thành về “Đầu tư công, “lấn át” đầu tư tư nhân? Góc nhìn từ mô hình thực nghiệm VECM” có kết quả của hàm hồi quy đồng liên kết như sau:

$$\ln(Y_i) = 6,57 + 0,23\ln(GI_i) + 0,33\ln(PI_i) + \hat{u}_i$$

(t-stat) (?) (30,46) (28,17)

Hàm trên sử dụng số liệu theo năm, từ 1986-2010 với nguồn số liệu của Tổng cục thống kê. Biến Y là GDP, biến GI là đầu tư khu vực nhà nước, và PI là đầu tư của khu vực tư nhân; các biến này đều lấy theo giá so sánh năm 1994.

a. Bạn hãy cho biết biến $\ln(GI)$ có ý nghĩa thống kê ở độ tin cậy 99% hay không?

Ho : $\beta_2 = 0$

Ha : $\beta_2 \neq 0$

T-stat[$\ln(GI)$]=30,46

$$t_{\frac{\alpha}{2}, (n-k)} = t_{\frac{0,01}{2}, (25-3)} = \text{TINV}(0.01, 22) = 2,82 \text{ (chú ý: từ 1986 đến 2010 có } n=25 \text{ quan sát)}$$

Do $|T\text{-stat}[\ln(GI)]| > 2,82$ nên bác bỏ Ho, chấp nhận Ha ở độ tin cậy 99%; hay biến $\ln(GI)$ có ý nghĩa thống kê ở độ tin cậy 99%.

b. Bạn hãy cho biết biến $\ln(PI)$ có ý nghĩa thống kê ở độ tin cậy 95% hay không?

Bạn có thể làm tương tự câu a, nhưng với độ tin cậy 95% (khi đó $t_{\frac{\alpha}{2}, (n-k)} = t_{\frac{0,05}{2}, (25-3)}$

=TINV(0.05, 22) = 2,073873058), hoặc áp dụng theo quy tắc “2-t”

Ho : $\beta_3 = 0$

Ha : $\beta_3 \neq 0$

Do bậc tự do (df) là 22, lớn hơn 20; và $|t\text{-stat}[\ln(PI)]| = 28,17$ (lớn hơn 2) nên bác bỏ Ho, chấp nhận Ha ở mức ý nghĩa 5% (hay ở độ tin cậy 95%). Nói cách khác biến $\ln(PI)$ có ý nghĩa thống kê ở độ tin cậy 95%.

Học viên cũng có thể suy luận là $|t\text{-stat}[\ln(PI)]| = 28,17$ (lớn hơn 2,82) nên biến $\ln(PI)$ có ý nghĩa thống kê ở độ tin cậy 99%; Và suy ra, biến này cũng có ý nghĩa thống kê ở độ tin cậy 95%.

c. Bạn nhận định gì về mức độ hiệu quả của đầu tư công so với đầu tư tư nhân?

Hệ số hồi quy đứng trước biến $\ln(\text{PI})$ bằng 0,33 cho ta biết hệ số co giãn của Y theo PI; còn hệ số hồi quy đứng trước biến $\ln(\text{GI})$ bằng 0,23 cho ta biết hệ số co giãn của Y theo GI.

Trong điều kiện các yếu tố khác không đổi, nếu PI tăng (hoặc giảm) 1% thì về mặt trung bình, Y sẽ tăng (hoặc giảm) 0,33 %.

Trong điều kiện cố định các yếu tố khác, nếu GI tăng (hoặc giảm) 1%, thì trung bình Y sẽ tăng (hoặc giảm) 0,23%.

Do $\varepsilon_Y^{PI} > \varepsilon_Y^{GI}$ nên có thể nói rằng đầu tư tư nhân (PI) hiệu quả hơn so với đầu tư công (GI).

d. Có người cho rằng hệ số co giãn của GDP theo đầu tư công là rất lớn, và bằng 0,42 ; bạn nhận định gì về phát biểu này.

Bạn có thể thực hiện kiểm định hai phía ở độ tin cậy mà bạn chọn. Ví dụ, nếu xét ở độ tin cậy 99%.

$$H_0 : \beta_2 = 0,42$$

$$H_a : \beta_2 \neq 0,42$$

$$Se(\hat{\beta}_2) = 0,23/30,46 = 0,00755$$

$$\text{Trị thống kê } T = (0,23 - 0,42)/0,0075 = -25,16$$

$$t_{\frac{\alpha}{2}, (n-k)} = t_{\frac{0,01}{2}, (25-3)} = \text{TINV}(0,01, 22) \simeq 2,82$$

Do $|T| > 2,82$ nên bác bỏ H_0 , chấp nhận H_a ở độ tin cậy 99%; hay $\beta_2 \neq 0,42$

