

KINH TẾ LƯỢNG - ECONOMETRICS

Tài liệu

- [1]. Nguyễn Quang Dong, Nguyễn Thị Minh (2012), *Giáo trình Kinh tế lượng*, NXB Thống kê
 - [2]. Nguyễn Cao Văn, Trần Thái Ninh, (1998), *Lý thuyết Xác suất và Thống kê toán*, NXB GD.(Tái bản các năm 2002, 2005)
 - [3]. Nguyễn Quang Dong, (2002), *Kinh tế lượng - Chương trình nâng cao*, NXB KHKT.
 - [4]. Nguyễn Quang Dong, (2002), *Bài tập Kinh tế lượng với sự trợ giúp của phần mềm Eviews*, NXB KHKT.
 - [5]. Nguyễn Khắc Minh, (2002), *Các phương pháp Phân tích & Dự báo trong Kinh tế*, NXB KHKT.
 - [6]. Graham Smith, (1996), *Econometric Analysis and Applications*, London University.
 - [7] D. Gujarati. *Basic Econometrics*. Third Edition. McGraw-Hill, Inc 1996.
 - [8] Maddala. *Introduction to Econometrics* . New york 1992.
 - [9] W. Green. *Econometric Analysis*. New york 2005.
-

BÀI MỞ ĐẦU

1. Khái niệm về Kinh tế lượng (*Econometrics*)

- Nhiều định nghĩa, tùy theo quan niệm của mỗi tác giả.
- *Econo + Metric*

Khái niệm: KTL nghiên cứu những mối quan hệ Kinh tế Xã hội; thông qua việc xây dựng, phân tích, đánh giá các mô hình để cho ra lời giải bằng số, hỗ trợ việc ra quyết định

***Kinh tế l- ượng là kinh tế học thực chứng
Econometrics – Pragmatic Economics***

- KTL sử dụng kết quả của :
 - + Lý thuyết kinh tế
 - + Mô hình toán kinh tế
 - + Thống kê, xác suất

2. Phương pháp luận (các bước tiến hành)

2.1. Đặt luận thuyết về vấn đề nghiên cứu

- Xác định phạm vi, bản chất, tính chất của các đối tượng và mối quan hệ giữa chúng.
- Xác định mô hình lý thuyết kinh tế hợp lý.

2.2. Xây dựng mô hình kinh tế toán:

- + Mỗi đối tượng đại diện bởi một hoặc một số biến số.
- + Mỗi mối quan hệ: Phương trình, hàm số, bất phương trình...
- + Giá trị các tham số: cho biết bản chất mối quan hệ.

2.3. Xây dựng mô hình kinh tế l- ợng t- ơng ứng

- Mô hình kinh tế toán: phụ thuộc hàm số
- Mô hình kinh tế l- ợng: phụ thuộc t- ơng quan và hồi quy

2.4. Thu thập số liệu

- Số liệu được dùng : từ thống kê.

2.5. Ước lượng các tham số của mô hình.

- Với bộ số liệu xác định và phương pháp cụ thể, kết quả ước lượng là những con số cụ thể.

2.6. Kiểm định mô hình.

- Bằng phương pháp kiểm định thống kê: kiểm định giá trị các tham số, bản chất mối quan hệ
- Kiểm định tính chính xác của mô hình.
- Nếu không phù hợp: quay lại các bước trên.
- Biến đổi, xây dựng mô hình mới để có kết quả tốt nhất.

2.7. Dự báo

- Dựa trên kết quả được cho là tốt : dự báo về mối quan hệ, về các đối tượng trong những điều kiện xác định.

2.8. Kiểm soát và Đề xuất chính sách.

- Dựa vào kết quả phân tích của mô hình mà đề xuất chính sách kinh tế.

Ví dụ: Nghiên cứu tính quy luật của tiêu dùng.

1. Xây dựng một luận thuyết kinh tế về tiêu dùng.

Trong tác phẩm: Lý thuyết về việc làm, lãi suất và tiền tệ, Keynes viết: "Luật tâm lý cơ bản . . . là một ng-ời sẽ tăng tiêu dùng khi thu nhập của ng-ời đó tăng lên, song không thể tăng nhiều bằng mức tăng của thu nhập"

2. Xây dựng mô hình kinh tế toán t- ơng ứng.

Ký hiệu: Y là tiêu dùng

X là thu nhập

Và giả sử Y phụ thuộc tuyến tính vào X . Ta có mô hình kinh tế toán sau đây:

$$Y = \beta_1 + \beta_2 X$$

Mô hình trên th- ờng đ- ợc gọi là *Hàm tiêu dùng* của Keynes và phải thoả mãn điều kiện:

$$0 < \beta_2 < 1$$

3. Xây dựng mô hình kinh tế l- ợng t- ơng ứng.

Mô hình kinh tế l- ợng t- ơng ứng có dạng:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Trong đó u_i là sai số ngẫu nhiên.

4. Thu thập số liệu thống kê.

Có số liệu sau về tổng mức tiêu dùng cá nhân (Y) và tổng thu nhập gộp GDP (X) của Mỹ giai đoạn 1980 — 1991 (đơn vị: tỷ USD) tính theo giá cố định năm 1987:

Năm	Y	X
1980	2447.1	3776.3
1981	2476.9	3843.1
1982	2503.7	3760.3
1983	2619.4	3906.6
1984	2746.1	4148.5
1985	2865.8	4279.8
1986	2969.1	4404.5
1987	3052.2	4539.9
1988	3162.4	4718.6
1989	3223.3	4838.0
1990	3260.4	4877.5
1991	3240.8	4821.0

Nguồn: Báo cáo kinh tế của tổng thống Mỹ, 1993.

5. Ước lượng mô hình.

Dùng phương pháp bình phương nhỏ nhất, tìm được các ước lượng sau:

$$\hat{b}_1 = -231,8 \quad \hat{b}_2 = 0,7194$$

Như vậy - ước lượng của hàm tiêu dùng là:

$$\hat{Y}_i = -231,8 + 0,7194X_i$$

6. Kiểm định mô hình:

$$H_0: \beta_2 = 0$$

$$H_1: \beta_2 > 0$$

$$H_0: \beta_2 = 1$$

$$H_1: \beta_2 < 1$$

H_0 : Mô hình có dạng tuyến tính

H_1 : Mô hình có dạng phi tuyến

H_0 : Sai số ngẫu nhiên phân phối chuẩn

H_1 : Sai số ngẫu nhiên không phân phối chuẩn . . .

Mục đích của kiểm định là kiểm chứng lại mô hình hoặc lý thuyết kinh tế.

7. Dự báo.

Chẳng hạn có cơ sở để cho rằng GDP của Mỹ vào năm 1994 là 6000 tỷ USD. Lúc đó có thể tìm được một dự báo điểm cho Tổng mức tiêu dùng cá nhân của Mỹ vào năm đó là:

$$\hat{y}_{1994} \approx -231,8 + 0,7194 \cdot 6000 = 4084,6 \text{ tỷ USD.}$$

Từ đó có thể xây dựng tiếp các dự báo bằng khoảng tin cậy.

8. Kiểm soát hoặc đề xuất chính sách.

Chẳng hạn chính phủ Mỹ tin rằng nếu có được tổng mức tiêu dùng cá nhân là 4000 tỷ USD thì sẽ duy trì được tỷ lệ thất nghiệp ở mức 6,5%. Từ đó để duy trì được tỷ lệ thất nghiệp nói trên cần phải có được GDP là:

$$GDP \approx (4000 + 231,8) / 0,7194 \approx 5882 \text{ tỷ USD.}$$

3. Số liệu dùng trong KTL

3.1. Phân loại

- Số liệu theo thời gian.
- Số liệu theo không gian.
- Số liệu chéo

3.1. Nguồn gốc

- Điều tra
- Mua
- Từ nguồn được phát hành : Niên giám thống kê

3.2. Tính chất của số liệu

- Số liệu ngẫu nhiên phi thực nghiệm.
- Phù hợp mục đích nghiên cứu.

Chú ý: Đặc điểm chung của các số liệu kinh tế xã hội là kém tin cậy

Bài 1. CÁC KHÁI NIỆM CƠ BẢN

1. Phân tích hồi qui – *Regression Analysis*

1.1. Định nghĩa

Phân tích hồi qui là phân tích mối liên hệ phụ thuộc giữa một biến gọi là biến phụ thuộc (biến được giải thích, biến nội sinh) phụ thuộc vào một hoặc một số biến khác gọi là (các) biến giải thích (biến độc lập, biến ngoại sinh, biến hồi qui).

1.2. Ví dụ Tiêu dùng và Thu nhập.

- Biến phụ thuộc (*dependent variable*) ký hiệu là Y
- Biến giải thích(*Explaine variable(s)*) / hồi qui (*regressor(s)*) ký hiệu là X , hoặc $X_2, X_3, \dots$
- Biến giải thích nhận những giá trị xác định, trong điều kiện đó biến phụ thuộc là một biến ngẫu nhiên.

Phân tích hồi qui nghiên cứu mối liên hệ phụ thuộc giữa biến phụ thuộc Y mà thực chất là một biến ngẫu nhiên, phụ thuộc vào các giá trị xác định của (các) biến giải thích như thế nào.

$$X = X_i \rightarrow (Y/X_i)$$

Cuu duong than cong . com

1.3. Mục đích hồi qui

- Ước lượng trung bình biến phụ thuộc trong những điều kiện xác định của biến giải thích.
- Ước lượng các tham số.
- Kiểm định về mối quan hệ.
- Dự báo giá trị biến phụ thuộc khi biến giải thích thay đổi.

(*) *Hồi qui : qui về trung bình*

Cuu duong than cong . com

1.4. So sánh với các quan hệ toán khác

- Quan hệ hàm số : $x \leftrightarrow y$
- Quan hệ tương quan ρ_{xy}
- Quan hệ nhân quả $X \rightarrow Y \rightarrow X$

cuu duong than cong . com

2. Mô hình hồi qui Tổng thể

- Tổng thể : toàn bộ những cá thể mang dấu hiệu nghiên cứu
- Phân tích hồi qui dựa trên toàn bộ tổng thể

Giả sử biến phụ thuộc Y chỉ phụ thuộc một biến giải thích X

2.1. Hàm hồi qui tổng thể (PRF: *Population Regression Function*).

Xét quan hệ hồi qui:

$X = X_i \rightarrow (Y/X_i)$	Biến ngẫu nhiên Y trong điều kiện $X = X_i$ ($i = 1 \div n$)
$\rightarrow \exists F(Y/X_i)$	Tồn tại Phân phối xác suất có điều kiện
$\rightarrow \exists E(Y/X_i)$	Tồn tại duy nhất giá trị Kỳ vọng có điều kiện
$X_i \rightarrow E(Y/X_i)$	Quan hệ hàm số
$E(Y/X_i) = f(X_i)$	Hàm hồi qui tổng thể (PRF)

Giả sử hàm hồi qui tổng thể có dạng tuyến tính

$$E(Y/X_i) = \beta_1 + \beta_2 X_i$$

β_1 và β_2 đ-ợc gọi là các hệ số hồi quy (regression coefficient)

Trong đó: $\beta_1 = E(Y/X_i = 0)$: hệ số chặn (INPT : intercept term)

$$\beta_2 = \frac{\sum E(Y / X_i)}{\sum X_i} : \text{hệ số góc (slope coefficient)}$$

→ Hàm hồi qui tổng thể cho biết mối quan hệ giữa biến phụ thuộc và biến giải thích về mặt trung bình trong tổng thể.

2.2. Phân loại

Hàm hồi qui tổng thể được gọi là tuyến tính nếu nó tuyến tính với tham số.

2.3. Sai số ngẫu nhiên.

- Xét giá trị cụ thể $Y_i \in (Y/X_i)$, thông thường $Y_i \neq E(Y/X_i)$
- Đặt $u_i = Y_i - E(Y/X_i)$: là sai số ngẫu nhiên (nhiều, yếu tố ngẫu nhiên: random errors)
- Tính chất của SSNN :
 - + Nhận những giá trị dương và âm.
 - + Kỳ vọng bằng 0: $E(u_i) = 0 \quad \forall i$

Bản chất của SSNN : đại diện cho tất cả những yếu tố không phải biến giải thích nhưng cũng tác động tới biến phụ thuộc:

- + Những yếu tố không biết.
- + Những yếu tố không có số liệu.
- + Những yếu tố không ảnh hưởng nhiều đến biến phụ thuộc.
- + Sai số của số liệu thống kê.
- + Sai lệch do chọn dạng hàm số.
- + Những yếu tố mà tác động của nó không mang tính hệ thống.

2.4. Mô hình hồi quy tổng thể

PRM: Population regression model

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (i = 1, N)$$

3. Mô hình hồi qui mẫu

- Không biết toàn bộ Tổng thể, nên dạng của PRF có thể biết nhưng giá trị β_j thì không biết.
- Mẫu : một bộ phận mang thông tin của tổng thể.
- $W = \{(X_i, Y_i), i = 1 \div n\}$ được gọi là một mẫu kích thước n , có n quan sát (*observation*).

3.1. Hàm hồi qui mẫu (SRF : *Sample Regression Function*)

- Trong mẫu W , tồn tại một hàm số mô tả xu thế biến động của biến phụ thuộc theo biến giải thích về mặt trung bình, $\hat{y} = \hat{f}(x)$ gọi là hàm hồi qui mẫu (SRF).

- Hàm hồi qui mẫu có dạng giống hàm hồi qui tổng thể

Nếu PRF có dạng $E(Y/X_i) = \beta_1 + \beta_2 X_i$

Thì SRF có dạng $\hat{y}_i = \hat{b}_1 + \hat{b}_2 X_i$

$\hat{b}_1$ và $\hat{b}_2$ gọi là các *hệ số hồi quy - ước l- ợng*

(*Estimated regression coefficients*)

$\hat{y}_i$ gọi là các *giá trị - ước l- ợng hay giá trị t- ơng hợp*

(*Fitted value*)

- Vì có vô số mẫu ngẫu nhiên, nên có vô số giá trị của $\hat{b}_1$ và

$\hat{b}_2 \rightarrow \hat{b}_j$ là biến ngẫu nhiên.

- Với một mẫu cụ thể w kích thước n , $\hat{b}_j$ sẽ là con số cụ thể.

