

Chương 6:
KIỂM ĐỊNH MÔ HÌNH

Th.S NGUYỄN PHƯƠNG

Bộ môn Toán kinh tế
Trường Đại học Ngân hàng TPHCM
Blog: <https://nguyenphuongblog.wordpress.com>
Email: nguyenphuong0122@gmail.com

Ngày 18 tháng 9 năm 2016

NỘI DUNG

- 1 Kỳ vọng của sai số ngẫu nhiên khác không
 - Nguyên nhân
 - Hậu quả của kỳ vọng sai số ngẫu nhiên khác không
 - Phát hiện về sự khác không của kỳ vọng sai số ngẫu nhiên
 - Một số biện pháp khắc phục
- 2 Phương sai sai số thay đổi (Heteroscedasticity)
 - Nguyên nhân
 - Hậu quả của phương sai sai số thay đổi
 - Phát hiện phương sai sai số thay đổi
 - Khắc phục hiện tượng phương sai sai số thay đổi
- 3 Đa cộng tuyến (Multicollinearity)
 - Bản chất đa cộng tuyến
 - Nguyên nhân và hậu quả
 - Cách phát hiện đa cộng tuyến cao
 - Khắc phục hiện tượng đa cộng tuyến
- 4 Sai số ngẫu nhiên không tuân theo quy luật chuẩn

Giả thiết 2 của mô hình hồi quy tuyến tính là

$$E(u|X_2, \dots, X_k) = 0.$$

Nếu giả thiết này thỏa mãn thì có

$$E(u) = 0 \quad \text{và} \quad \text{cov}(X_j, u) = 0, \forall j = 2, \dots, k.$$

Nguyên nhân

- Mô hình "thiếu biến quan trọng" (omit variable). Mô hình được cho là thiếu biến quan trọng Z nếu:
 - Biến Z có tác động đến biến phụ thuộc Y .
 - Biến Z có tương quan với $X_j, j = 2, 3, \dots, k$

Khi đó $\text{cov}(X_j, u) \neq 0$.

- Dạng hàm sai (functional form misspecification)
- Tính tác động đồng thời của số liệu
- Sai số đo lường của các biến độc lập.

Hậu quả của kỳ vọng sai số ngẫu nhiên khác không

- ▶ Ước lượng OLS sẽ là ước lượng chệch, tức là $E(\hat{\beta}_j) \neq \beta_j$.
- ▶ Nếu mô hình thiếu biến quan trọng Z thì UL OLS không vững.
- ▶ Các suy diễn thống kê không còn đáng tin cậy

- Nếu mô hình bỏ sót biến quan trọng:
Giả sử muốn biết mô hình

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + u$$

có bỏ sót "biến quan trọng Z" hay không ta hồi quy mô hình

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + \alpha_{k+1} Z + u.$$

Sau đó kiểm định cặp giả thuyết:

$$H_0 : \alpha_{k+1} = 0; \quad H_1 : \alpha_{k+1} \neq 0.$$

Nếu bác bỏ H_0 thì chấp nhận biến Z có tác động đến Y và mô hình đã thiếu "biến quan trọng Z".

- Nếu mô hình có dạng hàm sai
 - Sử dụng Kiểm định Ramsey (để phát hiện mô hình có dạng hàm sai do thiếu biến là hàm mũ của các biến có sẵn trong mô hình)
 - Sử dụng Kiểm định Davidson-MacKinnon (Kiểm định J)

Ví dụ 1.1

Sử dụng bộ số liệu ch5vd3.wf1 thu được kết quả ước lượng sau:

$$CT = 42,73 + 0,85TN + e$$

Để kiểm định xem mô hình có khuyến nghị bỏ sót biến TS hay không, ta thực hiện kiểm định.

Equation: UNTITLED Workfile: CH2VD5:Untitl...				
View	Proc	Object	Print Name Freeze Estimate Forecast Stats Resids	
Omitted Variables: TS				
F-statistic	15.76310	Prob. F(1,30)	0.0004	
Log likelihood ratio	13.93526	Prob. Chi-Square(1)	0.0002	
Test Equation:				
Dependent Variable: CT				
Method: Least Squares				
Date: 09/06/15 Time: 21:57				
Sample: 1 33				
Included observations: 33				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	18.86018	8.832144	2.135402	0.0410
TN	0.791224	0.015991	49.48055	0.0000
TS	0.015818	0.003984	3.970277	0.0004
R-squared	0.999545	Mean dependent var	1610.415	
Adjusted R-squared	0.999515	S.D. dependent var	557.2878	
S.E. of regression	12.27498	Akaike info criterion	7.939512	
Sum squared resid	4520.257	Schwarz criterion	8.075558	
Log likelihood	-128.0019	Hannan-Quinn criter.	7.985287	
F-statistic	32963.98	Durbin-Watson stat	1.912696	
Prob(F-statistic)	0.000000			

