



Giới thiệu một số vấn đề liên quan đến mô hình hồi quy

Bởi:

Phạm Trí Cao

Giới thiệu

Đa cộng tuyến

Bản chất của đa cộng tuyến

Đa cộng tuyến hoàn hảo: Các biến X_1, X_2, \dots, X_k được gọi là đa cộng tuyến hoàn hảo nếu tồn tại $\beta_1, \beta_2, \dots, \beta_k$ không đồng thời bằng không sao cho

$$\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k = 0 \quad (5.1)$$

Hiện tượng đa cộng tuyến hoàn hảo thường xảy ra do nhầm lẫn của nhà kinh tế lượng như trường hợp cái bẫy của biến giả mà chúng ta đã xem xét ở mục 4.7.3 chương 4.

Hiện tượng đa cộng tuyến mà chúng ta xét trong kinh tế lượng được hiểu với nghĩa rộng hơn đa cộng tuyến hoàn hảo như điều kiện (5.1). Các biến X_1, X_2, \dots, X_k được gọi là đa cộng tuyến không hoàn hảo nếu tồn tại $\beta_1, \beta_2, \dots, \beta_k$ sao cho

$$\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \beta = 0 \quad (5.2)$$

với β là sai số ngẫu nhiên.

Chúng ta có thể biểu diễn biến X_i theo các biến còn lại như sau

$$X_i = -\frac{\lambda_1}{\lambda_i} X_2 - \frac{\lambda_2}{\lambda_i} X_3 - \dots - \frac{\lambda_k}{\lambda_i} X_k - \frac{\epsilon}{\lambda_i}$$

$$\text{với } \beta_i \neq 0. \quad (5.3)$$

Vậy hiện tượng đa cộng tuyến xảy ra khi một biến là sự kết hợp tuyến tính của các biến còn lại và một nhiễu ngẫu nhiên.

Giới thiệu một số vấn đề liên quan đến mô hình hồi quy

Một số nguyên nhân gây ra hiện tượng đa cộng tuyến

Khi chọn các biến độc lập mỗi quan có quan hệ nhân quả hay có tương quan cao vì đồng phụ thuộc vào một điều kiện khác. Ví dụ số giường bệnh và số bác sĩ nếu đồng thời là biến độc lập của một hồi quy thì sẽ gây ra hiện tượng đa cộng tuyến gần hoàn hảo.

Khi số quan sát nhỏ hơn số biến độc lập. Một ví dụ điển hình là một nghiên cứu y khoa trên một số lượng nhỏ bệnh nhân nhưng lại khảo sát quá nhiều nhân tố tác động lên hiệu quả điều trị.

Cách thu thập mẫu. Ví dụ chỉ thu thập mẫu trên một số lớp giới hạn của tổng thể.

Chọn biến X_i có độ biến thiên nhỏ.

Hệ quả của đa cộng tuyến

Ví dụ 5.1

Ví dụ này lấy từ William E.Griffiths et al, Learning and Practicing Econometrics, John Wiley&Sons Inc, 1998, trang 433.

. Nghiên cứu của Klein và Golberger(1995) về quan hệ giữa tiêu dùng nội địa C_t , thu nhập từ lương W_t , thu nhập khác phi nông nghiệp P và thu nhập từ nông nghiệp A của nền kinh tế Hoa Kỳ từ năm 1928 đến 1950, với số liệu của các năm 1942 đến 1944 bị loại ra khỏi dữ liệu. Klein và Golberger thực hiện hồi quy tiêu dùng nội địa theo ba loại thu nhập như sau

$$C_t = \beta_1 + \beta_2 W_t + \beta_3 P_t + \beta_4 A_t + \beta_t(5.4)$$

Hồi quy này có thể gặp phải hiện tượng đa cộng tuyến vì các loại thu nhập có xu hướng cùng tăng theo sự phát triển của nền kinh tế.