3.2. Phần dư

- Thông thường $Y_i \neq \hat{y}_i$, đặt $e_i = Y_i - \hat{y}_i$ và gọi là phần dư (*residual*).
 - Bản chất của phần dư e_i giống sai số ngẫu nhiên u_i
- $\hat{y}_i, \hat{\beta}_1, \hat{\beta}_2, e_i$ là ước lượng điểm tương ứng của $E(Y/X_i), \beta_1, \beta_2, u_i$.

cuu duong than cong . com

3.3. Mô hình hồi quy mẫu

SRM: *Sample regression model*

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$$

Với mỗi mẫu cụ thể sẽ tìm được một SRF tương ứng nên phải tìm một ước lượng tốt nhất.

cuu duong than cong . com

Bài 2. ƯỚC LƯỢNG VÀ KIỂM ĐỊNH MÔ HÌNH HỒI QUI ĐƠN

1. Mô hình

- Mô hình hồi qui đơn (*Simple regression*) là mô hình một phương trình gồm một biến phụ thuộc (Y) và một biến giải thích (X).

- Mô hình có dạng: **PRF** $E(Y/X_i) = \beta_1 + \beta_2 X_i$
PRM $Y_i = \beta_1 + \beta_2 X_i + u_i$

- Với mẫu $W = \{(X_i, Y_i), i = 1 \div n\}$, tìm β_1, β_2 sao cho

$$\text{SRF} \quad \hat{Y}_i = \beta_1 + \beta_2 X_i$$

$$\text{SRM} \quad Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$$

phản ánh xu thế biến động về mặt trung bình của mẫu.

2. Phương pháp bình phương nhỏ nhất(*Ordinary least squares -OLS*)

2.1. Phương pháp

- Tìm β_1, β_2 sao cho $Q = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n e_i^2 \rightarrow \min$

Lấy đạo hàm riêng của Q theo $\hat{\beta}_1$ và $\hat{\beta}_2$ và cho bằng 0:

$$\partial Q / \partial \hat{\beta}_1 = -2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = 0$$

$$\partial Q / \partial \hat{\beta}_2 = -2 \sum X_i (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = 0$$

$$\Rightarrow \begin{aligned} \hat{\beta}_1 n + \hat{\beta}_2 \sum X_i &= \sum Y_i \\ \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2 &= \sum X_i Y_i \end{aligned}$$

Đặt: $\bar{x} = (\sum X_i)/n$; $\bar{y} = (\sum Y_i)/n$; $\bar{xy} = (\sum X_i Y_i)/n$; $\overline{x^2} = (\sum X_i^2)/n$

$$\Rightarrow \beta_2 = \frac{\overline{XY} - \bar{x}\bar{y}}{\overline{X^2} - (\bar{x})^2}; \quad \beta_1 = \bar{y} - \hat{b}_2 \bar{x}$$

Đặt $x_i = X_i - \bar{x}$; $y_i = Y_i - \bar{y}$; $\hat{y}_i = \hat{y}_i - \bar{y}$

$$\rightarrow \hat{b}_2 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

$$\rightarrow \hat{y}_i = \hat{b}_2 x_i$$

gọi là hàm hồi quy mẫu đi qua gốc toạ độ.

β_1 , β_2 ước lượng bằng phương pháp bình phương nhỏ nhất nên đ-ợc gọi là các ước lượng bình phương nhỏ nhất (OLS) của β_1 và β_2 .

2.2. Phương pháp OLS có các tính chất sau:

- SRF đi qua điểm trung bình mẫu $(\bar{x}, \bar{y})$
- Trung bình của các giá trị - ớc l- ợng bằng trung bình mẫu

$$\bar{\hat{y}} = \bar{y}$$

- Tổng các phần d- bằng không $\sum_{i=1}^n e_i = 0$

d. Các phần dư không tương quan với các giá trị của biến giải

$$\text{thích: } \text{cov}(e_i, X_i) = \sum_{i=1}^n e_i X_i = 0$$

e. Các phần dư không tương quan với các giá trị ước lượng của

$$\text{biến phụ thuộc } Y : \text{cov}(e_i, \hat{Y}_i) = \sum_{i=1}^n e_i \hat{Y}_i = 0$$

3. Các giả thiết cơ bản của OLS

Một ước lượng sẽ dùng được khi nó là tốt nhất. Để ước lượng OLS là tốt nhất thì tổng thể phải thỏa mãn một số giả thiết sau:

Giả thiết 1: Mô hình hồi quy có dạng tuyến tính đối với tham số.

Giả thiết 2: Biến giải thích là phi ngẫu nhiên

Giả thiết 3: Trung bình của các sai số ngẫu nhiên bằng 0

$$E(u_i) = 0 \quad \forall i$$

Giả thiết 4: Phương sai sai số ngẫu nhiên bằng nhau

$$\text{Var}(u_i) = \sigma^2 \quad \forall i$$

Giả thiết 5: Các sai số ngẫu nhiên không tương quan

$$\text{Cov}(u_i, u_j) = 0 \quad \forall i \neq j$$

Giả thiết 6: SSNN và biến giải thích không tương quan

$$\text{Cov}(u_i, X_i) = 0 \quad \forall i$$

Giả thiết 7: Các giá trị của biến giải thích phải khác nhau càng nhiều càng tốt

$$\text{Var}(X) > 0$$

Giả thiết 8: Kích thước mẫu phải lớn hơn số tham số cần ước lượng của mô hình.

$$n > k$$

Giả thiết 9: Mô hình được chỉ định đúng.

Giả thiết 10: Không có đa cộng tuyến giữa các biến giải thích của mô hình hồi quy bội.

Định lý Gaus-Markov: Nếu tổng thể thỏa mãn các giả thiết trên thì ước lượng OLS sẽ là ước lượng tuyến tính, không chệch, tốt nhất (trong số các ước lượng không chệch) của các tham số (Best Linear Unbiased Estimator - BLUE).

4. Các tham số của ước lượng OLS

Các ước lượng $\hat{\beta}_j$ là biến ngẫu nhiên tùy thuộc mẫu, nên có các tham số đặc trưng

$$\text{Kì vọng : } E(\beta_1) = \beta_1 \qquad E(\beta_2) = \beta_2$$

$$\text{Phương sai : } \text{Var}(\beta_1) = \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n x_i^2} s^2 \qquad \text{Var}(\beta_2) = \frac{1}{\sum_{i=1}^n x_i^2} s^2$$

$$\text{Độ lệch chuẩn : } SD(\hat{\beta}_j) = \sqrt{\text{Var}(\hat{\beta}_j)} \qquad (j = 1, 2)$$

Th- ờng thì σ^2 là phương sai của sai số ngẫu nhiên chưa biết, được ước lượng bởi $\hat{s}^2$

$$\hat{s}^2 = \frac{\sum_{i=1}^n e_i^2}{n - 2} \quad \text{với } 2 \text{ là số tham số cần phải ước lượng của mô hình.}$$

$$\hat{s} = \sqrt{\hat{s}^2} \quad \text{là độ lệch chuẩn của đường hồi quy :}$$

(Standard error of Regression)

Lúc đó ta thu đ- ợc:

$$Se(\hat{\beta}_1) = \frac{\hat{s} \sqrt{\sum_{i=1}^n x_i^2}}{\sqrt{n \sum_{i=1}^n x_i^2}}$$

$$Se(\hat{b}_2) = \frac{s}{\sqrt{\sum_{i=1}^n \hat{a}_{x_i}^2}}$$

Các sai số chuẩn phản ánh độ chính xác của ước lượng

$$\text{Cov}(\hat{b}_1, \hat{b}_2) = -\bar{x} \text{Var}(\hat{b}_2)$$

Hiệp phương sai phản ánh mối quan hệ giữa $\hat{b}_1$ và $\hat{b}_2$

Các tham số trên thường được cho trong ma trận sau:

$$\text{var-cov} = \begin{pmatrix} \text{var}(b_1) & \dots & \text{cov}(b_1, b_2) \\ \text{cov}(b_2, b_1) & \dots & \text{var}(b_2) \end{pmatrix}$$

5. Sự phù hợp của hàm hồi qui- Hệ số xác định R^2

$$Y_i = \hat{Y}_i + e_i \Rightarrow \begin{matrix} y_i = Y_i - \bar{Y} \\ \hat{y}_i = \hat{Y}_i - \bar{Y} \end{matrix} \Rightarrow y_i = \hat{y}_i + e_i$$

$$y_i^2 = \hat{y}_i^2 + e_i^2 + 2e_i\hat{y}_i$$

$$\sum_{i=1}^n y_i^2 = \sum_{i=1}^n \hat{y}_i^2 + \sum_{i=1}^n e_i^2 + 2 \sum_{i=1}^n e_i \hat{y}_i$$

$$\text{Do } \sum_{i=1}^n e_i \hat{y}_i = 0$$

$$\text{Nên } \sum_{i=1}^n y_i^2 = \sum_{i=1}^n \hat{y}_i^2 + \sum_{i=1}^n e_i^2$$

$$\text{Ký hiệu: } \sum_{i=1}^n y_i^2 = TSS \quad \sum_{i=1}^n \hat{y}_i^2 = ESS \quad \text{và} \quad \sum_{i=1}^n e_i^2 = RSS$$

Thì thu được hệ thức cơ bản của phương pháp phân tích phương sai (Analysis of Variance) sau đây:

$$TSS = ESS + RSS$$

TSS (Total Sum of Squares) : đo tổng biến động của biến phụ thuộc

ESS (Explained Sum of Squares): tổng biến động của biến phụ thuộc được giải thích bởi MH – biến giải thích.

RSS (Residual Sum of Squares) : tổng biến động của biến phụ thuộc được giải thích bởi các yếu tố nằm ngoài mô hình – Sai số ngẫu nhiên.

Đặt $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$ gọi là hệ số xác định, $0 \leq R^2 \leq 1$

Ý nghĩa: Hệ số xác định R^2 là tỉ lệ (hoặc tỉ lệ %) sự biến động của biến phụ thuộc được giải thích bởi biến giải thích (theo mô hình, trong mẫu).

6. Hệ số t-ơng quan R :

Là căn bậc hai của hệ số xác định và đo mức độ t-ơng quan tuyến tính giữa Y và X

Hệ số t-ơng quan th-ờng đ-ợc cho trong ma trận sau:

$$r = \begin{pmatrix} 1 & \dots & r & 0 \\ \vdots & & \vdots & \vdots \\ r & \dots & 1 & 0 \\ 0 & & 0 & 1 \end{pmatrix}$$

7. Phân phối xác suất của sai số ngẫu nhiên.

Muốn tiến hành các suy diễn thống kê, thì phải biết phân phối xác suất của các ước lượng, phân phối đó tùy thuộc phân phối xác suất của SSNN.

Giả thiết 11: Các SSNN u_i có phân phối chuẩn.

Cơ sở của giả thiết này là:

+ Do u_i thường là sự tổng hợp của một số lớn các nhân tố ngẫu nhiên độc lập và ảnh hưởng đều nhau nên theo hệ quả của định lý giới hạn trung tâm thì có thể xem là u_i phân phối chuẩn.

+ Phân phối chuẩn chỉ có hai tham số là μ và σ^2 nên dễ sử dụng.

+ Phân phối chuẩn có tính chất là nếu u_i phân phối chuẩn thì mọi hàm tuyến tính của nó cũng phân phối chuẩn.

+ Phân phối chuẩn có tính chất là tính độc lập và không tương quan là đồng nhất.

Kết hợp các giả thiết 3,4,5 và 11 ta có giả thiết chung là:

$$u_i \sim n.i.d(0, \sigma^2)$$

Mô hình thỏa mãn các giả thiết trên gọi là *mô hình hồi quy tuyến tính cổ điển* (*Classic Linear Regression Model - CLRM*).

8. Các tính chất của các ước lượng OLS.

a. Các ước lượng của CLRM là các ước lượng không chệch.

b. Các ước lượng của CLRM là các ước lượng vững

c. Các ước lượng của CLRM là các ước lượng hiệu quả nhất.

$$d. \hat{\beta}_1 \sim N(\beta_1, \text{var}(\hat{\beta}_1))$$

$$\Rightarrow U = \frac{\hat{\beta}_1 - \beta_1}{\text{sd}(\hat{\beta}_1)} \sim N(0,1)$$

$$\Rightarrow T = \frac{\hat{\beta}_1 - \beta_1}{\text{se}(\hat{\beta}_1)} \sim T(n-2)$$

$$e. \hat{\beta}_2 \sim N(\beta_2, \text{var}(\hat{\beta}_2))$$

$$\Rightarrow U = \frac{\hat{\beta}_2 - \beta_2}{\text{sd}(\hat{\beta}_2)} \sim N(0,1)$$

$$\Rightarrow T = \frac{\hat{\beta}_2 - \beta_2}{\text{se}(\hat{\beta}_2)} \sim T(n-2)$$

$$f. \chi^2 = \frac{(n-2)\hat{s}^2}{s^2} \sim \chi^2(n-2)$$

g. Các ước lượng của *CLRM* đều là *BLUE* hoặc *BUE*

$$h. Y_i \sim N(\beta_1 + \beta_2 X_i, \sigma^2) \quad i = 1, 2, \dots, N.$$

9. Suy diễn thống kê.

9.1. Ước lượng khoảng

Với độ tin cậy $1 - \alpha$ cho trước:

Khoảng tin cậy của các hệ số hồi quy β_1 và β_2

Khoảng tin cậy tổng quát:

$$\beta_j - Se(\beta_j)t_{\alpha/2}(n-2) < \beta_j < \beta_j + Se(\beta_j)t_{\alpha/2}(n-2)$$

Khoảng tin cậy đối xứng:

$$\beta_j - Se(\beta_j)t_{\alpha/2}(n-2) < \beta_j < \beta_j + Se(\beta_j)t_{\alpha/2}(n-2)$$

Khoảng tin cậy bên phải:

$$\beta_j - Se(\beta_j)t_{\alpha}(n-2) < \beta_j$$

Khoảng tin cậy bên trái:

$$\beta_j < \beta_j + Se(\beta_j)t_{\alpha}(n-2) \quad (j = 1, 2)$$

Khoảng tin cậy cho sai số ngẫu nhiên:

Khoảng tin cậy tổng quát:

$$\frac{\hat{s}^2(n-2)}{c_{\alpha/2}^2(n-2)} < \sigma^2 < \frac{\hat{s}^2(n-2)}{c_{1-\alpha/2}^2(n-2)}$$

Khoảng tin cậy hai phía:

$$\frac{\hat{s}^2(n-2)}{c_{\alpha/2}^2(n-2)} < \sigma^2 < \frac{\hat{s}^2(n-2)}{c_{1-\alpha/2}^2(n-2)}$$

Khoảng tin cậy bên phải:

$$\frac{\hat{s}^2(n-2)}{c_{\alpha}^{2(n-2)}} < \sigma^2$$

Khoảng tin cậy bên trái:

$$\sigma^2 < \frac{\hat{s}^2(n-2)}{c_{1-\alpha}^{2(n-2)}}$$

9.2. Kiểm định giả thuyết

Với mức ý nghĩa α cho trước, kiểm định mối quan hệ thứ tự của hệ số với các số thực cho trước

i. Cặp giả thuyết
$$\begin{cases} H_0: b_j = b_j^* \\ H_1: b_j \neq b_j^* \end{cases} \quad j = 1, 2$$

$$\text{Tiêu chuẩn kiểm định: } T_{qs} = \frac{\hat{b}_j - b_j^*}{se(\hat{b}_j)}$$

Nếu $|T_{qs}| > t_{\alpha/2}(n-2)$ thì bác bỏ H_0 , ngược lại: chưa có cơ sở bác bỏ H_0 .

ii. Cặp giả thuyết
$$\begin{cases} H_0: b_j = b_j^* \\ H_1: b_j > b_j^* \end{cases} \quad \text{Nếu } T_{qs} > t_{\alpha}(n-2) : \text{ bác bỏ } H_0$$

iii. Cặp giả thuyết
$$\begin{cases} H_0: b_j = b_j^* \\ H_1: b_j < b_j^* \end{cases} \quad \text{Nếu } T_{qs} < -t_{\alpha}(n-2) : \text{ bác bỏ } H_0$$

Trường hợp đặc biệt
$$\begin{cases} H_0: \beta_j = 0 \\ H_1: \beta_j \neq 0 \end{cases} \rightarrow T_{qs} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

Các kiểm định trên đ-ợc gọi là Kiểm định T.