Có thể kiểm định bằng khoảng tin cậy, hoặc có thể thực hiện kiểm định 1 phía nếu học viên muốn thực hiện và lập luận theo cách này.

e. Bạn hãy ước lượng khoảng tin cậy 95% của hệ số hồi quy đứng trước biến $\ln(\text{PI})$

$$Se(\hat{\beta}_3) = 0,33/28,17 = 0,01173$$

$$t_{\frac{\alpha}{2}, (n-k)} = t_{\frac{0,05}{2}, (25-3)} = \text{TINV}(0,05, 22) \simeq 2,074$$

Khoảng tin cậy 95% của β_3 sẽ là $0,33 \pm 2,074 \cdot 0,01173$, hay (0,3057; 0,3543)

Bài 2 (50 điểm)

Mùa Noel đến, hơi lạnh một chút, được nhóm phân công, sau một hồi hì hục tìm kiếm lý thuyết và dữ liệu, anh Tèo và chị Tý đã thu thập được một số bài báo, một số dữ liệu từ 1990-2007 của biến Export (xuất khẩu), exrate (tỷ giá giữa đồng USD và đồng VND, ví dụ 1 USD = 16000 VND), GDPVN (GDP của Việt Nam), GDPWorld (GDP của thế giới), FDI (FDI vào Việt Nam) từ nguồn dữ liệu của Cơ quan thống kê Liên Hiệp Quốc và Tổng Cục thống kê của Việt Nam. Mục tiêu của anh Tèo, chị Tý là đi tìm một mô hình để dự báo cho giá trị xuất khẩu ở năm tiếp theo, và họ cũng muốn diễn giải các hệ số hồi quy. Đồ án nhóm sắp đến hạn nộp! Bạn cũng là một thành viên của nhóm, hãy cùng anh Tèo, chị Tý trả lời các câu hỏi sau:

a. Bạn kỳ vọng như thế nào về mối quan hệ giữa xuất khẩu và các biến exrate, GDPVN, GDPWorld, FDI

Có thể tìm được nhiều bài nghiên cứu xung quanh vấn đề này, ví dụ nghiên cứu của Fugazza (2004)¹. Ông đề cập đến mô hình dạng ln-ln. Và Mối liên hệ thuận chiều giữa Export các biến exrate, GDPVN, GDPWorld, FDI cũng như mối liên hệ giữa Export và một số biến khác.

Trong trường hợp này, ta có thể kỳ vọng rằng Export có mối liên hệ cùng chiều với các biến exrate, GDPVN, GDPWorld, FDI (căn cứ theo nghiên cứu của Marco Fugazza (2004))

Bạn có thể trả lời câu hỏi này tốt hơn khi kết hợp nhiều lý thuyết, nghiên cứu trước ... và bình luận kỹ hơn về mối quan hệ giữa các biến.

Tính toán các hệ số tương quan trên dữ liệu thu thập được (Xem Phụ lục 1 về dữ liệu được thu thập)

Covariance Analysis: Ordinary
Date: 12/12/13 Time: 19:12
Sample: 1 18
Included observations: 18

Correlation					
Probability	EXPORT	EXRATE	GDPVN	GDPWORLD	FDI
EXPORT	1,000000 -----				
EXRATE	0,819991 0,0000	1,000000 -----			
GDPVN	0,987702 0,0000	0,866231 0,0000	1,000000 -----		
GDPWORLD	0,991038 0,0000	0,813526 0,0000	0,987740 0,0000	1,000000 -----	
FDI	0,750029 0,0003	0,414765 0,0870	0,739510 0,0005	0,779658 0,0001	1,000000 -----

Hệ số tương quan sau khi chuyển về dạng Ln

Covariance Analysis: Ordinary
Date: 12/17/13 Time: 09:31
Sample: 1990 2007

¹ Marco Fugazza (2004), "Export performance and its determinants supply and demand constraints", Associate Economic Affairs Officer, UNITED NATIONS

Included observations: 18

Correlation					
Probability	LN_EXPORT	LN_EXRATE	LN_GDPVN	LN_GDPWO RLD	LN_FDI
LN_EXPORT	1,000000 -----				
LN_EXRATE	0,917366 0,0000	1,000000 -----			
LN_GDPVN	0,994281 0,0000	0,898772 0,0000	1,000000 -----		
LN_GDPWORLD	0,950480 0,0000	0,823194 0,0000	0,950999 0,0000	1,000000 -----	
LN_FDI	0,704596 0,0011	0,558612 0,0160	0,759850 0,0003	0,764020 0,0002	1,000000 -----

b. Bạn hãy xây dựng hàm hồi quy:

$$\ln(\text{Export}_i) = \hat{\beta}_1 + \hat{\beta}_2 \ln(\text{exrate}_i) + \hat{\beta}_3 \ln(\text{GDPVN}_i) + \hat{\beta}_4 \ln(\text{GDPWorld}_i) + \hat{u}_i$$

Bạn có nhận xét gì từ kết quả ước lượng. Nếu mô hình trên là mô hình U (Mô hình không giới hạn), mô hình R (mô hình giới hạn) của bạn sẽ là gì (xây dựng mô hình từ tổng quát đến đơn giản)? Bạn chọn mô hình U hay mô hình R?