Năm	C	W	P	A
1928	52,8	39,21	17,73	4,39
1929	62,2	42,31	20,29	4,60
1930	58,6	40,37	18,83	3,25
1931	56,6	39,15	17,44	2,61
1932	51,6	34,00	14,76	1,67
1933	51,1	33,59	13,39	2,44
1934	54	36,88	13,93	2,39
1935	57,2	39,27	14,67	5,00
1936	62,8	45,51	17,20	3,93
1937	65	46,06	17,15	5,48
1938	63,9	44,16	15,92	4,37
1939	67,5	47,68	17,59	4,51
1940	71,3	50,79	18,49	4,90
1941	76,6	57,78	19,18	6,37
1945	86,3	78,97	19,12	8,42
1946	95,7	73,54	19,76	9,27
1947	98,3	74,92	17,55	8,87
1948	100,3	74,01	19,17	9,30
1949	103,2	75,51	20,20	6,95
1950	108,9	80,97	22,12	7,15

Bảng 5.1. Số liệu thu nhập và tiêu dùng của nền kinh tế Hoa Kỳ

Kết quả hồi quy như sau

$$\hat{C} = 8,133 + 1,059W + 0,452P + 0,121A \quad (5.5)$$

$$t\text{-Stat}(0,91)(6,10)(0,69)(0,11)$$

$$\text{Khoảng } 95\%(-10,78;27,04)(0,69;1,73)(-0,94;1,84)(-2,18;2,43)$$

$$R^2 = 0,95 \quad F = 107,07 > F(3,16,99\%) = 5,29.$$

Mô hình này có tính giải thích cao thể hiện qua R^2 rất cao và thống kê F cao. Tuy nhiên một số hệ số lại không khác không với ý nghĩa thống kê thể hiện qua t-stat thấp, nghĩa

Giới thiệu một số vấn đề liên quan đến mô hình hồi quy

là ước lượng khoảng cho các hệ số chứa 0. W với hệ số có t-stat lớn thì ý nghĩa kinh tế lại rất lạ: nếu thu nhập từ lương tăng 1 USD thì tiêu dùng tăng 1,059 USD. Để tìm hiểu lý do gây ra hiện tượng trên chúng ta phải dùng lý thuyết của đại số ma trận, ở đây chỉ minh họa bằng mô hình hồi quy ba biến. Phương sai của ước lượng hệ số β_2 là

$$\text{var}(\hat{\beta}_2) = \frac{1}{\sum_{i=1}^n x_{2i}^2 (1 - r_{23}^2)} \sigma^2$$

Khi X_2 và X_3 có hiện tượng cộng tuyến thì

cao làm cho phương sai của ước lượng β_2 cao. Ước lượng β_2 theo phương pháp bình phương tối thiểu trở nên không hiệu quả.

Hệ quả của đa cộng tuyến

Ước lượng các hệ số không hiệu quả do phương sai của ước lượng lớn. Mô hình có đa cộng tuyến có t-stat nhỏ và một số hệ số của thể có dấu trái với lý thuyết hay có giá trị không phù hợp. R^2 thể hiện độ phù hợp của dữ liệu và F thể hiện ý nghĩa chung của các hệ số có thể rất cao.

Giá trị ước lượng của các hệ số rất nhạy cảm đối với việc tăng hoặc bớt một hoặc quan sát hoặc loại bỏ biến có mức ý nghĩa thấp.

Mặc dù việc phân tích tác động riêng phần của một biến khó khăn nhưng tính chính xác của dự báo có thể vẫn cao khi bản chất của đa cộng tuyến vẫn không đổi đối với quan sát mới.

Biện pháp khắc phục

Nếu mục tiêu của phân tích hồi quy là dự báo thì trong một số trường hợp chúng ta không cần khắc phục hiện tượng đa cộng tuyến.