Kiểm định bằng P-value:

Với kiểm định bên phải: $P\text{-value} = P(T > T_{qs})$

Với kiểm định bên trái: $P\text{-value} = P(T < T_{qs})$

Với kiểm định hai phía: $P\text{-value} = 2P(T > |T_{qs}|)$

Nếu cho tr-ớc α thì quy tắc kết luận nh- sau:

Nếu $P\text{-value} < \alpha$ thì bác bỏ H_0

Nếu $P\text{-value} > \alpha$ thì thừa nhận H_0

iii. Cặp giả thuyết
$$\begin{cases} H_0: s^2 = s_0^2 \\ H_1: s^2 \neq s_0^2 \end{cases}$$

Tiêu chuẩn kiểm định:

$$\chi^2 = \frac{(n-2)s^2}{s_0^2}$$

Nếu $\chi_{qs} < c_{1-\alpha/2}^2$ hoặc $\chi_{qs} > c_{\alpha/2}^2$ thì bác bỏ H_0

Cặp giả thuyết
$$\begin{cases} H_0: s^2 = s_0^2 \\ H_1: s^2 > s_0^2 \end{cases}$$

Nếu $\chi_{qs} > c_{\alpha}^2$ thì bác bỏ H_0

Cặp giả thuyết
$$\begin{cases} H_0: s^2 = s_0^2 \\ H_1: s^2 < s_0^2 \end{cases}$$

Nếu $\chi_{qs} < c_{1-\alpha}^2$ thì bác bỏ H_0

Các kiểm định trên đ- ợc gọi là kiểm định Khi bình ph- ơng (χ^2)

Các kiểm định trên cũng có thể tiến hành bằng ph- ơng pháp P-value.

10. Kiểm định về sự thích hợp của mô hình.

Cặp giả thuyết

$$\begin{array}{l} H_0: R^2 = 0 \\ H_1: R^2 > 0 \end{array} \quad \begin{array}{l} \text{Biến giải thích không giải thích cho } Y \\ \text{Biến giải thích có giải thích cho } Y \end{array} \Leftrightarrow \begin{array}{l} H_0: b_2 = 0 \\ H_1: b_2 \neq 0 \end{array}$$

Kiểm định F:
$$F_{qs} = \frac{ESS / 1}{RSS / (n - 2)} = \frac{R^2 / 1}{(1 - R^2) / (n - 2)}$$

- Nếu $F_{qs} > F_{\alpha}(1; n - 2)$ thì bác bỏ H_0 : biến giải thích giải thích được cho sự biến động của biến phụ thuộc, hàm hồi qui được gọi là phù hợp.
- Ngược lại, Y không phụ thuộc vào biến giải thích, hàm hồi qui không phù hợp.

Vì hai cặp giả thiết tương đương, kiểm định F tương đương kiểm định T

$$F_{qs} = (T_{qs})^2.$$

Kiểm định F nói trên cũng có thể tiến hành bằng phương pháp P-value

11. Dự báo

Là ước lượng khoảng cho giá trị trung bình và cá biệt của biến phụ thuộc khi biến giải thích nhận giá trị xác định $X = X_0$

11.1. Dự báo giá trị trung bình

Khoảng tin cậy tổng quát:

$$\hat{y}_0 - Se(\hat{y}_0)t_{\alpha/2}(n-2) < E(Y/X_0) < \hat{y}_0 + Se(\hat{y}_0)t_{\alpha/2}(n-2)$$

Khoảng tin cậy đối xứng:

$$\hat{y}_0 - Se(\hat{y}_0)t_{\alpha/2}(n-2) < E(Y/X_0) < \hat{y}_0 + Se(\hat{y}_0)t_{\alpha/2}(n-2)$$

Khoảng tin cậy bên phải:

$$\hat{y}_0 - Se(\hat{y}_0)t_{\alpha}(n-2) < E(Y/X_0)$$

Khoảng tin cậy bên trái:

$$E(Y/X_0) < \hat{Y}_0 + Se(\hat{Y}_0)t_{\alpha}(n-2)$$

$$\text{Với } \hat{Y}_0 = \beta_1 + \beta_2 X_0 \quad \text{và} \quad Se(\hat{Y}_0) = \hat{s} \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{S X_i^2}}$$

11.2. Dự báo giá trị cá biệt

$$\hat{Y}_0 - Se(\hat{Y}_0 - Y_0)t_{\alpha_2}(n-2) < Y_0 < \hat{Y}_0 + Se(\hat{Y}_0 - Y_0)t_{\alpha_1}(n-2)$$

$$\text{Với } Se(\hat{Y}_0 - Y_0) = \hat{s} \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{S X_i^2}}$$

Ví dụ 1: Hồi quy hàm tiêu dùng Keynes và cho nhận xét.

Dependent Variable: Y
Method: Least Squares
Date: 02/16/09 Time: 09:07
Sample: 1980 1991
Included observations: 12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-231.7951	94.52751	-2.452144	0.0341
X	0.719433	0.021750	33.07810	0.0000
R-squared	0.990943	Mean dependent var		2880.600
Adjusted R-squared	0.990038	S.D. dependent var		314.4417
S.E. of regression	31.38488	Akaike info criterion		9.881541
Sum squared resid	9850.106	Schwarz criterion		9.962359
Log likelihood	-57.28925	F-statistic		1094.160
Durbin-Watson stat	1.284183	Prob(F-statistic)		0.000000

Ví dụ 2: Với các số liệu về lãi suất cổ phiếu của công ty IBM và của thị trường chứng khoán Mỹ từ tháng 1 năm 1978 đến tháng 12 năm 1987

(tập số liệu ch2bt1) hãy ước lượng mô hình SIM và cho nhận xét

Mô hình SIM(Single Index Model) có dạng:

$$R_i = \alpha + \beta R_m$$

Trong đó: R_i là lợi tức của công ty i

R_m là lợi tức của chỉ số thị trường.

α thể hiện tác động của các yếu tố khác ngoài R_m đối với R_i

β đo mức độ nhạy cảm của chứng khoán i trước những dao động của thị trường.

Nếu chứng khoán ít nhạy cảm trước những biến động của chỉ số thị trường thì $0 < \beta < 1$ và được gọi là chứng khoán tự vệ. Nếu chứng khoán nhạy cảm trước những biến động của chỉ số thị trường thì $\beta > 1$ và gọi là chứng khoán năng động. Mô hình SIM giả định rằng lợi tức của mỗi chứng khoán đều có ba bộ phận hợp thành:

α đại diện cho phần lợi tức không phụ thuộc vào lãi suất thị trường

β đo lường mức độ nhạy cảm của lợi tức chứng khoán đang xét trước những thay đổi của lợi tức của chỉ số chứng khoán.

Như vậy mô hình SIM chỉ ra hai loại rủi ro khác nhau trong đầu tư chứng khoán:

- Rủi ro thị trường, được đo bằng β , là rủi ro liên quan đến sự biến động của toàn bộ thị trường và không thể giảm thiểu bằng cách đa dạng hóa trong phạm vi thị trường.
- Rủi ro riêng chỉ liên quan đến cổ phiếu đang xét. Nó có thể đo bằng sai số chuẩn của β . Nhà đầu tư chỉ có thể loại trừ rủi ro riêng bằng cách đa dạng hóa danh mục đầu tư.

Ta chuyển sang mô hình kinh tế lượng:

$$R_i = \alpha + \beta R_m + U_i$$

Kết quả hồi quy như sau:

Dependent Variable: IBM
Method: Least Squares
Date: 02/16/09 Time: 09:54

Sample: 1978:01 1987:12
Included observations: 120

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003278	0.004703	0.697054	0.4871
MARKET	0.453024	0.067675	6.694125	0.0000
R-squared	0.275235	Mean dependent var	0.009617	
Adjusted R-squared	0.269093	S.D. dependent var	0.059024	
S.E. of regression	0.050461	Akaike info criterion	-3.118687	
Sum squared resid	0.300471	Schwarz criterion	-3.072229	
Log likelihood	189.1212	F-statistic	44.81131	
Durbin-Watson stat	1.882724	Prob(F-statistic)	0.000000	

Ví dụ 3: Hãy thu thập số liệu của Việt nam để hồi quy các mô hình sau:

$$(a) \text{ FDI}_t = \beta_1 + \beta_2 \text{GDP}_t + U_t$$

$$(b) \ln \text{FDI}_t = \beta_1 + \beta_2 \ln \text{GDP}_t + U_t$$

Và cho biết mô hình nào phù hợp hơn.

Bài 3. MÔ HÌNH HỒI QUI BỘI (*Multiple regression*)

1. Mô hình hồi qui 3 biến.

1.1. Mô hình:

Mô hình hồi qui trong đó biến phụ thuộc Y phụ thuộc vào 2 biến giải thích X_2, X_3 có dạng

$$\text{PRF: } E(Y / X_{2i}, X_{3i}) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} \quad (1)$$

Đồ thị là một mặt phẳng

$$\text{PRM: } Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

trong đó: β_1 gọi là hệ số chặn (*intercept*)

β_j ($j = 2, 3$) gọi là hệ số góc riêng phần (*partial slope*)

Giả sử mọi giả thiết của OLS đều thoả mãn, lúc đó với mẫu kích thước n đã được lập từ tổng thể sẽ xác định được:

$$\text{SRF: } \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} \quad (2)$$

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

$$\text{Tìm } \hat{\beta}_j \ (j = \overline{1, 3}) \text{ sao cho } Q = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n e_i^2 \rightarrow \min$$

$$\Rightarrow \partial Q / \partial \hat{\beta}_1 = 0$$

$$\partial Q / \partial \hat{\beta}_2 = 0$$

$$\partial Q / \partial \hat{\beta}_3 = 0$$

$$\begin{aligned} \Rightarrow \hat{\beta}_1 n + \hat{\beta}_2 \sum X_{2i} + \hat{\beta}_3 \sum X_{3i} &= \sum Y_i \\ \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i} X_{3i} &= \sum X_{2i} Y_i \\ \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{2i} X_{3i} + \hat{\beta}_3 \sum X_{3i}^2 &= \sum X_{3i} Y_i \end{aligned}$$

$$\begin{aligned} \text{Ký hiệu: } \bar{Y} &= (\sum Y_i)/n & \bar{X}_2 &= (\sum X_{2i})/n & \bar{X}_3 &= (\sum X_{3i})/n \\ y_i &= Y_i - \bar{Y} & x_{2i} &= X_{2i} - \bar{X}_2 & x_{3i} &= X_{3i} - \bar{X}_3 \end{aligned}$$

$$\begin{aligned} \Rightarrow \hat{\beta}_1 &= \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3 \\ \hat{\beta}_2 &= \frac{\sum x_{2i} y_i \sum x_{3i}^2 - \sum x_{3i} y_i \sum x_{2i} x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i} x_{3i})^2} \\ \hat{\beta}_3 &= \frac{\sum x_{3i} y_i \sum x_{2i}^2 - \sum x_{2i} y_i \sum x_{2i} x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i} x_{3i})^2} \end{aligned}$$

$$\Rightarrow \hat{y}_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} \rightarrow \text{Hàm hồi quy mẫu đi qua gốc tọa độ.}$$

1.2. Các tham số của các ước lượng OLS.

$$E(\hat{\beta}_j) = \beta_j \quad j = \overline{1, 3}$$

$$\text{Var}(\hat{b}_1) = \frac{\frac{1}{n} + \frac{\bar{x}_2^2 \sum x_{3i}^2 + \bar{x}_3^2 \sum x_{2i}^2 - 2\bar{x}_2\bar{x}_3 \sum x_{2i}x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i}x_{3i})^2}}{\sigma^2}$$

$$\text{Var}(\hat{b}_2) = \frac{\sum x_{3i}^2}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i}x_{3i})^2} \sigma^2 = \frac{s^2}{\sum x_{2i}^2 (1 - r_{23}^2)}$$

$$\text{Var}(\hat{b}_3) = \frac{\sum x_{2i}^2}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i}x_{3i})^2} \sigma^2 = \frac{s^2}{\sum x_{3i}^2 (1 - r_{23}^2)}$$

$$\text{Se}(\hat{b}_j) = \sqrt{\text{Var}(\hat{b}_j)} \quad \text{trong đó} \quad s^2 \gg \hat{s}^2 = \frac{RSS}{n-3}$$

$$\text{Cov}(\hat{b}_2, \hat{b}_3) = \frac{-r_{23}s^2}{(1-r_{23}^2)\sqrt{\sum x_{2i}^2}\sqrt{\sum x_{3i}^2}}$$

1.3. Hệ số xác định bội R^2

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

Với mô hình ba biến:

$$R^2 = \frac{\hat{b}_2 \sum x_{2i}y_i + \hat{b}_3 \sum x_{3i}y_i}{\sum y_i^2}$$

1.4. Hệ số t-ơng quan.

a. Hệ số tương quan bội R : Là căn bậc hai của hệ số xác định bội và đo mức độ t-ơng quan tuyến tính chung giữa Y , X_2 và X_3 .

b. Hệ số t-ơng quan cặp r_{ij} : Đo mức độ t-ơng quan tuyến tính giữa biến i và biến j của mô hình.

$$r_{12}^2 = \frac{(\sum x_{2i} y_i)^2}{\sum x_{2i}^2 \sum y_i^2}$$

$$r_{13}^2 = \frac{(\sum x_{3i} y_i)^2}{\sum x_{3i}^2 \sum y_i^2}$$

$$r_{23}^2 = \frac{(\sum x_{2i} x_{3i})^2}{\sum x_{2i}^2 \sum x_{3i}^2}$$

c. Hệ số t-ơng quan riêng phần $r_{ij,k}$: Đo mức độ t-ơng quan tuyến tính giữa biến i và biến j của mô hình với điều kiện biến k không đổi.

$$r_{12,3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

$$r_{13,2} = \frac{r_{13} - r_{12} r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}}$$

$$r_{23,1} = \frac{r_{23} - r_{12} r_{13}}{\sqrt{(1 - r_{12}^2)(1 - r_{13}^2)}}$$

Ví dụ: Bảng sau đây cho Tỷ lệ lạm phát Y(%), Tỷ lệ thất nghiệp X₂(%) và Tỷ lệ lạm phát kỳ vọng X₃(%) của Mỹ giai đoạn 1970- 1982:

Năm	Y	X ₂	X ₃
1970	5.92	4.9	4.78
1971	4.30	5.9	3.84
1972	3.30	5.6	3.13
1973	6.23	4.9	3.44
1974	10.97	5.6	6.84
1975	9.14	8.5	9.47
1976	5.77	7.7	6.51
1977	6.45	7.1	5.92
1978	7.60	6.1	6.08
1979	11.47	5.8	8.09
1980	13.46	7.1	10.01
1981	10.24	7.6	10.81
1982	5.99	9.7	8.00

- Hồi quy Y với X_2 và cho nhận xét.
- Hồi quy Y với X_2 và X_3 và so sánh với kết quả thu được ở phần a.
- Hãy phân tích kết quả thu được ở mô hình 3 biến.