Gõ lệnh: `ls log(export) c log(exrate) log(gdpvn) log(gdpworld) log(fdi)`

Dependent Variable: LOG(EXPORT)

Method: Least Squares

Date: 12/12/13 Time: 23:08

Sample: 1990 2007

Included observations: 18

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-13,86905	3,362210	-4,124979	0,0012
LOG(EXRATE)	0,307557	0,179007	1,718127	0,1095
LOG(GDPVN)	1,287176	0,109309	11,77560	0,0000
LOG(GDPWORLD)	0,488243	0,215132	2,269506	0,0409
LOG(FDI)	-0,143023	0,035852	-3,989241	0,0015
R-squared	0,996576	Mean dependent var		9,363134
Adjusted R-squared	0,995522	S,D, dependent var		1,013042
S.E. of regression	0,067789	Akaike info criterion		-2,314686
Sum squared resid	0,059740	Schwarz criterion		-2,067360
Log likelihood	25,83217	Hannan-Quinn criter,		-2,280583
F-statistic	945,8664	Durbin-Watson stat		2,292505
Prob(F-statistic)	0,000000			

Viết hàm hồi quy

Từ Bảng trên, bạn có thể viết được hàm hồi quy mẫu. Biến $\log(\text{exrate})$ có ý nghĩa ở mức ý nghĩa 11%, các biến giải thích khác đều có ý nghĩa ở mức ý nghĩa 10%. Dấu của các hệ số hồi quy $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\beta}_4$ đều phù hợp. Hệ số $\hat{\beta}_5 = -0,143$ mang dấu âm (trái dấu so với kỳ vọng). Bên cạnh đó, mô

hình có hệ số xác định cao, đồng thời trị tuyệt đối hệ số tương quan giữa các biến giải thích cũng có giá trị rất lớn (rất gần 1), điều này có thể do bị hiện tượng đa cộng tuyến.

Nếu ta loại biến $\log(\text{GDPVN})$ ra khỏi mô hình thì hệ số hồi quy, cũng như P-value của $\log(\text{FDI})$ thay đổi rất đáng kể, và mô hình này vẫn trực trặc do đa cộng tuyến, và $\log(\text{FDI})$ vẫn không có ý nghĩa thống kê.

Mô hình loại biến $\log(\text{GDPVN})$

Dependent Variable: LOG(EXPORT)

Method: Least Squares

Date: 12/12/13 Time: 23:10

Sample: 1990 2007

Included observations: 18

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-48,24852	5,488504	-8,790832	0,0000
LOG(EXRATE)	1,815691	0,411634	4,410936	0,0006
LOG(GDPWORLD)	2,324714	0,487741	4,766288	0,0003
LOG(FDI)	0,025114	0,108241	0,232023	0,8199
R-squared	0,960051	Mean dependent var		9,363134
Adjusted R-squared	0,951491	S,D, dependent var		1,013042
S.E. of regression	0,223121	Akaike info criterion		0,030927
Sum squared resid	0,696963	Schwarz criterion		0,228787
Log likelihood	3,721661	Hannan-Quinn criter,		0,058209
F-statistic	112,1491	Durbin-Watson stat		0,839665
Prob(F-statistic)	0,000000			

Nếu ta loại biến $\ln(\text{exrate})$ ra khỏi mô hình thì dấu của $\log(\text{FDI})$ vẫn là âm, trái dấu kỳ vọng; và mô hình bị mất đi một biến quan trọng về mặt lý thuyết (thường được đề cập trong các nghiên cứu, và lý thuyết về kinh tế học).