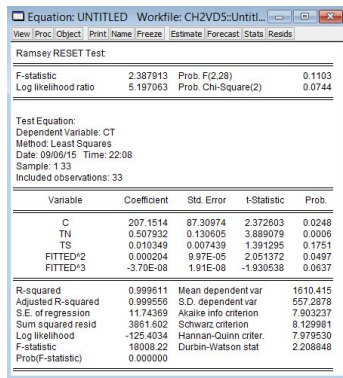
Hình: Kiểm định thiếu biến

Ví dụ 1.2

Trong ví dụ trước, ta thấy mô hình bỏ sót biến tài sản (TS). Khi thêm biến TS vào mô hình:

$$CT = \beta_1 + \beta_2 TN + \beta_3 TS + u$$

Câu hỏi: có vấn đề về dạng hàm sai hay không?



Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	207.1514	87.30974	2.372603	0.0248
TN	0.507932	0.130605	3.889079	0.0006
TS	0.010349	0.007439	1.391295	0.1751
FITTED^2	0.000204	9.97E-05	2.051372	0.0497
FITTED^3	-3.70E-08	1.91E-08	-1.930538	0.0637

R-squared	0.999611	Mean dependent var	1610.415
Adjusted R-squared	0.999556	S.D. dependent var	557.2878
S.E. of regression	11.74369	Akaike info criterion	7.903237
Sum squared resid	3861.602	Schwarz criterion	8.129981
Log likelihood	-125.4034	Hannan-Quinn criter.	7.979530
F-statistic	18008.22	Durbin-Watson stat	2.208848
Prob(F-statistic)	0.000000		

Hình: Kiểm định Ramsy RESET

Một số biện pháp khắc phục khi mô hình có kỳ vọng của sai số ngẫu nhiên khác 0

- Nếu mô hình thiếu biến quan trọng Z (đã biết) được phát hiện từ kiểm định t thì ta thêm biến Z vào mô hình.
- Nếu mô hình có dạng hàm sai được phát hiện từ kiểm định Ramsey thì xét các mô hình thay thế.
- Nếu mô hình thiếu biến không quan sát được thì có thể sử dụng hai phương pháp:
 - Sử dụng biến đại diện (proxy variable)
 - Sử dụng biến công cụ (instrumental variable)

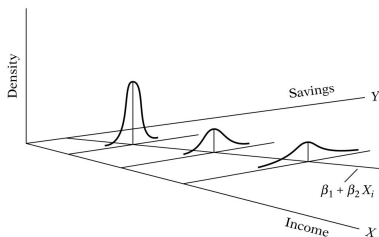
Xét mô hình hồi quy:

$$Y = \beta_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u.$$

Định lý Gause-Markov khẳng định rằng để ước lượng OLS là tốt nhất thì phương sai sai số trong mô hình hồi quy phải bằng nhau tại mọi quan sát. Khi giả thiết này không thỏa mãn thì mô hình có hiện tượng phương sai sai số thay đổi, tức là:

$$\text{var}(u_j | X_{2i}, X_{3i}, \dots, X_{ki}) = \sigma_i^2, i = 1, 2, \dots, n;$$

nghĩa là tại các bộ giá trị $(X_{2i}, X_{3i}, \dots, X_{ki})$ khác nhau thì phương sai của sai số ngẫu nhiên u_i nhận các giá trị khác nhau, kí hiệu là σ^2 .



Nguyên nhân

- ▶ Do bản chất của các hiện tượng kinh tế
 - ✓ Số liệu theo không gian có quy mô khác nhau
 - quan sát các doanh nghiệp có quy mô quá lớn - quá nhỏ
 - ✓ Số liệu theo thời gian qua các giai đoạn có mức biến động khác nhau
 - giai đoạn kinh tế ổn định - giai đoạn khủng hoảng
 - ✓ Mỗi quan hệ sản có hiện tượng PSSS thay đổi
 - mỗi quan hệ của thu nhập – chi tiêu, chỉ số thị trường chứng khoán
- ▶ Do số liệu không phản ánh đúng bản chất của hiện tượng kinh tế
- ▶ Do kỹ thuật thu thập, xử lý dữ liệu ngày càng được hoàn thiện nên sai số ngày càng ít
- ▶ Do hành vi của con người có sự tiếp thu từ quá khứ
- ▶ Do định dạng không đúng dạng hàm của mô hình