Dependent Variable: Y
 Method: Least Squares
 Date: 02/17/09 Time: 09:40
 Sample: 1970 1982
 Included observations: 13

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.127172	4.285283	1.429817	0.1806
X2	0.244934	0.630456	0.388502	0.7051
R-squared	0.013536	Mean dependent var	7.756923	
Adjusted R-squared	-0.076143	S.D. dependent var	3.041892	
S.E. of regression	3.155577	Akaike info criterion	5.276858	
Sum squared resid	109.5343	Schwarz criterion	5.363773	
Log likelihood	-32.29958	F-statistic	0.150934	
Durbin-Watson stat	0.969568	Prob(F-statistic)	0.705058	

Dependent Variable: Y
 Method: Least Squares
 Date: 02/17/09 Time: 09:35
 Sample: 1970 1982
 Included observations: 13

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.193357	1.594789	4.510538	0.0011
X2	-1.392472	0.305018	-4.565214	0.0010
X3	1.470032	0.175786	8.362633	0.0000
R-squared	0.876590	Mean dependent var	7.756923	
Adjusted R-squared	0.851907	S.D. dependent var	3.041892	
S.E. of regression	1.170605	Akaike info criterion	3.352092	
Sum squared resid	13.70316	Schwarz criterion	3.482465	
Log likelihood	-18.78860	F-statistic	35.51521	
Durbin-Watson stat	2.225465	Prob(F-statistic)	0.000029	

2. Mô hình hồi quy tổng quát k biến - Dạng ma trận của mô hình

2.1. Mô hình

Mô hình hồi qui trong đó biến phụ thuộc Y phụ thuộc vào $k - 1$ biến giải thích $X_2, \dots, X_k$ có dạng

$$PRF: E(Y_i) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} \quad (1)$$

$$PRM: Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i \quad (2)$$

trong đó: β_1 gọi là hệ số chặn

β_j ($j=2, k$) gọi là các hệ số góc riêng phần

Với mẫu $W = \{(X_{2i}, X_{3i}, \dots, X_{ki}, Y_i); i = 1 \div n\}$,

$$SRF: \hat{Y}_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} \quad (3)$$

$$SRM: Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + e_i \quad (4)$$

2.2. Dạng ma trận

$$Y_1 = \beta_1 + \beta_2 X_{21} + \dots + \beta_k X_{k1} + u_1$$

$$Y_2 = \beta_1 + \beta_2 X_{22} + \dots + \beta_k X_{k2} + u_2$$

...

$$Y_{n-1} = \beta_1 + \beta_2 X_{2n-1} + \dots + \beta_k X_{kn-1} + u_{n-1}$$

$$Y_n = \beta_1 + \beta_2 X_{2n} + \dots + \beta_k X_{kn} + u_n$$

$\Leftrightarrow$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{n-1} \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{21} & \dots & X_{k1} \\ 1 & X_{22} & \dots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n-1} & \dots & X_{kn-1} \\ 1 & X_{2n} & \dots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \\ u_n \end{bmatrix}$$

$$\mathbf{Y}_{(n \times 1)} = \mathbf{X}_{(n \times k)} \times \boldsymbol{\beta}_{(k \times 1)} + \mathbf{U}_{(n \times 1)}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U} \rightarrow E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$$

Tương tự, đặt $\hat{\mathbf{Y}}_n = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_{n-1} \\ \hat{Y}_n \end{bmatrix}$; $\hat{\mathbf{b}}_n = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \vdots \\ \hat{b}_k \end{bmatrix}$; $\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_{n-1} \\ e_n \end{bmatrix}$, thì

$$\hat{\mathbf{Y}}_n = \mathbf{X}_n \hat{\mathbf{b}}_n$$

$$\mathbf{Y} = \mathbf{X}_n \hat{\mathbf{b}}_n + \mathbf{e}$$

2.3. Phương pháp bình phương nhỏ nhất

Tìm $\hat{\mathbf{b}}_n$ sao cho $\sum_{i=1}^n e_i^2 = \mathbf{e}'\mathbf{e} \rightarrow \min$

$$\Leftrightarrow (\mathbf{Y} - \mathbf{X}_n \hat{\mathbf{b}}_n)' (\mathbf{Y} - \mathbf{X}_n \hat{\mathbf{b}}_n) \rightarrow \min \Leftrightarrow \mathbf{X}'_n \mathbf{X}_n \hat{\mathbf{b}}_n = \mathbf{X}'_n \mathbf{Y}$$

Nếu tồn tại $(\mathbf{X}'_n \mathbf{X}_n)^{-1}$ thì $\hat{\mathbf{b}}_n = (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{Y}$

Khi đó $\hat{\mathbf{b}}_n = (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{Y}$ là ước lượng tuyến tính không chệch tốt nhất của $\boldsymbol{\beta}$

2.4. Các tham số của ước lượng

Kì vọng: $E(\hat{\mathbf{b}}_n) = \boldsymbol{\beta}$

Phương sai – hiệp phương sai

$$Cov(\hat{\beta}) = \begin{bmatrix} Var(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) & \dots & Cov(\hat{\beta}_1, \hat{\beta}_k) \\ Cov(\hat{\beta}_2, \hat{\beta}_1) & Var(\hat{\beta}_2) & \dots & Cov(\hat{\beta}_2, \hat{\beta}_k) \\ \dots & \dots & \dots & \dots \\ Cov(\hat{\beta}_k, \hat{\beta}_1) & Cov(\hat{\beta}_k, \hat{\beta}_2) & \dots & Var(\hat{\beta}_k) \end{bmatrix} = \sigma^2 (X'X)^{-1}$$

Với σ^2 được ước lượng bởi $s^2 = \frac{ESS}{n - k}$

2.5. Sự phù hợp của hàm hồi qui

Hệ số xác định bội.

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

Cho biết tỉ lệ sự biến động của biến phụ thuộc được giải thích bởi *tất cả* các biến giải thích có trong mô hình.

R^2 có các tính chất sau:

$$+ 0 \leq R^2 \leq 1$$

Tính chất này dùng để đánh giá mức độ thích hợp của hàm hồi quy.

+ Giá trị của R^2 đồng biến với số biến giải thích của mô hình.

Tuy nhiên không thể lấy điều đó để xem xét việc đưa thêm biến giải thích vào mô hình.

2.6. Hệ số xác định bội hiệu chỉnh.

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

$\bar{R}^2$ có các tính chất sau:

+ $\bar{R}^2$ có thể nhận giá trị âm.

+ Khi số biến giải thích của mô hình tăng lên thì $\bar{R}^2$ tăng chậm hơn R^2 .

$$\bar{R}^2 \leq R^2 \leq 1$$

Tính chất này được dùng làm căn cứ xem xét việc đưa thêm biến giải thích vào mô hình. Khi đưa thêm biến vào mô hình mà $\bar{R}^2$ còn tăng hoặc khi giá trị t của kiểm định về sự bằng không của hệ số hồi quy tương ứng với biến đưa thêm còn lớn hơn 1 thì việc đưa thêm biến còn hợp lý.

2.7. Hệ số tương quan.

a. Hệ số tương quan bội R .

b. Hệ số tương quan cặp r_{ij} ($i, j = \overline{1, k}$)

Các hệ số tương quan cặp thường được cho trong ma trận sau:

$$r_{ij} = \begin{matrix} & \begin{matrix} 1 & \dots & r_{12} & \dots & r_{13} & \dots & r_{1k} \end{matrix} & \\ \begin{matrix} r_{21} & 1 & \dots & r_{23} & \dots & r_{2k} \end{matrix} & & & & & \\ \vdots & & \ddots & & & \\ \begin{matrix} r_{k1} & \dots & r_{k2} & \dots & r_{k3} & \dots & 1 \end{matrix} & & & & & \end{matrix}$$

c. Hệ số tương quan riêng phần $r_{12,34 \dots k} \dots r_{k-1k,12 \dots k-2}$

→ Các hệ số tương quan cặp được gọi là hệ số tương quan riêng phần bậc 0.

3. Suy diễn thống kê.

3.1. Ước lượng khoảng

i. Khoảng tin cậy cho từng hệ số

$$\beta_j - Se(\beta_j)t_{\alpha/2}(n-k) < \beta_j < \beta_j + Se(\beta_j)t_{\alpha/2}(n-k) \quad (j = \overline{1, k})$$

→ Khoảng tin cậy đối xứng, bên phải, bên trái.

ii. Khoảng tin cậy cho hai hệ số

$$(\hat{b}_i \pm \hat{b}_j) - Se(\hat{b}_i \pm \hat{b}_j)t_{\alpha/2}(n-k) < \beta_i \pm \beta_j < (\hat{b}_i \pm \hat{b}_j) + Se(\hat{b}_i \pm \hat{b}_j)t_{\alpha/2}(n-k)$$

$$\text{Với } Se(\hat{b}_i \pm \hat{b}_j) = \sqrt{Var(\hat{b}_i \pm \hat{b}_j)} = \sqrt{Var(\hat{b}_i) \pm 2Cov(\hat{b}_i, \hat{b}_j) + Var(\hat{b}_j)}$$

iii. Khoảng tin cậy cho sai số ngẫu nhiên

$$\frac{\hat{s}^2(n-k)}{c_{\frac{\alpha}{2}}^{2(n-k)}} < \sigma^2 < \frac{\hat{s}^2(n-k)}{c_{1-\frac{\alpha}{2}}^{2(n-k)}}$$

→ Khoảng tin cậy hai phía, bên phải, bên trái.

3.2. Kiểm định giả thuyết:

a. Kiểm định T:

Cặp giả thuyết	Tiêu chuẩn kiểm định	Miền bác bỏ H_0
$\begin{cases} H_0: b_j = b_j^* \\ H_1: b_j \neq b_j^* \end{cases}$	$T_{qs} = \frac{\hat{b}_j - b_j^*}{Se(\hat{b}_j)}$	$ T_{qs} > t_{\frac{\alpha}{2}}^{(n-k)}$
$\begin{cases} H_0: b_j = b_j^* \\ H_1: b_j > b_j^* \end{cases}$		$T_{qs} > t_{\alpha}^{(n-k)}$
$\begin{cases} H_0: b_j = b_j^* \\ H_1: b_j < b_j^* \end{cases}$		$T_{qs} < -t_{\alpha}^{(n-k)}$
$\begin{cases} H_0: b_i \pm b_j = a \\ H_1: b_i \pm b_j \neq a \end{cases}$	$T_{qs} = \frac{\hat{b}_i \pm \hat{b}_j - a}{Se(\hat{b}_i \pm \hat{b}_j)}$	$ T_{qs} > t_{\frac{\alpha}{2}}^{(n-k)}$

b. Kiểm định χ^2 :

Đối với σ^2 việc kiểm định cũng tiến hành t-ơng tự với tiêu chuẩn kiểm định:

$$\chi^2 = \frac{(n - k) s^2}{s_0^2}$$

4. Kiểm định sự phù hợp của hàm hồi qui

$$\begin{matrix} H_0: R^2 = 0 \\ H_1: R^2 > 0 \end{matrix} \Leftrightarrow \begin{matrix} H_0: b_2 = \dots = b_k = 0 \\ H_1: b_j \neq 0 : (j = 2, \dots, k) \end{matrix}$$

Tất cả các biến giải thích không giải thích cho Y

Ít nhất một biến giải thích có giải thích cho Y

$$F_{qs} = \frac{ESS / (k - 1)}{RSS / (n - k)} = \frac{R^2 / (k - 1)}{1 - R^2 / (n - k)}$$

$F_{qs} > F_{\alpha}(k - 1; n - k)$ thì bác bỏ H_0 : hàm hồi qui là phù hợp.

Tất cả các kiểm định trên cũng đều có thể tiến hành bằng ph-ơng pháp *P-value*.

Ví dụ: Với tập số liệu đã cho, hãy tìm các -ớc l-ợng $\hat{\beta}$ bằng ph-ơng pháp ma trận và các tham số t-ơng ứng của mô hình. Hãy tiến hành các -ớc l-ợng và kiểm định cần thiết.