Mô hình loại biến $\log(\text{exrate})$

Dependent Variable: LOG(EXPORT)

Method: Least Squares

Date: 12/12/13 Time: 23:12

Sample: 1990 2007

Included observations: 18

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-11,08025	3,143024	-3,525347	0,0034
LOG(GDPVN)	1,421544	0,081519	17,43817	0,0000
LOG(GDPWORLD)	0,428298	0,226600	1,890107	0,0796
LOG(FDI)	-0,168153	0,034940	-4,812584	0,0003
R-squared	0,995798	Mean dependent var		9,363134
Adjusted R-squared	0,994898	S,D, dependent var		1,013042
S.E. of regression	0,072361	Akaike info criterion		-2,221165
Sum squared resid	0,073306	Schwarz criterion		-2,023304
Log likelihood	23,99048	Hannan-Quinn criter,		-2,193882
F-statistic	1105,971	Durbin-Watson stat		2,420143
Prob(F-statistic)	0,000000			

Nếu ta loại biến $\log(\text{GDPWorld})$ ra khỏi mô hình thì hệ số hồi quy, cũng như P-value của $\log(\text{exrate})$ thay đổi rất đáng kể, và mô hình này vẫn trục trặc do đa cộng tuyến, và $\log(\text{exrate})$ vẫn không có ý nghĩa thống kê. $\log(\text{FDI})$ dù có ý nghĩa thống kê nhưng lại ngược dấu kỳ vọng.

Mô hình loại biến $\log(\text{GDPWorld})$

Dependent Variable: LOG(EXPORT)

Method: Least Squares

Date: 12/12/13 Time: 23:13

Sample: 1990 2007

Included observations: 18

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6,699024	1,309851	-5,114341	0,0002
LOG(EXRATE)	0,241671	0,201125	1,201599	0,2495
LOG(GDPVN)	1,467015	0,085732	17,11161	0,0000
LOG(FDI)	-0,133307	0,040530	-3,289082	0,0054
R-squared	0,995219	Mean dependent var		9,363134
Adjusted R-squared	0,994195	S,D, dependent var		1,013042
S.E. of regression	0,077187	Akaike info criterion		-2,092040
Sum squared resid	0,083410	Schwarz criterion		-1,894179
Log likelihood	22,82836	Hannan-Quinn criter,		-2,064757
F-statistic	971,4323	Durbin-Watson stat		1,873863
Prob(F-statistic)	0,000000			

Nếu ta loại biến $\log(\text{FDI})$ thì kết quả hồi quy như sau:

Mô hình loại biến $\log(\text{FDI})$

Dependent Variable: LOG(EXPORT)

Method: Least Squares

Date: 12/12/13 Time: 23:26

Sample: 1990 2007

Included observations: 18

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-14,27361	4,829672	-2,955400	0,0104
LOG(EXRATE)	0,598890	0,234871	2,549865	0,0231
LOG(GDPVN)	1,113512	0,144093	7,727713	0,0000
LOG(GDPWORLD)	0,385767	0,306956	1,256751	0,2294
R-squared	0,992384	Mean dependent var		9,363134
Adjusted R-squared	0,990752	S,D, dependent var		1,013042
S.E. of regression	0,097421	Akaike info criterion		-1,626419
Sum squared resid	0,132872	Schwarz criterion		-1,428559
Log likelihood	18,63777	Hannan-Quinn criter,		-1,599137
F-statistic	608,0750	Durbin-Watson stat		1,052482
Prob(F-statistic)	0,000000			

Mô hình hồi quy ở trên vẫn có khả năng bị đa cộng tuyến, tuy nhiên dấu của các hệ số hồi quy là hợp lý và các biến giải thích đều có ý nghĩa thống kê ở mức ý nghĩa 5% ngoại trừ **LOG(GDPWORLD)**.

*Chương 5 của Ramanathan đề cập rằng: khi một mô hình bị cộng tuyến, nếu một biến rất quan trọng về mặt lý thuyết và P-value của nó nhỏ hơn 0,25 thì có thể được xem xét giữ lại trong mô hình, ở đây xem xét giữ lại biến **LOG(GDPWORLD)**.*

Vậy ta lựa chọn mô hình này.²

c. Anh Tèo dẫn đo suy nghĩ không biết nên chọn mô hình nào trong 2 mô hình sau:

$$\text{Mô hình 1: } \ln(\text{Export}_i) = \hat{\beta}_1 + \hat{\beta}_2 \ln(\text{exrate}_i) + \hat{\beta}_3 \ln(\text{GDPVN}_i) + \hat{\beta}_4 \ln(\text{GDPWorld}_i) + \hat{u}_i$$

$$\text{Mô hình 2: } \text{Export}_i = \hat{\alpha}_1 + \hat{\alpha}_2 \text{exrate}_i + \hat{\alpha}_3 \text{GDPVN}_i + \hat{\alpha}_4 \text{GDPWorld}_i + v_i$$

Bạn hãy đưa ra lời khuyên cho anh Tèo là nên chọn mô hình nào?