Nếu mục tiêu của phân tích là xét tác động riêng phần của từng biến số lên biến phụ thuộc để quyết định chính sách thì đa cộng tuyến trở thành một vấn đề nghiêm trọng. Sau đây là một số biện pháp khắc phục.

Dùng thông tin tiên nghiệm. Ví dụ khi hồi quy hàm sản xuất Cobb-Douglas

$$\ln(Y_i) = \beta_1 + \beta_2 \ln(K_i) + \beta_3 \ln(L_i) + \beta_4 \ln(I_i) \quad (5.6)$$

Giới thiệu một số vấn đề liên quan đến mô hình hồi quy

Chúng ta có thể gặp hiện tượng đa cộng tuyến do K và L cùng tăng theo quy mô sản xuất. Nếu ta biết là hiệu suất không đổi theo quy mô thì ta có thêm thông tin $\beta_2 + \beta_3 = 1$. Với thông tin tiên nghiệm này chúng ta chuyển mô hình hồi quy (5.6) thành

$$\ln(Y_i) = \beta_1 + \beta_2 \ln(K_i) + (1 - \beta_2) \ln(L_i) + \beta_i \quad (5.7)$$

Bỏ đi một biến có đa cộng tuyến. Đây là cách làm đơn giản nhất. Ví dụ trong mô hình có biến giải thích là số bác sĩ và số giường bệnh thì ta có thể bỏ đi biến số giường bệnh. Nếu biến bị bỏ đi thực sự cần phải có trong mô hình thì chúng ta lại gặp phải một vấn đề khác, đó là ước lượng chệch đối với các hệ số còn lại. Vấn đề này chúng ta sẽ tiếp tục xem xét ở cuối chương.

Chuyển dạng dữ liệu

Giả sử chúng ta hồi quy trên dữ liệu chuỗi thời gian

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_t \quad (5.8)$$

Và chúng ta gặp phải hiện tượng đa cộng tuyến do X_{1t} và X_{3t} có thể cùng tăng hoặc giảm theo từng năm. Ta có thể tối thiểu tác động đa cộng tuyến này bằng kỹ thuật hồi quy trên sai phân bậc nhất như sau:

Ta có

$$Y_{t-1} = \beta_1 + \beta_2 X_{2,t-1} + \beta_3 X_{3,t-1} + \beta_{t-1} \quad (5.9)$$

Từ (5.8) và (5.9) ta xây dựng mô hình hồi quy

$$(Y_t - Y_{t-1}) = \beta_2 (X_{2t} - X_{2,t-1}) + \beta_3 (X_{3t} - X_{3,t-1}) + \beta_t \quad (5.10)$$

Với $\beta_t = \beta_t - \beta_{t-1}$.

Một vấn đề mới nảy sinh là β_t có thể có tính tương quan chuỗi, và như thế không tuân theo giả định của mô hình hồi quy tuyến tính cổ điển. Nếu hiện tượng tương quan chuỗi là nghiêm trọng thì mô hình (5.10) còn kém hơn cả mô hình (5.8).

Tăng thêm quan sát. Giải pháp này thích hợp cho hiện tượng đa cộng tuyến do cỡ mẫu nhỏ. Đôi khi chỉ cần tăng thêm một số quan sát là ta khắc phục được hiện tượng đa cộng tuyến. Một lần nữa chúng ta lại có sự đánh đổi. Tăng dữ liệu đôi khi đồng nghĩa với việc tăng chi phí, nhất là đối với dữ liệu sơ cấp. Mặt khác nếu là dữ liệu không có kiểm soát, chúng ta phải biết chắc rằng các điều kiện khác tương tự với khi ta thu thập dữ liệu gốc.

Khắc phục hiện tượng đa cộng tuyến đòi hỏi các kỹ thuật phức tạp và đôi khi cũng không mang lại hiệu quả như ta mong muốn. Mặt khác, hầu hết các mô hình hồi quy bội đều

Giới thiệu một số vấn đề liên quan đến mô hình hồi quy

có tính cộng tuyến nhất định nên chúng ta phải cẩn thận trong việc xây dựng mô hình và giải thích kết quả. Chúng ta sẽ nghiên cứu nguyên tắc xây dựng mô hình ở cuối chương.