5. Kiểm định thu hẹp hồi qui (Kiểm định Wald):

5.1. Thủ tục:

Xét mô hình:

$$E(Y/X_2, \dots, X_{k-m}, \dots, X_k) = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (UR)$$

Nghi ngờ m biến giải thích $X_{k-m+1}, \dots, X_k$ không giải thích cho Y

$$\begin{matrix} H_0: b_{k-m+1} = b_{k-m+2} = \dots = b_k = 0 \\ H_1: b_j \neq 0 : (j = k - m + 1, \dots, k) \end{matrix}$$

Tất cả m biến giải thích không giải thích cho Y

Ít nhất một biến giải thích có giải thích cho Y

Nếu giả thuyết H_0 là đúng thì mô hình trở thành:

$$E(Y/X_2, \dots, X_{k-m}) = \beta_1 + \beta_2 X_2 + \dots + \beta_{k-m} X_{k-m} \quad (R)$$

Tiêu chuẩn kiểm định:

$$F_{qs} = \frac{(RSS_r - RSS_{ur}) / m}{RSS_{ur} / (n - k)} = \frac{(R_{ur}^2 - R_{\beta}^2) / m}{(1 - R_{ur}^2) / (n - k)}$$

Nếu $F_{qs} > F_{\alpha}(m, n - k)$ bác bỏ H_0

- Trường hợp $m = 1$: $F_{qs} = (T_{qs})^2$ với T_{qs} ứng với hệ số duy nhất cần kiểm định.
- Trường hợp $m = k - 1$: F_{qs} trong kiểm định thu hẹp chính là F_{qs} trong kiểm định sự phù hợp.
- Kiểm định thu hẹp hồi qui còn dùng cho những trường hợp khác.

Ví dụ: Với tệp số liệu ch3bt5 hãy ước lượng hàm cầu về thịt lợn phụ thuộc vào giá và thu nhập theo dạng tuyến tính và tuyến tính lôga và cho nhận xét. Có thể bỏ đi biến nào ra khỏi mô hình hay không?

5.2. Các dạng thu hẹp hồi qui

Ví dụ $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad (UR)$

$$H_0: \beta_3 = 1; H_1: \beta_3 \neq 1.$$

$$H_0 \text{ đúng} \Rightarrow Y_i = \beta_1 + \beta_2 X_{2i} + X_{3i} + u_i$$

$$\Leftrightarrow Y_i - X_{3i} = \beta_1 + \beta_2 X_{2i} + u_i \Leftrightarrow Y_i^* = \beta_1 + \beta_2 X_{2i} + u_i \quad (R)$$

$$H_0: \beta_2 = \beta_3; H_1: \beta_2 \neq \beta_3.$$

$$H_0 \text{ đúng} \Rightarrow Y_i = \beta_1 + \beta_2 X_{2i} + \beta_2 X_{3i} + u_i$$

$$\Leftrightarrow Y_i = \beta_1 + \beta_2(X_{2i} + X_{3i}) + u_i \Leftrightarrow Y_i = \beta_1 + \beta_2 X_i^* + u_i \quad (R)$$

$$H_0: \beta_2 + \beta_3 = a; H_1: \beta_2 + \beta_3 \neq a$$

$$H_0 \text{ đúng} \Rightarrow Y_i = \beta_1 + \beta_2 X_{2i} + (a - \beta_2) X_{3i} + u_i$$

$$\Leftrightarrow Y_i - aX_{3i} = \beta_1 + \beta_2(X_{2i} - X_{3i}) + u_i \Leftrightarrow Y_i^* = \beta_1 + \beta_2 X_i^* + u_i \quad (R)$$

6. Dự báo.

6.1. Dự báo giá trị trung bình

$$\hat{Y}_0 - Se(\hat{Y}_0) t_{\alpha/2}(n-k) < E(Y/X_0) < \hat{Y}_0 + Se(\hat{Y}_0) t_{\alpha/1}(n-k)$$

$$\text{Với } \hat{Y}_0 = \mathbf{X}_0' \hat{\beta} \quad \text{và} \quad Se(\hat{Y}_0) = \hat{s} \sqrt{\mathbf{x}_0' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0}$$

6.2. Dự báo giá trị cá biệt

$$\hat{Y}_0 - Se(Y_0) t_{\alpha/2}(n-k) < Y_0 < \hat{Y}_0 + Se(Y_0) t_{\alpha/1}(n-k)$$

$$\text{Với } Se(Y_0) = \hat{s} \sqrt{1 + \mathbf{x}_0' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0}$$

7. Một số mô hình Kinh tế

7.1. Hàm thu nhập – chi tiêu

Y_i : Thu nhập

C_i : Chi tiêu

$$C_i = \beta_1 + \beta_2 Y_i + u_i$$

- C là chi tiêu cho tiêu dùng : $\beta_1 > 0; 1 > \beta_2 > 0$
- C là chi tiêu cho hàng hóa thông thường
- C là chi tiêu cho hàng hóa cao cấp
- C là chi tiêu cho hàng hóa thứ cấp

7.2. Hàm cầu

Q_i : cầu về hàng hóa

P_i : giá cả hàng hóa

PT_i : giá hàng hóa thay thế

PB_i : giá hàng hóa bổ sung

$$Q_i = \beta_1 + \beta_2 P_i + \beta_3 PT_i + \beta_4 PB_i + u_i$$

7.3. Hàm tăng tr- ởng

r : tỷ lệ tăng tr- ởng

t : thời gian

Y_t và Y_0 là giá trị của biến Y tại thời kỳ t và thời kỳ gốc

$$Y_t = Y_0(1+r)^t$$

$$\Rightarrow \ln Y_t = \ln Y_0 + t \cdot \ln(1+r)$$

$$\Rightarrow Y'_t = \beta_1 + \beta_2 \cdot t$$

7.4. Hàm chi phí – sản lượng

Q_i : sản lượng

TC_i : tổng chi phí, MC_i : chi phí cận biên, AC_i : chi phí trung bình,

FC_i : chi phí cố định

$$TC_i = \beta_1 + \beta_2 Q_i + \beta_3 Q_i^2 + \beta_4 Q_i^3 + u_i$$

$$\rightarrow FC_i = \beta_1 + u_i$$

$$\rightarrow MC_i = \beta_2 + 2\beta_3 Q_i + 3\beta_4 Q_i^2 + u_i$$

$$\rightarrow AC_i = \frac{\beta_1}{Q_i} + \beta_2 + \beta_3 Q_i + \beta_4 Q_i^2 + u_i$$

7.5. Hàm hyperbol

$$Y = b_1 + b_2 \frac{1}{X}$$

Là hàm phi tuyến đối với X song tuyến tính đối với tham số

a. Nếu cả β_1 và β_2 đều dương, khi đó đồ thị cong xuống với mức tiệm cận dưới là β_1 . Hàm này thường được dùng để phân tích chi phí trung bình để sản xuất 1 sản phẩm.

b. Nếu $\beta_1 > 0$ và $\beta_2 < 0$ khi đó đồ thị cong lên với mức tiệm cận trên là β_1 . Hàm này dùng để phân tích sự phụ thuộc của chi tiêu vào thu nhập.

c. Nếu $\beta_1 < 0$ và $\beta_2 > 0$ thì đó là đường cong Philips, nó cong xuống và tiệm cận β_1 ở dưới trục hoành.

7.4. Hàm mũ – Hàm Loga tuyến tính

Mô hình kinh tế có dạng $Y_i = \beta_0 X_{2i}^{\beta_2} X_{3i}^{\beta_3}$

$$\Leftrightarrow \ln Y_i = \ln \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i}$$

Xét mô hình $\ln Y_i = \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + v_i$

$$\Leftrightarrow E(Y / X_{2i}, X_{3i}) = e^{\beta_1} X_{2i}^{\beta_2} X_{3i}^{\beta_3}$$

$$\beta_1 : E(Y/X_{2i} = X_{3i} = 1) = e^{\beta_1}$$

$\beta_2 = \varepsilon_{E(Y)/X_2}$: Khi X_2 thay đổi 1%, yếu tố khác không đổi, thì $E(Y)$ thay đổi β_2 %

Ví dụ mô hình : $E(Q_i) = e^{\beta_1} K_i^{\beta_2} L_i^{\beta_3}$

7.5. Hàm nửa Loga

Mô hình : $Y_i = \beta_1 + \beta_2 \ln X_i + u_i$

$$\beta_1 = E(Y/X = 1)$$

β_2 : Khi X tăng 1% thì $E(Y)$ thay đổi β_2 đơn vị.

Mô hình : $\ln Y_i = \beta_1 + \beta_2 X_i + u_i$

β_2 : Khi X tăng 1 đơn vị thì $E(Y)$ thay đổi β_2 %.

7.6. Hàm chi phí – lợi ích

C_i : chi phí

U_i : lợi ích

$$U_i = \beta_1 + \beta_2 C_i + \beta_3 C_i^2 + u_i$$

Bài 4. ĐA CỘNG TUYẾN

1. Bản chất của đa cộng tuyến (*Multicollinearity*)

1.1. Hiện tượng :

Xét MH: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i$

Gt 10: Các biến giải thích không có quan hệ cộng tuyến.

Nếu giả thiết bị vi phạm $\rightarrow$ hiện tượng đa cộng tuyến.

Có hai dạng đa cộng tuyến:

i. Đa cộng tuyến hoàn hảo (Perfect Multicollinearity) :

$\exists \lambda_j \neq 0 (j \neq 1)$ sao cho:

$$\lambda_2 X_{2i} + \dots + \lambda_k X_{ki} = 0 \quad \forall i$$

→ Ma trận X là suy biến, không có lời giải duy nhất.

ii. Đa cộng tuyến không hoàn hảo (Imperfect Multicollinearity) :

$\exists \lambda_j \neq 0 (j \neq 1)$ sao cho:

$$\lambda_2 X_{2i} + \dots + \lambda_k X_{ki} + v_i = 0$$

với v_i là SSNN có phương sai dương → vẫn có lời giải.

1.2. Nguyên nhân

Đa cộng tuyến hoàn hảo gần như không bao giờ xảy ra

Đa cộng tuyến không hoàn hảo thường xuyên xảy ra, do các nguyên nhân:

- Bản chất các biến giải thích có quan hệ tương quan với nhau.
- Do số liệu mẫu không ngẫu nhiên.
- Do kích thước mẫu không đủ.
- Do quá trình làm tròn số liệu.

2. Hậu quả

2.1. Đa cộng tuyến hoàn hảo : không giải được

Vì lúc đó $b_j = \frac{0}{0} \quad \forall j$ và

$$\text{Var}(b_j) = \infty \quad \forall j$$

2.2. Đa cộng tuyến không hoàn hảo:

- Các ước lượng có phương sai lớn, là ước lượng không hiệu quả.
- Khoảng tin cậy rộng không còn ý nghĩa.

- Các kiểm định T có thể sai.
- Các kiểm định T và kiểm định F có thể cho kết luận mâu thuẫn nhau.
- Các ước lượng có thể sai về dấu.
- Mô hình trở nên nhạy cảm với mỗi sự thay đổi của số biến giải thích và của tệp số liệu.

3. Phát hiện đa cộng tuyến.

3.1. Sự mâu thuẫn giữa kiểm định T và F

Có mâu thuẫn: Kiểm định F có ý nghĩa, tất cả các kiểm định T về các hệ số góc không có ý nghĩa.

→ có Đa cộng tuyến.

→ Điều ngược lại chưa chắc đúng.

3.2. Hồi qui phụ

Nghi ngờ biến giải thích X_j phụ thuộc tuyến tính vào các biến giải thích khác, hồi qui mô hình hồi qui phụ:

$$X_j = \alpha_1 + \alpha_2 X_2 + \dots + \alpha_{j-1} X_{j-1} + \alpha_{j+1} X_{j+1} + \dots + v_i \quad (*)$$

$$\begin{matrix} \vdots \\ n-1 \\ n \end{matrix} \quad \begin{matrix} R^2 = 0 \\ R^2 = 0 \end{matrix}$$

Mô hình ban đầu không có Đa cộng tuyến

Mô hình ban đầu có Đa cộng tuyến

$$\rightarrow F_{qs} = \frac{R^2}{1 - R^2} \cdot \frac{n - k}{k - 1} ; F_{qs} > F_{\alpha}(k - 1, n - k) \text{ thì bác bỏ } H_0.$$

3.3. Độ đo Theil

Dùng để so sánh mức độ đa cộng tuyến không hoàn hảo giữa các mô hình.

B-ớc 1: Hồi qui mô hình ban đầu tìm đ-ợc R^2

B-ớc 2: Bỏ biến X_j ra khỏi mô hình, hồi qui thu đợc R^2_{-j} ($j=2, k$)

$$m = R^2 - \frac{1}{k} (R^2 - R^2_{-j}) \text{ đợc gọi là độ đo Theil}$$

Ví dụ: Sử dụng tệp số liệu ch5bt4 về Tiêu dùng Y, Thu nhập X_2 và Tài sản có khả năng chuyển đổi cao X_3 của 25 hộ gia đình Mỹ để kiểm định hiện t-ợng đa cộng tuyến giữa các biến giải thích.

Kết quả hồi qui Y theo X_2 và X_3 nh- sau:

Dependent Variable: Y
Method: Least Squares
Date: 11/19/08 Time: 10:12
Sample: 1 25
Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	33.87971	19.11513	1.772403	0.0902
X2	-26.00263	34.95897	-0.743804	0.4649
X3	6.709261	8.740550	0.767602	0.4509
R-squared	0.741695	Mean dependent var		169.3680
Adjusted R-squared	0.718213	S.D. dependent var		79.05857
S.E. of regression	41.96716	Akaike info criterion		10.42382
Sum squared resid	38747.34	Schwarz criterion		10.57008
Log likelihood	-127.2977	F-statistic		31.58532
Durbin-Watson stat	2.785912	Prob(F-statistic)		0.000000

Hồi quy phụ của X_2 theo X_3 có kết quả sau:

Dependent Variable: X2
Method: Least Squares
Date: 11/19/08 Time: 10:17
Sample: 1 25
Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.104988	0.111892	-0.938303	0.3578
X3	0.250022	0.000157	1594.459	0.0000
R-squared	0.999991	Mean dependent var	159.4480	
Adjusted R-squared	0.999991	S.D. dependent var	81.46980	
S.E. of regression	0.250315	Akaike info criterion	0.144426	
Sum squared resid	1.441126	Schwarz criterion	0.241936	
Log likelihood	0.194678	F-statistic	2542299.	
Durbin-Watson stat	2.245068	Prob(F-statistic)	0.000000	

Hồi quy phụ của X_3 theo X_2 có kết quả sau:

Dependent Variable: X3
Method: Least Squares
Date: 11/19/08 Time: 10:19
Sample: 1 25
Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.425686	0.447288	0.951703	0.3511
X2	3.999613	0.002508	1594.459	0.0000
R-squared	0.999991	Mean dependent var	638.1560	
Adjusted R-squared	0.999991	S.D. dependent var	325.8492	
S.E. of regression	1.001168	Akaike info criterion	2.916830	
Sum squared resid	23.05376	Schwarz criterion	3.014340	
Log likelihood	-34.46038	F-statistic	2542299.	
Durbin-Watson stat	2.245125	Prob(F-statistic)	0.000000	

Để tính độ đo Theil ta hồi quy Y lần l- ợt với X_2 và X_3 . Kết quả nh- sau:

Dependent Variable: Y

Method: Least Squares
 Date: 11/19/08 Time: 10:23
 Sample: 1 25
 Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	36.73575	18.58133	1.977025	0.0601
X2	0.831821	0.104206	7.982448	0.0000
R-squared	0.734777	Mean dependent var	169.3680	
Adjusted R-squared	0.723246	S.D. dependent var	79.05857	
S.E. of regression	41.59070	Akaike info criterion	10.37025	
Sum squared resid	39785.09	Schwarz criterion	10.46776	
Log likelihood	-127.6281	F-statistic	63.71948	
Durbin-Watson stat	2.919889	Prob(F-statistic)	0.000000	

Dependent Variable: Y
 Method: Least Squares
 Date: 11/19/08 Time: 10:24
 Sample: 1 25
 Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	36.60968	18.57637	1.970766	0.0609
X3	0.208034	0.026033	7.991106	0.0000
R-squared	0.735199	Mean dependent var	169.3680	
Adjusted R-squared	0.723686	S.D. dependent var	79.05857	
S.E. of regression	41.55758	Akaike info criterion	10.36866	
Sum squared resid	39721.74	Schwarz criterion	10.46617	
Log likelihood	-127.6082	F-statistic	63.85778	
Durbin-Watson stat	2.916396	Prob(F-statistic)	0.000000	

Hãy dùng các kết quả trên tính độ đo Theil.