Mô hình 1:

Dependent Variable: LOG(EXPORT)

Method: Least Squares

Date: 12/12/13 Time: 23:22

Sample: 1990 2007

Included observations: 18

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-14,27361	4,829672	-2,955400	0,0104
LOG(EXRATE)	0,598890	0,234871	2,549865	0,0231
LOG(GDPVN)	1,113512	0,144093	7,727713	0,0000
LOG(GDPWORLD)	0,385767	0,306956	1,256751	0,2294
R-squared	0,992384	Mean dependent var		9,363134
Adjusted R-squared	0,990752	S,D, dependent var		1,013042
S.E. of regression	0,097421	Akaike info criterion		-1,626419
Sum squared resid	0,132872	Schwarz criterion		-1,428559
Log likelihood	18,63777	Hannan-Quinn criter,		-1,599137
F-statistic	608,0750	Durbin-Watson stat		1,052482
Prob(F-statistic)	0,000000			

Mô hình 2:

Dependent Variable: EXPORT

Method: Least Squares

Date: 12/12/13 Time: 19:21

Sample: 1 18

Included observations: 18

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-22109,72	10149,95	-2,178308	0,0470
EXRATE	-0,284098	0,427060	-0,665241	0,5167
GDPVN	0,412258	0,236722	1,741529	0,1035
GDPWORLD	0,000926	0,000411	2,256130	0,0406
R-squared	0,985794	Mean dependent var		17577,87
Adjusted R-squared	0,982750	S,D, dependent var		15232,03
S.E. of regression	2000,569	Akaike info criterion		18,23338
Sum squared resid	56031843	Schwarz criterion		18,43124
Log likelihood	-160,1004	Hannan-Quinn criter,		18,26066
F-statistic	323,8340	Durbin-Watson stat		0,468177
Prob(F-statistic)	0,000000			

² Chú ý: Học viên có thể có những lập luận khác, tuy vậy cần phải có lập luận hợp lý để chọn ra được mô hình tốt nhất

Dễ dàng lựa chọn mô hình 1 là hợp lý hơn, bởi lẽ mô hình 1 giữ được các biến quan trọng về mặt lý thuyết, đồng thời các biến giải thích đều có ý nghĩa thống kê, và có dấu phù hợp (trừ biến LOG(GDPWORLD) tuy nhiên vẫn được lựa chọn)³

Trong khi mô hình 2 có khả năng bị đa cộng tuyến mạnh hơn, có 2 biến giải thích không có ý nghĩa thống kê ở mức ý nghĩa 5%, biến exrate lại trái dấu kỳ vọng và Pvalue rất cao, lên tới 51,67%.

d. Giả sử anh Tèo đã chọn Mô hình 1 ở câu c. Và chị Tý lại làm Tèo phân vân. Chị Tý nói rằng: khi học môn Kinh tế học, chỉ có tỷ giá là có thể ảnh hưởng đến xuất khẩu thôi, dạng tuyến tính theo các biến hay dạng ln theo các biến cũng được, chứ GDPVN và GDPWorld đâu có liên quan gì đến xuất khẩu. Tèo không biết có nên bỏ đi cả hai biến ln(GDPVN) và ln(GDPWorld) hay không. Anh chị hãy gỡ rối giúp Tèo.

Từ Mô hình 1 ở câu c, Tèo có thể thực hiện kiểm định Wald để đưa ra kết luận.

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	138,6230	(2, 14)	0,0000
Chi-square	277,2461	2	0,0000

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err,
C(3)	1,113512	0,144093
C(4)	0,385767	0,306956

Restrictions are linear in coefficients.

Mô hình 1: $\ln(\text{Export}_i) = \beta_1 + \beta_2 \ln(\text{exrate}_i) + \beta_3 \ln(\text{GDPVN}_i) + \beta_4 \ln(\text{GDPWorld}_i) + u_i$

Ho : $\beta_3 = \beta_4 = 0$

Ha : Trong hai tham số β_3, β_4 ; có ít nhất một tham số khác 0

P-value(Wald)=0,000 (<0,05) nên bác bỏ Ho, chấp nhận Ha ở mức ý nghĩa 5%. Nói cách khác, không nên loại đi cả hai biến ln(GDPVN) và ln(GDPWorld)

e. Nếu anh Tèo xây dựng mô hình từ tổng quát đến đơn giản (Hendry/LSE), thì chị Tý lại thích làm ngược lại (tuần trước, lúc ngồi ở gốc cây me, Tèo hứa là sau khi làm xong bài tập kinh tế lượng sẽ dẫn chị đi uống bia mà lại quên !). Giờ đây, chị muốn “tìm ra chân lý” theo cách riêng của chị, đó là xây dựng mô hình từ đơn giản đến tổng quát. Chị đã ước lượng được mô hình dưới đây:

$$\ln(\text{Export}_i) = \hat{\beta}_1 + \hat{\beta}_2 \ln(\text{exrate}_i) + \hat{u}_i$$

³ Giống như đã bình luận ở trên: *Chương 5 của Ramanathan đề cập rằng: khi một mô hình bị cộng tuyến, nếu một biến rất quan trọng về mặt lý thuyết và P-value của nó nhỏ hơn 0,25 thì có thể được xem xét giữ lại trong mô hình.*

Và chị đang phân vân không biết có nên đưa biến $\ln(\text{GDPVN})$ vào tiếp hay không? Bạn hãy tư vấn giúp chị Tỷ.