Phương sai của sai số thay đổi - HETEROSKEDASTICITY

Bản chất của phương sai của sai số thay đổi

Giả định của mô hình hồi quy tuyến tính cổ điển là phương sai của sai số hồi quy không đổi qua các quan sát. Trong thực tế sai số hồi quy có thể tăng lên hoặc giảm đi khi giá trị biến độc lập X tăng lên. Tổng quát, thay cho giả định

$$\mathbf{E}(e_i^2) = \sigma^2$$

chúng ta giả định

$$\mathbf{E}(e_i^2) = \sigma_i^2$$

(5.11)

Thường gặp phương sai không đồng nhất ở dữ liệu chéo và dữ liệu bảng. Nguyên nhân phương sai không đồng nhất rất đa dạng, sau đây là một số trường hợp điển hình:

Gọi Y là số phế phẩm trong 100 sản phẩm của một thợ học việc, X là số giờ thực hành. Khi số giờ thực hành càng lớn thì số phế phẩm càng nhỏ và càng ít biến động. Chúng ta có trường hợp phương sai giảm dần khi X tăng dần.

Khi thu nhập(X) tăng thì chi tiêu cho các mặt hàng xa xỉ tăng và mức biến động càng lớn. Chúng ta có trường hợp phương sai tăng dần khi X tăng dần.

Khi cải thiện phương pháp thu thập số liệu thì phương sai giảm.

Phương sai của sai số tăng do sự xuất hiện của điểm nằm ngoài, đó là các trường hợp bất thường với dữ liệu rất khác biệt(rất lớn hoặc rất nhỏ so với các quan sát khác).

Phương sai thay đổi khi không xác định dạng mô hình, nếu một biến quan trọng bị bỏ sót thì phương sai của sai số lớn và thay đổi. Tình trạng này giảm hẳn khi đưa biến bị bỏ sót vào mô hình.

Hệ quả của phương sai thay đổi khi sử dụng ước lượng OLS

Xét hồi quy

$$Y_i = \beta_1 + \beta_2 X_i + e_i \quad (5.12)$$

Giới thiệu một số vấn đề liên quan đến mô hình hồi quy

với

Sử dụng phương pháp bình phương tối thiểu thông thường (OLS) chúng ta có

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2} = \beta_2 + \frac{\sum_{i=1}^n x_i \varepsilon_i}{\sum_{i=1}^n x_i^2} \quad (5.13)$$

$$E(\hat{\beta}_2) = \beta_2 + \frac{\sum_{i=1}^n x_i E(\varepsilon_i)}{\sum_{i=1}^n x_i^2} = \beta_2$$

vậy ước lượng theo OLS không chệch.

$$\text{var}(\hat{\beta}_2) = \frac{\sum_{i=1}^n x_i^2 \sigma_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2}$$

Chúng ta không chưa rõ là OLS có cho ước lượng hiệu quả hay không.

Ước lượng bình phương tối thiểu có trọng số (WLS)

Đặt

, chia hai vế của (5,12) cho w_i chúng ta có mô hình hồi quy

$$\frac{Y_i}{w_i} = \beta_1 \frac{1}{w_i} + \beta_2 \frac{X_i}{w_i} + \frac{\varepsilon_i}{w_i} \quad (5.14)$$

Ta viết lại mô hình (5.13) như sau

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \varepsilon_i^* \quad (5.15)$$

Mô hình (5.14) không có tung độ gốc và phương sai đồng nhất.