4. Khắc phục đa cộng tuyến.

4.1. Dùng thông tin tiên nghiệm.

Ví dụ: Xét mô hình $TD_i = \beta_1 + \beta_2 TN_i + \beta_3 SK_i + u_i$

Để thấy TD_i có cộng tuyến với SK_i

Nếu có thể cho rằng $\beta_3 = 0,1\beta_2$

Thì mô hình trở thành $TD_i = \beta_1 + \beta_2(TN_i + 0,1SK_i) + u_i$

Và đã khắc phục đ-ợc đa cộng tuyến.

4.2. Bỏ bớt biến nếu có thể.

Lúc đó việc lựa chọn biến bị loại khỏi mô hình có thể căn cứ vào kết quả của hồi quy phụ.

4.3. Tăng kích thước mẫu hoặc lấy mẫu mới nếu có thể.

4.4. Đổi dạng của mô hình.

Ví dụ thay vì hồi quy mô hình $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$

Ng-ời ta hồi quy mô hình $\ln Y_i = \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$

4.5. Dùng sai phân cấp 1.

Xét mô hình $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$

Tại thời điểm $t-1$ mô hình có dạng:

$$Y_{t-1} = \beta_1 + \beta_2 X_{2t-1} + \beta_3 X_{3t-1} + u_{t-1}$$

Lấy sai phân ta có:

$$Y_t - Y_{t-1} = \beta_2 (X_{2t} - X_{2t-1}) + \beta_3 (X_{3t} - X_{3t-1}) + (u_t - u_{t-1}) \quad (*)$$

Mô hình (*) đ-ợc gọi là mô hình sai phân cấp 1.

4.6. Giảm cộng tuyến trong hồi quy đa thức.

Có thể giảm cộng tuyến trong hồi quy đa thức bằng cách lấy sai phân của các biến trong mô hình so với giá trị trung bình của chúng.

Bài 5. PHƯƠNG SAI CỦA SAI SỐ THAY ĐỔI

1. Bản chất của hiện tượng phương sai của sai số thay đổi

MH ban đầu: $Y_i = \beta_1 + \beta_2 X_i + u_i$

1.1. Hiện tượng

Gt 4: Phương sai các sai số ngẫu nhiên là đồng nhất

$Var(u_i) \equiv \sigma^2$ không đổi.

Nếu gt được thỏa mãn $\rightarrow$ PSSS đồng đều (*homoscedasticity*).

Gt không thỏa mãn : $Var(u_i) = \sigma_i^2$ không đồng nhất

$\rightarrow$ PSSS thay đổi (*heteroscedasticity*).

1.2. Nguyên nhân

- Bản chất hiện tượng Kinh tế xã hội.

- Số liệu không đúng bản chất hiện tượng.
- Quá trình xử lý số liệu.

2. Hậu quả

- Các ước lượng là không chệch, nhưng không hiệu quả → không phải là tốt nhất.
- Các kiểm định T, F có thể sai, khoảng tin cậy rộng.

3. Phát hiện

$Var(u_i) = \sigma_i^2$ là không biết. Dùng ước lượng của nó là e_i^2 để phân tích đánh giá.

3.1. Đồ thị phần dư

Dùng đồ thị của e_i , $|e_i|$ hoặc e_i^2 để đánh giá.

3.2. Kiểm định Park

Giả thiết: $\sigma_i^2 = \sigma^2 X_i^{\alpha_2}$

Trong đó σ^2 là một hằng số

→ MH hồi qui phụ $\ln e_i^2 = \alpha_1 + \alpha_2 \ln X_i + v_i$ (*)

B-ớc 1. Hồi quy mô hình ban đầu tìm đ-ợc e_i

B-ớc 2. Hồi quy mô hình (*)

B-ớc 3. Kiểm định cặp giả thuyết $H_0: \alpha_2 = 0$; $H_1: \alpha_2 \neq 0$

3.3. Kiểm định Glejer

Tùy vào giả thiết mà thực hiện hồi qui phụ để kiểm định

Gt : $\sigma_i^2 = \sigma^2 X_i$, do đó hồi qui mô hình hồi qui phụ

$$|e_i| = \alpha_1 + \alpha_2 X_i + v_i (*)$$

$\begin{matrix} \vdots \\ n_0 : a_2 = 0 : R_*^2 = 0 \\ \vdots \end{matrix}$ Mô hình đầu có PSSS đồng đều

$\begin{matrix} \vdots \\ n_1 : a_2^{-1} : 0 : R_*^{2-1} = 0 \\ \vdots \end{matrix}$ Mô hình đầu có PSSS thay đổi

Dùng kiểm định T hoặc F để kiểm định

Tương tự

Gt : $\sigma_i^2 = \sigma^2 X_i^2 \rightarrow$ MH hồi qui phụ $|e_i| = \alpha_1 + \alpha_2 X_i^2 + v_i$

Gt : $\sigma_i^2 = \sigma^2 \sqrt{X_i} \rightarrow$ MH hồi qui phụ $|e_i| = \alpha_1 + \alpha_2 \sqrt{X_i} + v_i$

Gt : $\sigma_i^2 = \sigma^2 \frac{1}{X_i} \rightarrow$ MH hồi qui phụ $|e_i| = \alpha_1 + \alpha_2 \frac{1}{X_i} + v_i$

3.4. Kiểm định White

Dùng cho mô hình nhiều biến giải thích. Hồi qui bình phương phần dư theo tổ hợp bậc cao dần của các biến giải thích.

VD : MH ban đầu $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$

$\rightarrow$ MH hồi qui phụ :

$$e^2 = \alpha_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_2^2 + \alpha_5 X_3^2 + \alpha_6 X_2 X_3 (+ \dots) + v_i (*)$$

$$\begin{matrix} \vdots \\ n_0 : R_*^2 = 0 \\ \vdots \\ n_1 : R_*^{2-1} = 0 \\ \vdots \end{matrix}$$

Kiểm định $\chi^2 : c_{qs}^2 = n R_*^2$, nếu $c_{qs}^2 > c_a^2(k_* - 1)$ thì bác bỏ H_0

Ví dụ: Tập số liệu ch6bt3 bao gồm các biến D88 là nợ n-ớc ngoài và Y88 là tổng sản phẩm trong n-ớc của 73 n-ớc đang phát triển. Hãy kiểm định hiện tượng sai số thay đổi.

Kết quả hồi quy D88 theo Y88 nh- sau:

Dependent Variable: D88
Method: Least Squares
Date: 11/19/08 Time: 18:28
Sample: 1 73
Included observations: 73

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3521.957	1066.371	3.302752	0.0015
Y88	0.275643	0.017613	15.64971	0.0000

R-squared 0.775255 Mean dependent var 10982.11
Adjusted R-squared 0.772090 S.D. dependent var 17071.75
S.E. of regression 8150.046 Akaike info criterion 20.87645
Sum squared resid 4.72E+09 Schwarz criterion 20.93920
Log likelihood -759.9904 F-statistic 244.9136
Durbin-Watson stat 2.081218 Prob(F-statistic) 0.000000

Kiểm định Park cho kết quả sau:

Dependent Variable: LOG(E^2)
Method: Least Squares
Date: 11/19/08 Time: 18:54
Sample: 1 73
Included observations: 73

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.549893	1.619225	5.897819	0.0000
LOG(Y88)	0.675378	0.177575	3.803343	0.0003

R-squared 0.169255 Mean dependent var 15.62156
Adjusted R-squared 0.157554 S.D. dependent var 2.521699
S.E. of regression 2.314538 Akaike info criterion 4.543312
Sum squared resid 380.3532 Schwarz criterion 4.606065
Log likelihood -163.8309 F-statistic 14.46542
Durbin-Watson stat 2.267623 Prob(F-statistic) 0.000299

Kiểm định Gleijer cho kết quả sau:

Dependent Variable: SQR(E)
Method: Least Squares
Date: 11/19/08 Time: 18:57
Sample(adjusted): 4 72
Included observations: 20
Excluded observations: 49 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	48.17599	11.98700	4.019020	0.0008
Y88	0.000774	0.000210	3.688858	0.0017

R-squared 0.430518 Mean dependent var 81.15995
Adjusted R-squared 0.398880 S.D. dependent var 46.05027
S.E. of regression 35.70366 Akaike info criterion 10.08302
Sum squared resid 22945.52 Schwarz criterion 10.18260
Log likelihood -98.83023 F-statistic 13.60767
Durbin-Watson stat 0.184354 Prob(F-statistic) 0.001680

Kiểm định White cho kết quả sau:

White Heteroskedasticity Test:

F-statistic	32.12701	Probability	0.000000
Obs*R-squared	34.93782	Probability	0.000000

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 11/19/08 Time: 18:59

Sample: 1 73

Included observations: 73

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-32588022	20003195	-1.629141	0.1078
Y88	5760.776	817.7877	7.044341	0.0000
Y88^2	-0.016020	0.003069	-5.219110	0.0000
R-squared	0.478600	Mean dependent var	64603429	
Adjusted R-squared	0.463703	S.D. dependent var	1.84E+08	
S.E. of regression	1.35E+08	Akaike info criterion	40.31714	
Sum squared resid	1.27E+18	Schwarz criterion	40.41127	
Log likelihood	-1468.576	F-statistic	32.12701	
Durbin-Watson stat	2.049448	Prob(F-statistic)	0.000000	

3. Khắc phục

Dựa trên giả thiết về sự thay đổi của PSSS thay đổi mà khắc phục

3.1. Nếu biết σ_i^2 – Dùng WLS- *Ph- ơng pháp bình ph- ơng nhỏ nhất có trọng số.*

Chia hai vế mô hình cho σ_i

$$\frac{Y_i}{\sigma_i} = b_1 \frac{1}{\sigma_i} + b_2 \frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i} \Leftrightarrow Y_i' = \beta_1 X_{0i} + \beta_2 X_i' + u_i'$$

$Var(u_i') = 1$ không đổi

3.2. Nếu chưa biết σ_i^2 – Dùng GLS □ *Ph- ơng pháp bình ph- ơng nhỏ nhất tổng quát.*

Tùy thuộc vào tính chất của σ_i^2 mà biến đổi mô hình gốc sao cho ph- ơng sai của sai số ngẫu nhiên trở nên đồng đều.

$$Gt 1 : \sigma_i^2 = \sigma^2 X_i$$

Lúc đó chia hai vế cho $\sqrt{X_i}$

$$\frac{Y_i}{\sqrt{X_i}} = b_1 \frac{1}{\sqrt{X_i}} + b_2 \sqrt{X_i} + \frac{u_i}{\sqrt{X_i}}$$

$$\Rightarrow \text{PSSS sẽ bằng } \sigma^2$$

$$\text{Gt 2 : } \sigma_i^2 = \sigma^2 X_i^2$$

Lúc đó chia hai vế cho X_i

$$\text{Gt 3 : } \sigma_i^2 = \sigma^2 E(Y_i)^2$$

Lúc đó chia hai vế cho $\hat{Y}_i$

Chú ý: Có thể thay đổi dạng hàm để khắc phục ph- ơng sai của sai số thay đổi.

a. $\ln Y_i = \beta_1 + \beta_2 \ln X_i + u_i$

b. $\ln Y_i = \beta_1 + \beta_2 X_i + u_i$

c. $Y_i = \beta_1 + \beta_2 \ln X_i + u_i$

Ví dụ: Tiếp tục sử dụng tệp số liệu trên, giả sử $\sigma_i^2 = \sigma^2 Y_{88i}$ lúc đó ta hồi quy mô hình sau:

$$\frac{D_{88i}}{\text{sqr}(Y_{88})} = b_1 \frac{1}{\text{sqr}(Y_{88})} + b_2 \text{sqr}(Y_{88}) + \frac{u_i}{\text{sqr}(Y_{88})}$$

Kết quả hồi quy nh- sau:

Dependent Variable: D88/SQR(Y88)

Method: Least Squares

Date: 11/19/08 Time: 21:57

Sample: 1 73

Included observations: 73

Variable	Coefficient	Std. Error	t-Statistic	Prob.
1/SQR(Y88)	1024.466	270.9094	3.781581	0.0003
SQR(Y88)	0.367922	0.028848	12.75364	0.0000
R-squared	0.308884	Mean dependent var	64.11631	
Adjusted R-squared	0.299150	S.D. dependent var	45.42712	
S.E. of regression	38.03013	Akaike info criterion	10.14165	
Sum squared resid	102686.7	Schwarz criterion	10.20440	
Log likelihood	-368.1702	Durbin-Watson stat	2.013380	

White Heteroskedasticity Test:

F-statistic	2.248267	Probability	0.072848
Obs*R-squared	8.526662	Probability	0.074083

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 11/20/08 Time: 22:00

Sample: 1 73

Included observations: 73

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-735.5216	5167.096	-0.142347	0.8872
1/SQR(Y88)	-3833.260	285895.2	-0.013408	0.9893
(1/SQR(Y88))^2	910480.8	4365419.	0.208567	0.8354
SQR(Y88)	19.78189	28.10322	0.703901	0.4839
(SQR(Y88))^2	-0.018487	0.040348	-0.458186	0.6483
R-squared	0.116804	Mean dependent var	1406.667	
Adjusted R-squared	0.064851	S.D. dependent var	3300.866	
S.E. of regression	3192.040	Akaike info criterion	19.04074	
Sum squared resid	6.93E+08	Schwarz criterion	19.19762	
Log likelihood	-689.9871	F-statistic	2.248267	
Durbin-Watson stat	2.115989	Prob(F-statistic)	0.072848	

cuu duong than cong . com

Bài 6. TỰ TƯƠNG QUAN

1. Hiện tượng tự tương quan (*Autocorrelation or Serial correlation*)

1.1. Hiện tượng

MH ban đầu: $Y_t = \beta_1 + \beta_2 X_t + u_t$

Gt 5: Các sai số ngẫu nhiên không tương quan

$$\text{Cov}(u_i, u_j) = 0 \quad (i \neq j) \quad \text{hoặc} \quad \text{Cov}(u_t, u_{t-p}) = 0 \quad (p \neq 0)$$

Nếu gt bị vi phạm : hiện tượng tự tương quan bậc p

Xét trường hợp $p = 1$

u_t và u_{t-1} có cùng trung bình và phương sai

$$\rightarrow u_t = \rho u_{t-1} + \varepsilon_t \quad (-1 \leq \rho \leq 1, \quad \varepsilon_t \text{ thỏa mãn các giả thiết của OLS})$$

- $\rho = -1$ tự tương quan âm hoàn hảo
- $-1 < \rho < 0$ tự tương quan âm
- $\rho = 0$ không có tự tương quan
- $0 < \rho < 1$ tự tương quan dương
- $\rho = 1$ tự tương quan dương hoàn hảo

$\Rightarrow$ L-ợc đồ $AR(1)$

Tổng quát : tự tương quan bậc p :

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \varepsilon_t \quad \text{với } \rho_p \neq 0$$

$\Rightarrow$ L-ợc đồ $AR(p)$: *Autoregressive Procedure order p*.