Cách 1:

Mô hình đang có của chị Tỷ

Dependent Variable: LOG(EXPORT)

Method: Least Squares

Date: 12/12/13 Time: 23:33

Sample: 1990 2007

Included observations: 18

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-27.84895	4.037753	-6.897141	0.0000
LOG(EXRATE)	3.938909	0.427271	9.218752	0.0000
R-squared	0.841561	Mean dependent var		9.363134
Adjusted R-squared	0.831659	S.D. dependent var		1.013042
S.E. of regression	0.415645	Akaike info criterion		1.186470
Sum squared resid	2.764176	Schwarz criterion		1.285400
Log likelihood	-8.678231	Hannan-Quinn criter.		1.200111
F-statistic	84.98539	Durbin-Watson stat		0.969194
Prob(F-statistic)	0.000000			

Nếu đưa tiếp biến GDPVN vào:

Dependent Variable: LOG(EXPORT)

Method: Least Squares

Date: 12/12/13 Time: 23:34

Sample: 1990 2007

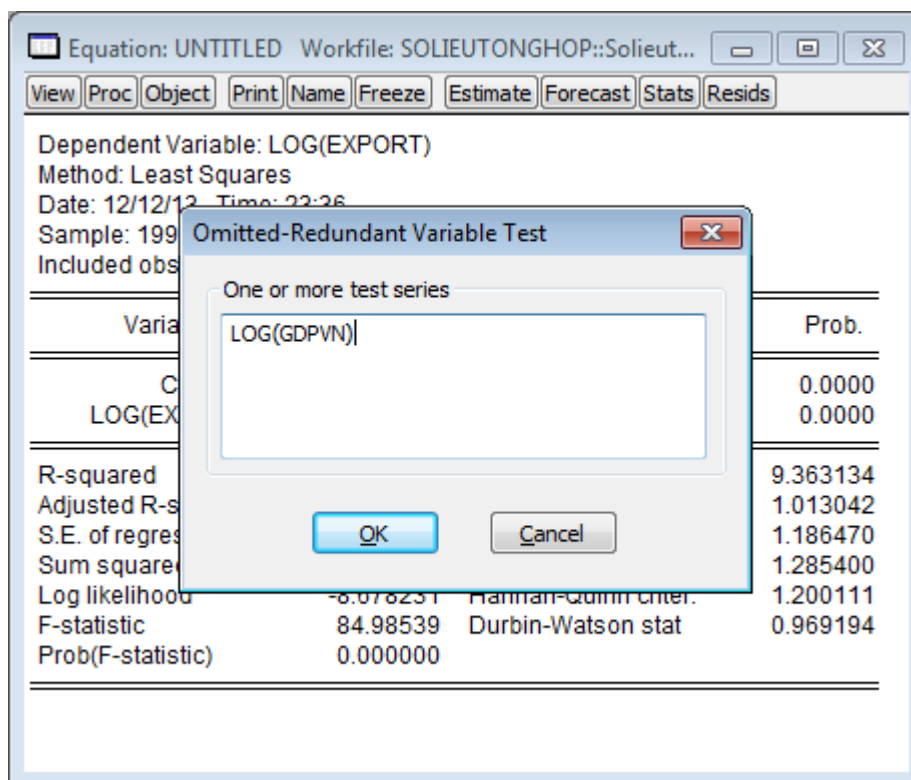
Included observations: 18

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-8.504508	1.529755	-5.559394	0.0001
LOG(EXRATE)	0.530217	0.232796	2.277598	0.0378
LOG(GDPVN)	1.267116	0.077778	16.29154	0.0000
R-squared	0.991525	Mean dependent var		9.363134
Adjusted R-squared	0.990395	S.D. dependent var		1.013042
S.E. of regression	0.099285	Akaike info criterion		-1.630636
Sum squared resid	0.147862	Schwarz criterion		-1.482241
Log likelihood	17.67573	Hannan-Quinn criter.		-1.610175
F-statistic	877.4292	Durbin-Watson stat		0.997754
Prob(F-statistic)	0.000000			

Vì mô hình 2 có R^2 đã hiệu chỉnh lớn hơn so với mô hình 1, và biến $\ln(\text{gdpvn})$ có ý nghĩa thống kê nên mô hình 2 tốt hơn.

Cách 2:

Hoặc sau khi ước lượng mô hình 1, trên Eviews bạn thực hiện kiểm định bỏ sót biến: Equation: View\Coefficient tests\Omitted Variables-Likelihood Ratio → nhập biến $\log(\text{gdpvn})$.