1.2. Nguyên nhân

- Bản chất, tính quán tính trong hiện tượng kinh tế xã hội
- Hiện tượng mạng nhện trong kinh tế
- Quá trình xử lý, nội suy số liệu
- Mô hình thiếu biến hoặc dạng hàm sai

2. Hậu quả

- Các ước lượng là không chệch nhưng không còn là ước lượng tốt nhất.

3. Phát hiện

3.1. Quan sát đồ thị của e_t theo e_{t-1}

Bước 1. Hồi quy mô hình gốc để tìm e_t và e_{t-1}

Bước 2. Vẽ đồ thị của e_t theo e_{t-1} và nhận xét.

3.2. Kiểm định Durbin – Watson

Kiểm định Durbin $\square$ Watson dựa trên thống kê

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \cong 2(1 - \hat{\rho})$$

$$\text{với } \hat{\rho} = \frac{\sum_{i=1}^n e_i e_{i-1}}{\sum_{i=1}^n e_i^2} \text{ là ước lượng cho } \rho$$

Chú ý: Kiểm định DW sẽ chỉ áp dụng được nếu thỏa mãn các điều kiện sau:

- + Mô hình phải có hệ số chặn.
- + Biến giải thích phải là phi ngẫu nhiên.
- + Nếu có hiện tượng tự tương quan thì đó chỉ là mô hình AR(1).
- + Mô hình không chứa biến trễ của biến phụ thuộc làm biến giải thích.
- + Không có quan sát nào bị mất trong tập số liệu.

Do không tìm được chính xác phân phối xác suất của d nên dựa vào tính chất của nó để kết luận.

$$\text{Do } -1 \leq \hat{\rho} \leq 1 \rightarrow 0 \leq d \leq 4$$

Với $n, k' = k - 1$ và $\alpha = 0,05$ cho trước, tra bảng $\rightarrow d_L$ và d_U

Tự tương quan dương $\rho > 0$	Không có kết luận	Không có tự tương quan $\rho = 0$	Không có kết luận	Tự tương quan âm $\rho < 0$
0	d_L	d_U	2	4
			$4 - d_U$	$4 - d_L$
				4

Hạn chế:

- + Vẫn còn hai miền không có kết luận $\Rightarrow$ Dùng kiểm định DW cải biên.
- + Không áp dụng được với các mô hình tự hồi quy.

3.3. Hồi qui phụ

Kiểm định

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \varepsilon_t$$

$$\text{MH hồi qui phụ : } e_t = (\alpha_0) + \alpha_1 e_{t-1} + \dots + \alpha_p e_{t-p} + v_t (*)$$

$$\begin{cases} n_0 : R^2 = 0 \\ n_1 : R^2 > 0 \end{cases}$$

Không có tự tương quan đến bậc p

Có tự tương quan ở bậc tương ứng

$$\Leftrightarrow \begin{cases} n_0 : a_1 = \dots = a_p = 0 \\ n_1 : a_{j-1} = 0 : (j-1, 0) \end{cases}$$

Kiểm định T hoặc F

3.4. Kiểm định Breusch- Goldfrey.

MH hồi qui phụ

$$e_t = [\beta_1 + \beta_2 X_t] + \alpha_1 e_{t-1} + \dots + \alpha_p e_{t-p} + v_t (*)$$

$$\begin{cases} n_0 : a_1 = \dots = a_p = 0 \\ n_1 : a_{j-1} = 0 : (j-1, 0) \end{cases}$$

Kiểm định χ^2 : $c_{qs}^2 = n_* R_*^2 = (n - p) R_*^2$, nếu $c_{qs}^2 > c_{\alpha}^2(p)$ thì bác bỏ H_0

Kiểm định F:

Hồi qui

$$e_t = [\beta_1 + \beta_2 X_t] + v_t (**)$$

$$F_{qs} = \frac{R_*^2 - R_{**}^2}{1 - R_*^2} \cdot \frac{n_* - k_*}{k_* - 1} \quad \text{Nếu } F_{qs} > F_{\alpha}(k_* - 1; n_* - k_*) \text{ thì bác bỏ } H_0$$

Ví dụ: Tập số liệu ch7bt4 gồm hai biến Tiêu dùng và thu nhập của Nigeria. Hãy kiểm định hiện t- ơng tự t- ơng quan của mô hình.

Hồi quy Tiêu dùng theo Thu nhập đ- ợc kết quả sau:

Dependent Variable: CONS
Method: Least Squares
Date: 11/19/08 Time: 22:08
Sample: 1960 1986
Included observations: 27

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	155.2239	203.4712	0.762879	0.4527
GDP	0.597069	0.060594	9.853648	0.0000
R-squared	0.795240	Mean dependent var	2037.449	
Adjusted R-squared	0.787050	S.D. dependent var	789.2231	
S.E. of regression	364.1989	Akaike info criterion	14.70446	
Sum squared resid	3316021.	Schwarz criterion	14.80045	
Log likelihood	-196.5103	F-statistic	97.09438	

Durbin-Watson stat 0.462830 Prob(F-statistic) 0.000000

Theo báo cáo thì thống kê $DW=0,46283$. Với $n=27$, $k'=1$ tra bảng được $d_l= 1,316$ và $d_u= 1,496$ suy ra mô hình có tự tương quan dương.

Dùng phương pháp hồi quy phụ cho kết quả sau:

Dependent Variable: E

Method: Least Squares

Date: 11/21/08 Time: 08:46

Sample(adjusted): 1963 1986

Included observations: 24 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.256609	52.50711	0.004887	0.9961
E(-1)	0.843549	0.221024	3.816555	0.0011
E(-2)	0.020077	0.293835	0.068326	0.9462
E(-3)	-0.173636	0.233916	-0.742300	0.4665
R-squared	0.604411	Mean dependent var	-8.530418	
Adjusted R-squared	0.545072	S.D. dependent var	378.2292	
S.E. of regression	255.1093	Akaike info criterion	14.07227	
Sum squared resid	1301615.	Schwarz criterion	14.26862	
Log likelihood	-164.8673	F-statistic	10.18582	
Durbin-Watson stat	2.022007	Prob(F-statistic)	0.000278	

Kiểm định BG cho kết quả sau:

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	34.31433	Probability	0.000005
Obs*R-squared	15.88781	Probability	0.000067

Test Equation:

Dependent Variable: RESID

Method: Least Squares

Date: 11/19/08 Time: 22:11

Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-60.84700	133.6292	-0.455342	0.6530
GDP	0.021511	0.039844	0.539890	0.5942
RESID(-1)	0.777523	0.132732	5.857844	0.0000
R-squared	0.588437	Mean dependent var	-8.42E-14	
Adjusted R-squared	0.554140	S.D. dependent var	357.1264	
S.E. of regression	238.4630	Akaike info criterion	13.89074	
Sum squared resid	1364750.	Schwarz criterion	14.03473	
Log likelihood	-184.5250	F-statistic	17.15717	
Durbin-Watson stat	1.846319	Prob(F-statistic)	0.000024	

4. Khắc phục

Mục đích là chuyển MH ban đầu có khuyết tật tự tương quan thành MH mới có cùng hệ số cũ nhưng không có tự tương quan

Mô hình ban đầu: $Y_t = \beta_1 + \beta_2 X_t + u_t$

Có tự tương quan : $u_t = \rho u_{t-1} + \varepsilon_t$ với $\rho \neq 0$, ε_t thỏa mãn các giả thiết OLS.

4.1. Khi ρ đã biết □ Dùng phương pháp sai phân tổng quát.

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

$$\Leftrightarrow Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \Leftrightarrow \rho Y_{t-1} = \rho \beta_1 + \beta_2 \rho X_{t-1} + \rho u_{t-1}$$

$\rightarrow Y_t - \rho Y_{t-1} = \beta_1(1 - \rho) + \beta_2(X_t - \rho X_{t-1}) + u_t - \rho u_{t-1}$ (phương trình sai phân tổng quát)

$$\rightarrow Y_t^* = \beta_1^* + \beta_2 X_t^* + \varepsilon_t$$

$$\text{Ước lượng bằng OLS} \rightarrow \hat{\beta}_1^* \rightarrow \hat{\beta}_1 = \frac{\hat{\beta}_1^*}{1 - \rho} \text{ và } \hat{\beta}_2$$

4.2. Khi ρ chưa biết

Ước lượng ρ bằng các phương pháp khác nhau

i. Từ thống kê Durbin-Watson

$$d \cong 2(1 - \hat{\rho}) \rightarrow \hat{\rho} = 1 - \frac{d}{2}$$

ii. Từ hồi qui phụ

$$e_t = (\alpha_0) + \alpha_1 e_{t-1} + v_t \rightarrow \text{lấy } \hat{r} = \hat{a}_1$$

iii. Phương pháp Cochran-Orcutt

Hồi qui mô hình ban đầu: $Y_t = \beta_1 + \beta_2 X_t + u_t \rightarrow \hat{b}_1^{(1)}, \hat{b}_2^{(1)}, e_t^{(1)}$

Hồi qui mô hình $e_t^{(1)} = \alpha + \rho e_{t-1}^{(1)} + v_t \rightarrow \hat{r}^{(1)}$

Lấy $\hat{r}^{(1)}$ thay vào phương trình sai phân tổng quát $\rightarrow \hat{b}_1^{(2)}, \hat{b}_2^{(2)}, e_t^{(2)}$

Hồi qui mô hình $e_t^{(2)} = \alpha + \rho e_{t-1}^{(2)} + v_t \rightarrow \hat{r}^{(2)}$

Lấy $\hat{r}^{(2)}$ thay vào phương trình sai phân tổng quát $\rightarrow \hat{b}_1^{(3)}, \hat{b}_2^{(3)}, e_t^{(3)}$

...

Quá trình lặp cho đến khi $\hat{r}$ ở hai bước kế tiếp chênh lệch nhau không đáng kể, β_1 và β_2 ở bước cuối cùng là ước lượng cho β_1 và β_2

iiii. Phương pháp Durbin □ Watson hai bước.

Trước hết viết lại mô hình sai phân tổng quát dưới dạng:

$$Y_t = \beta_1(1 - \rho) + \beta_2 X_t - \beta_2 \rho X_{t-1} + \rho Y_{t-1} + (u_t - \rho u_{t-1})$$

Bước 1. Ước lượng mô hình trên thu được $\hat{r}$

Bước 2. Thiết lập mô hình sai phân tổng quát và ước lượng nó để tìm $\hat{b}_1$ và $\hat{b}_2$

iiii. Đưa thêm biến vào mô hình.

Một trong các nguyên nhân của tự tương quan là mô hình bị chỉ định sai do bỏ sót biến giải thích vì vậy việc đưa thêm biến vào mô hình cũng có thể khắc phục tự tương quan.

Ví dụ: Tiếp tục ví dụ đã cho, ta tìm cách khắc phục hiện tượng tự tương quan của mô hình.

Nếu sử dụng thống kê DW, từ báo cáo ta có $d = 0,46283$

$$\Rightarrow \hat{r} = 1 - \frac{d}{2} = 0,768585.$$

Thiết lập các biến sai phân tổng quát:

$$\text{Cons1}_t = \text{cons}_t - 0,768585\text{cons}_{t-1}$$

$$\text{gdp1}_t = \text{gdp}_t - 0,768585\text{gdp}_{t-1}$$

Hồi quy Cons1 với gdp1 thu được kết quả sau:

Dependent Variable: CONS1

Method: Least Squares

Date: 11/20/08 Time: 22:28

Sample(adjusted): 1961 1986

Included observations: 26 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	143.5202	123.3592	1.163433	0.2561
GDP1	0.467009	0.140706	3.319044	0.0029

R-squared 0.314600 Mean dependent var 523.5229

Adjusted R-squared 0.286042 S.D. dependent var 277.1436

S.E. of regression 234.1754 Akaike info criterion 13.82382

Sum squared resid 1316115. Schwarz criterion 13.92060

Log likelihood -177.7097 F-statistic 11.01605

Durbin-Watson stat 1.718728 Prob(F-statistic) 0.002875

Kiểm định DW cho thấy mô hình sai phân tổng quát đã không có tự tương quan.

Để sử dụng phương pháp lặp Cochrane- Orcutt , ta khai báo thêm vào danh sách biến giải thích bậc của tự tương quan, chẳng hạn ở đây là AR(1) và hồi quy bằng OLS thu được kết quả sau:

Dependent Variable: CONS
Method: Least Squares
Date: 11/20/08 Time: 22:37
Sample(adjusted): 1961 1986
Included observations: 26 after adjusting endpoints
Convergence achieved after 39 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1290.375	864.3354	1.492910	0.1491
GDP	0.320087	0.180066	1.777612	0.0887
AR(1)	0.895753	0.097838	9.155465	0.0000

R-squared	0.917399	Mean dependent var	2068.732
Adjusted R-squared	0.910216	S.D. dependent var	787.5967
S.E. of regression	235.9953	Akaike info criterion	13.87367
Sum squared resid	1280957.	Schwarz criterion	14.01883
Log likelihood	-177.3577	F-statistic	127.7230
Durbin-Watson stat	1.884378	Prob(F-statistic)	0.000000

Inverted AR Roots .90

Kết quả thu được là - ước lượng ở bậc cuối cùng, đồng thời kết quả cũng cho - ước lượng của bậc độ AR(1):

$$cons_t = 1290,375 + 0,320087gdp_t + e_t$$

$$e_t = 0,895753e_{t-1} + e_t$$

Kết quả cũng cho - ước lượng của hệ số tự tương quan bậc nhất bằng 0,895753 chứng tỏ mức độ tự tương quan khá cao.