Omitted Variables: LOG(GDPVN)

F-statistic	265,4142	Prob, F(1,15)	0,0000
Log likelihood ratio	52,70792	Prob, Chi-Square(1)	0,0000

Test Equation:

Dependent Variable: LOG(EXPORT)

Method: Least Squares

Date: 12/12/13 Time: 23:37

Sample: 1990 2007

Included observations: 18

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-8,504508	1,529755	-5,559394	0,0001
LOG(EXRATE)	0,530217	0,232796	2,277598	0,0378
LOG(GDPVN)	1,267116	0,077778	16,29154	0,0000

R-squared	0,991525	Mean dependent var	9,363134
Adjusted R-squared	0,990395	S,D, dependent var	1,013042
S.E. of regression	0,099285	Akaike info criterion	-1,630636
Sum squared resid	0,147862	Schwarz criterion	-1,482241
Log likelihood	17,67573	Hannan-Quinn criter,	-1,610175
F-statistic	877,4292	Durbin-Watson stat	0,997754
Prob(F-statistic)	0,000000		

Bảng kết quả cho thấy P-value(F)=0,000 (<0,05) nên mô hình 1 đã bị bỏ sót biến ln(GDPVN) -> Cần bổ sung thêm biến này vào mô hình.

Cách 3:

Cách này chỉ áp dụng với trường hợp mẫu lớn, mẫu trong bài không đủ lớn tuy nhiên cách này vẫn được nêu ra để tham khảo.

Kiểm định nhân tử Lagrange

$$(R): \ln(\text{Export}_i) = \hat{\beta}_1 + \hat{\beta}_2 \ln(\text{exrate}_i) + \hat{u}_i$$

$$(U): \ln(\text{Export}_i) = \hat{\beta}_1 + \hat{\beta}_2 \ln(\text{exrate}_i) + \hat{\beta}_3 \ln(\text{GDPVN}) + \hat{v}_i$$

Kiểm định giả thuyết

$$H_0: \hat{\beta}_3 = 0$$

$$H_a: \hat{\beta}_3 \neq 0$$

-Bước 1: Ước lượng mô hình (R)

Dependent Variable: LOG(EXPORT)

Method: Least Squares

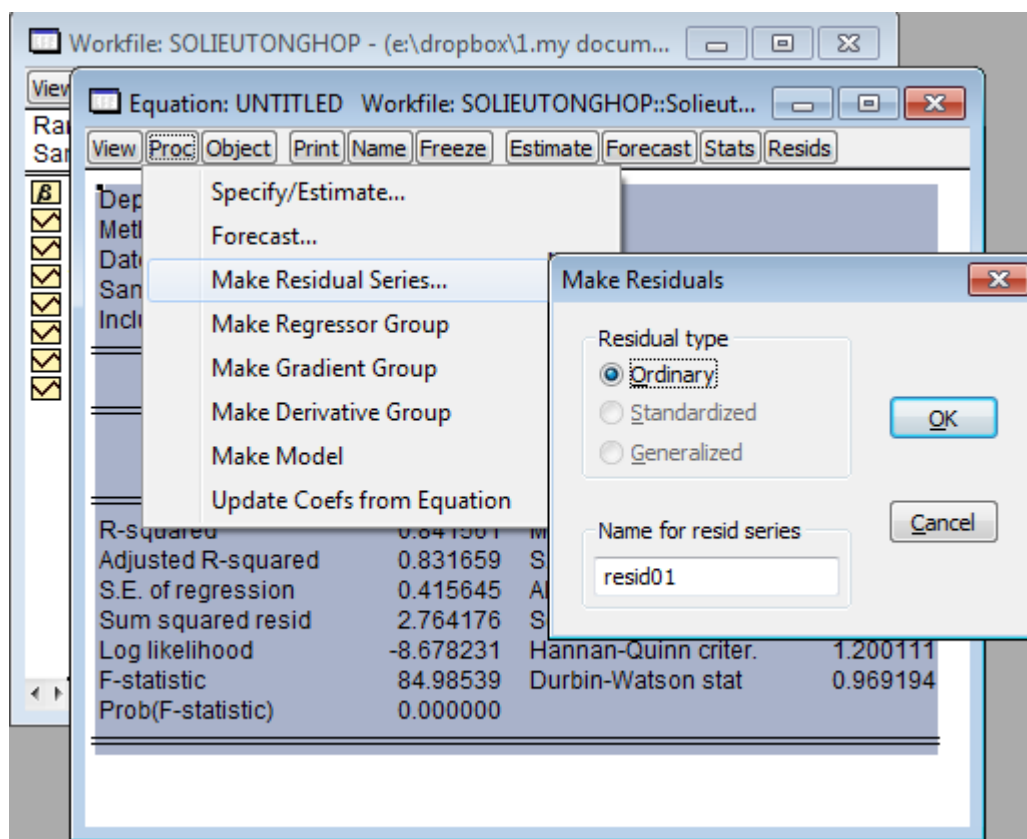
Date: 12/13/13 Time: 09:22

Sample: 1990 2007

Included observations: 18

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-27,84895	4,037753	-6,897141	0,0000
LOG(EXRATE)	3,938909	0,427271	9,218752	0,0000
R-squared	0,841561	Mean dependent var		9,363134
Adjusted R-squared	0,831659	S,D, dependent var		1,013042
S.E. of regression	0,415645	Akaike info criterion		1,186470
Sum squared resid	2,764176	Schwarz criterion		1,285400
Log likelihood	-8,678231	Hannan-Quinn criter,		1,200111
F-statistic	84,98539	Durbin-Watson stat		0,969194
Prob(F-statistic)	0,000000			

-Bước 2: Tính phần dư $\hat{u}_R$



Ta được phần dư resid01.

-Bước 3: Ước lượng mô hình: $\hat{u}_{Ri} = \hat{\alpha}_1 + \hat{\alpha}_2 \ln(exrate_i) + \hat{\alpha}_3 \ln(GDPVN) + \hat{\theta}_i$

Dependent Variable: RESID01
Method: Least Squares
Date: 12/13/13 Time: 09:30
Sample: 1990 2007
Included observations: 18

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	19,34444	1,529755	12,64545	0,0000
LOG(EXRATE)	-3,408692	0,232796	-14,64237	0,0000
LOG(GDPVN)	1,267116	0,077778	16,29154	0,0000
R-squared	0,946508	Mean dependent var		1,81E-14
Adjusted R-squared	0,939375	S,D, dependent var		0,403235
S.E. of regression	0,099285	Akaike info criterion		-1,630636
Sum squared resid	0,147862	Schwarz criterion		-1,482241
Log likelihood	17,67573	Hannan-Quinn criter,		-1,610175
F-statistic	132,7071	Durbin-Watson stat		0,997754
Prob(F-statistic)	0,000000			

-Bước 4: giả sử mẫu lớn, $n \cdot R^2 = 18 \cdot 0,946508 = 17,037144$

Tra bảng phân phối Chi-Square với bậc tự do là 1, mức ý nghĩa 5% ta có $\chi^2 = 3,841$

Vì $n \cdot R^2 > \chi^2$ nên bác bỏ giả thuyết $H_0 \Rightarrow$ nên đưa biến LOG(GDPVN) vào mô hình.

PHỤ LỤC 1: BẢNG DỮ LIỆU THU THẬP

Year	GDP World (triệu USD)	Export (triệu USD)	GDPVN (triệu USD)	FDI (triệu USD)	Exrate (VND/USD)
1990	22,595,357.99	1,709.76	6,471.74	856.27	6,482.80
1991	23,710,647.31	2,362.65	7,642.40	1,284.40	10,037.03
1992	25,366,624.45	3,428.35	9,867.00	2,077.60	11,202.19
1993	25,866,212.80	3,785.84	13,180.96	2,829.80	10,640.96
1994	27,610,498.78	5,537.74	16,281.18	4,262.10	10,965.67
1995	30,454,517.43	6,804.16	20,736.26	7,925.20	11,038.25
1996	31,079,757.22	10,077.15	24,657.51	9,635.30	11,032.58
1997	30,989,037.75	11,570.33	26,843.62	5,955.60	11,683.33
1998	30,753,804.29	12,203.04	27,209.60	4,873.40	13,268.00
1999	31,952,428.72	14,332.18	28,683.73	2,282.50	13,943.17
2000	32,859,658.10	17,155.09	31,172.63	2,762.80	14,167.75
2001	32,672,873.83	17,850.12	32,685.20	3,265.70	14,725.17
2002	33,990,572.18	19,913.09	35,064.11	2,993.40	15,279.50
2003	38,156,597.67	23,452.27	39,552.50	3,172.70	15,509.58
2004	42,949,906.74	29,862.57	45,427.85	4,534.30	15,746.00
2005	46,509,522.89	36,712.10	57,633.26	6,840.00	15,858.92
2006	50,386,809.93	45,011.65	66,371.66	12,004.50	15,994.25
2007	56,676,991.14	55,038.05	77,414.43	21,348.80	16,105.13

Nguồn: Tổng cục thống kê; Cơ quan thống kê Liên Hiệp Quốc truy cập tại địa chỉ:

<http://www.gso.gov.vn/default.aspx?tabid=392&idmid=3&ItemID=14346>

<http://unstats.un.org/unsd/snaama/dnlList.asp>