Ta cũng có thể khắc phục tự tương quan của mô hình bằng cách đưa thêm biến trễ vào mô hình vì theo lý thuyết kinh tế thì tiêu dùng thường còn phụ thuộc vào tiêu dùng ở kỳ trước, vì vậy có thể thử nghiệm mô hình sau:

$$Cons_t = \beta_1 + \beta_2gdp_t + \alpha Cons_{t-1} + u_t$$

Kết quả như sau:

Dependent Variable: CONS
Method: Least Squares
Date: 11/20/08 Time: 23:07
Sample(adjusted): 1961 1986
Included observations: 26 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.962031	122.4302	0.032362	0.9745
GDP	0.207495	0.064345	3.224707	0.0038
CONS(-1)	0.695319	0.094013	7.396025	0.0000

R-squared	0.937124	Mean dependent var	2068.732
Adjusted R-squared	0.931656	S.D. dependent var	787.5967
S.E. of regression	205.8988	Akaike info criterion	13.60081
Sum squared resid	975068.9	Schwarz criterion	13.74598
Log likelihood	-173.8106	F-statistic	171.3986
Durbin-Watson stat	1.919159	Prob(F-statistic)	0.000000

5. Mô hình ARCH(Autoregressive Conditional Heterocedasticity Model)

Hiện tượng tự tương quan thời gian xảy ra đối với các số liệu thời gian trong khi hiện tượng phương sai của sai số thay đổi thời gian xảy ra đối với các số liệu chéo. Tuy nhiên hiện tượng phương sai của sai số thay đổi cũng vẫn thời gian xảy ra đối với các số liệu thời gian.

Trong kinh tế khi phân tích và dự báo các chuỗi thời gian như giá cổ phiếu, tỷ lệ lạm phát, tỷ giá hối đoái... người ta nhận thấy rằng khả năng dự báo các biến này trong các giai đoạn khác nhau là khác nhau đáng kể. Có giai đoạn sai số dự báo rất nhỏ, ngược lại có giai đoạn lại rất lớn. Sự biến động này nằm trong tính không chắc chắn của thị trường tài chính, sự nhạy cảm với các tin đồn, sự thay đổi chính sách tiền tệ hay thuế... Điều đó làm cho phương sai của sai số dự báo thay đổi giữa các giai đoạn khác nhau tức là có hiện tượng tự tương quan chuỗi trong phương sai của sai số dự báo.

Vì hành vi của sai số dự báo có thể giả thiết phụ thuộc vào hành vi của các sai số ngẫu nhiên U_t nên ta có thể áp dụng hiện tượng tự tương quan chuỗi đối với phương sai của U_t . Để mô tả hiện tượng này ENGLE đã xây dựng mô hình ARCH. Tổng cơ bản của mô hình này là phương sai của sai số ngẫu nhiên tại thời điểm t (s_t^2) phụ thuộc vào bình phương của sai số ngẫu nhiên tại thời điểm $t-1$ — U_{t-1}^2 .

Xét mô hình k biến:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + u_t$$

Và giả thiết rằng

$$U_t \sim N(0; a_0 + a_1 u_{t-1}^2 + \dots + a_p u_{t-p}^2)$$

Tức là U_t có phân phối chuẩn với kỳ vọng bằng 0 phương sai bằng $a_0 + a_1 u_{t-1}^2 + \dots + a_p u_{t-p}^2$. L -ợc đồ trên gọi là ARCH(1).

Một cách tổng quát L -ợc đồ ARCH(p) có dạng:

$$\text{var}(u_t) = s_t^2 = a_0 + a_1 u_{t-1}^2 + a_2 u_{t-2}^2 + \dots + a_p u_{t-p}^2$$

Nếu $\alpha_1 = \alpha_2 = \dots = \alpha_p$ (không có t-ơng quan chuỗi trong sai số của ph-ơng sai) thì ta có ph-ơng sai của sai số là đồng đều.

Theo Engle để kiểm định cặp giả thuyết:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_p$$

H_1 : Có ít nhất một hệ số khác không

Thì tiến hành hồi quy phụ:

$$e_t^2 = a_0 + a_1 e_{t-1}^2 + a_2 e_{t-2}^2 + \dots + a_p e_{t-p}^2 + e_t$$

Trong đó e_t là phần d- của mô hình gốc và dùng kiểm định F về sự thích hợp của mô hình hồi quy phụ.

Ngoài ra có thể dùng tiêu chuẩn kiểm định

$$\chi^2 = nR^2 \sim \chi^2(p)$$

Với R^2 là hệ số xác định của mô hình hồi quy phụ.

Nếu $\chi_{qs}^2 > \chi_{\alpha}^{2(p)}$ thì bác bỏ H_0 .

Chẳng hạn với mô hình trên kiểm định ARCH cho kết quả sau:

ARCH Test:

F-statistic	10.59560	Probability	0.003360
Obs*R-squared	7.963025	Probability	0.004774

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 11/20/08 Time: 23:10

Sample(adjusted): 1961 1986

Included observations: 26 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	57717.39	37683.49	1.531636	0.1387
RESID^2(-1)	0.551280	0.169359	3.255089	0.0034
R-squared	0.306270	Mean dependent var	126786.4	
Adjusted R-squared	0.277365	S.D. dependent var	186797.1	
S.E. of regression	158792.4	Akaike info criterion	26.86239	
Sum squared resid	6.05E+11	Schwarz criterion	26.95916	
Log likelihood	-347.2110	F-statistic	10.59560	
Durbin-Watson stat	2.129578	Prob(F-statistic)	0.003360	

Kết quả cho thấy mô hình có ARCH(1)

Nếu có hiện tượng ARCH thì áp dụng phương pháp GLS để khắc phục hiện tượng phương sai của sai số thay đổi.

Với mô hình đang xét, phương pháp GLS cho kết quả sau:

Dependent Variable: CONS

Method: ML - ARCH (Marquardt)

Date: 11/20/08 Time: 23:17

Sample: 1960 1986

Included observations: 27

Convergence achieved after 73 iterations

Variance backcast: ON

	Coefficient	Std. Error	z-Statistic	Prob.
C	422.5544	58.32670	7.244614	0.0000
GDP	0.460517	0.018060	25.49888	0.0000
Variance Equation				
C	2738.626	5899.843	0.464186	0.6425
ARCH(1)	1.927399	0.650321	2.963764	0.0030
R-squared	0.709273	Mean dependent var	2037.449	
Adjusted R-squared	0.671352	S.D. dependent var	789.2231	
S.E. of regression	452.4448	Akaike info criterion	13.79384	
Sum squared resid	4708244.	Schwarz criterion	13.98581	
Log likelihood	-182.2168	F-statistic	18.70396	
Durbin-Watson stat	0.300423	Prob(F-statistic)	0.000002	

Khái quát hóa của mô hình ARCH là các lớp mô hình GARCH, trong đó phương sai có điều kiện của U_t không chỉ phụ thuộc vào bình phương của sai số ngẫu nhiên trong quá khứ mà còn phụ thuộc vào phương sai có điều kiện trong quá khứ. Đó là các lớp mô hình GARCH, TARARCH, E-GARCH, GARCH-M...

Chú ý rằng kiểm định DW khi có hiện tượng ARCH có thể mất chính xác vì các hiện tượng tự tương quan và ARCH có thể kết hợp với nhau, do đó đối với các số liệu tài chính cần tiến hành kiểm định ARCH trước khi đưa ra kết luận dựa vào giá trị DW.

Bài 7. CHỈ ĐỊNH MÔ HÌNH

1. Các loại sai lầm chỉ định và hậu quả.

1.1. Mô hình thừa biến giải thích

Ví dụ: Mô hình đúng: $Y_i = \beta_1 + \beta_2 X_i + u_i$

Mô hình sai: $Y_i = \alpha_1 + \alpha_2 X_i + \alpha_3 Z_i + v_i$

Nếu mô hình thừa biến giải thích thì các ước lượng vẫn là không chệch và vững, nhưng không hiệu quả, khoảng tin cậy rộng.

Kiểm định bằng cách bỏ bớt biến số nghi là không cần thiết và dùng kiểm định với hệ số tương ứng để kết luận

1.2. Mô hình thiếu biến.

Ví dụ: Mô hình đúng: $Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_i + u_i$

Mô hình sai: $Y_i = \alpha_1 + \alpha_2 X_i + v_i$

Nếu mô hình thiếu biến thì các ước lượng sẽ bị chệch nên không đáng tin cậy.

1.3. Dạng hàm sai.

Ví dụ : Mô hình đúng: $Y_i = \beta_1 + \beta_2 X_i + u_i$

Mô hình sai: $\ln Y_i = \beta_1 + \beta_2 \ln X_i + v_i$

Có thể kiểm định thiếu biến và dạng hàm sai bằng các kiểm định sau:

1. Kiểm định Ramsey

$$\text{Mô hình ban đầu: } Y_i = \beta_1 + \beta_2 X_i + u_i \quad (1)$$

Nếu cho rằng mô hình thiếu biến Z_i nào đó thì:

B1: Hồi qui mô hình ban đầu thu được các giá trị ước lượng $\hat{Y}_i$

B2: Hồi qui MH hồi qui phụ :

$$Y_i = [\beta_1 + \beta_2 X_i] + \alpha_1 \hat{Y}_i^2 + \dots + \alpha_m \hat{Y}_i^{m+1} + u_i \quad (2)$$

$$\left\{ \begin{array}{l} H_0: a_1 = \dots = a_m = 0 \\ H_1: a_j \neq 0, j = 1, m \end{array} \right. \quad \begin{array}{l} \text{MH (1) không thiếu} \\ \text{biến} \end{array}$$

MH (1) thiếu biến

$$F_{qs} = \frac{R_{(2)}^2 - R_{(1)}^2}{1 - R_{(2)}^2} \cdot \frac{n - k_{(2)}}{k_{(2)} - 1}$$

Nếu $F_{qs} > F_{\alpha}(k_{(2)} - 1; n - k_{(2)})$ bác bỏ H_0

2. Kiểm định nhân tử Lagrange (LM)

B1: Hồi qui mô hình ban đầu thu được các phần dư e_i và giá trị ước lượng $\hat{Y}_i$

B2: Hồi qui MH hồi qui phụ :

$$e_i = [\beta_1 + \beta_2 X_i] + \alpha_1 \hat{Y}_i^2 + \dots + \alpha_m \hat{Y}_i^{m+1} + v_i (*)$$

$$\left\{ \begin{array}{l} H_0: a_1 = \dots = a_m = 0 \\ H_1: a_j \neq 0, j = 1, m \end{array} \right. \quad \begin{array}{l} \text{MH (1) có dạng hàm đúng} \\ \text{MH (1) có dạng hàm sai} \end{array}$$

Kiểm định χ^2 : $c_{qs}^2 = n R_{*}^2$, nếu $c_{qs}^2 > c_{\alpha}^2(p)$ thì bác bỏ H_0 .

Ví dụ: Trở lại thí dụ với tệp số liệu ch7bt4.

Kiểm định Ramsey cho kết quả sau:

Ramsey RESET Test:

F-statistic	0.348918	Probability	0.560248
Log likelihood ratio	0.389707	Probability	0.532453

Test Equation:

Dependent Variable: CONS

Method: Least Squares

Date: 11/19/08 Time: 22:18

Sample: 1960 1986

Included observations: 27

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-268.6193	746.5686	-0.359805	0.7221
GDP	0.954119	0.607571	1.570384	0.1294
FITTED^2	-0.000152	0.000257	-0.590693	0.5602
R-squared	0.798175	Mean dependent var	2037.449	
Adjusted R-squared	0.781356	S.D. dependent var	789.2231	
S.E. of regression	369.0360	Akaike info criterion	14.76410	
Sum squared resid	3268502.	Schwarz criterion	14.90809	
Log likelihood	-196.3154	F-statistic	47.45732	
Durbin-Watson stat	0.428978	Prob(F-statistic)	0.000000	

Kiểm định bằng nhân tử Lagrange cho kết quả sau:

Dependent Variable: E

Method: Least Squares

Date: 11/21/08 Time: 08:53

Sample: 1960 1986

Included observations: 27

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-423.8432	746.5686	-0.567722	0.5755
GDP	0.357051	0.607571	0.587669	0.5622
CONSF^2	-0.000152	0.000257	-0.590693	0.5602
R-squared	0.014330	Mean dependent var	-8.42E-14	
Adjusted R-squared	-0.067809	S.D. dependent var	357.1264	
S.E. of regression	369.0360	Akaike info criterion	14.76410	
Sum squared resid	3268502.	Schwarz criterion	14.90809	
Log likelihood	-196.3154	F-statistic	0.174459	
Durbin-Watson stat	0.428978	Prob(F-statistic)	0.840967	

Từ đó $c_{qs}^2 = n * R_s^2 = 27 * 0,01433 = 0,38691$

Mà $\chi^2(1) = 3,84146 \Rightarrow$ Mô hình chỉ định đúng.

2. Phân phối xác suất của sai số ngẫu nhiên

Các suy diễn thống kê (khoảng tin cậy, kiểm định giả thiết) phụ thuộc giả thiết SSNN phân phối chuẩn. Nếu SSNN không phân phối chuẩn thì các ước lượng vẫn là ước lượng tốt nhất, nhưng các phân tích không dùng được.

H_0 : SSNN phân phối chuẩn

H_1 : SSNN không phân phối chuẩn

Kiểm định Jarque – Bera:

Với S là hệ số bất đối xứng (*skewness*), K là hệ số nhọn (*kurtosis*) của e_i

$$JB = c_{qs}^2 = n \frac{\hat{S}^2}{6} + \frac{(K - 3)^2}{24}$$

Nếu $c_{qs}^2 > c_a^2(2)$ thì bác bỏ H_0

Ví dụ: Với tệp số liệu *ch7bt4* kiểm định Jarque-Bera cho kết quả sau:

Nếu mô hình không có khuyết tật nào thì các ước lượng là ước lượng tuyến tính, không chệch, tốt nhất; các kết quả hồi qui là đáng tin cậy và sử dụng để phân tích được.