Giới thiệu một số vấn đề liên quan đến mô hình hồi quy

$$\text{var}(\hat{\varepsilon}_i) = \text{var}\left(\frac{\varepsilon_i}{w_i}\right) = \frac{w_i^2 \sigma^2}{w_i^2} = \sigma^2$$

Vậy ước lượng hệ số của (5.15) theo OLS là ước lượng hiệu quả (BLUE).

Kết quả ước lượng β_2 của (5.15) theo OLS như sau

$$\hat{\beta}_{2, \text{OLS}} = \frac{\sum_{i=1}^n \left(\frac{X_i Y_i}{w_i^2}\right) \sum_{i=1}^n \left(\frac{1}{w_i^2}\right) - \sum_{i=1}^n \left(\frac{Y_i}{w_i^2}\right) \sum_{i=1}^n \left(\frac{X_i}{w_i^2}\right)}{\sum_{i=1}^n \left(\frac{X_i^2}{w_i^2}\right) \sum_{i=1}^n \left(\frac{1}{w_i^2}\right) - \left(\sum_{i=1}^n \left(\frac{X_i}{w_i^2}\right)\right)^2}$$

(5.16)

Ước lượng (5.16) hoàn toàn khác với (5.13). Chúng ta biết ước lượng theo WLS (5.16) là ước lượng hiệu quả vậy ước lượng theo OLS (5.13) là không hiệu quả.

Phương sai đúng của hệ số ước lượng β_2 là

$$\text{var}(\hat{\beta}_2) = \frac{\sum_{i=1}^n x_i^2 \sigma_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2}$$

nhưng các phần mềm máy tính báo cáo phương sai là

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}$$

Từ phương sai của sai số bị tính sai này các trị thống kê t-stat và sai số chuẩn của hệ số ước lượng phần mềm cung cấp là vô dụng.

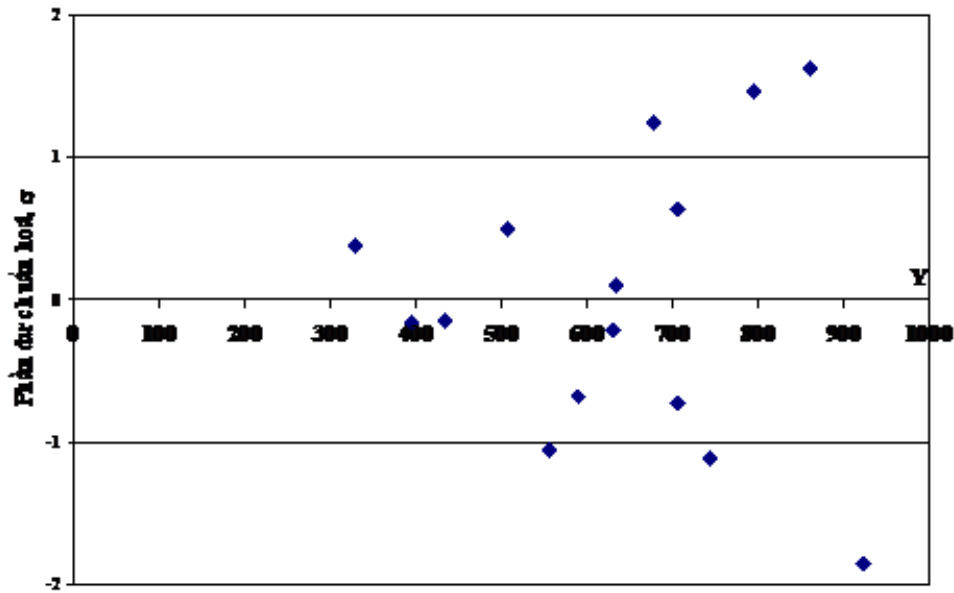
Tóm lại, với sự hiện diện của phương sai của sai số thay đổi mặc dù ước lượng các hệ số theo OLS vẫn không chệch nhưng ước lượng không hiệu quả và các trị thống kê như t-stat không chính xác.

Phát hiện và khắc phục

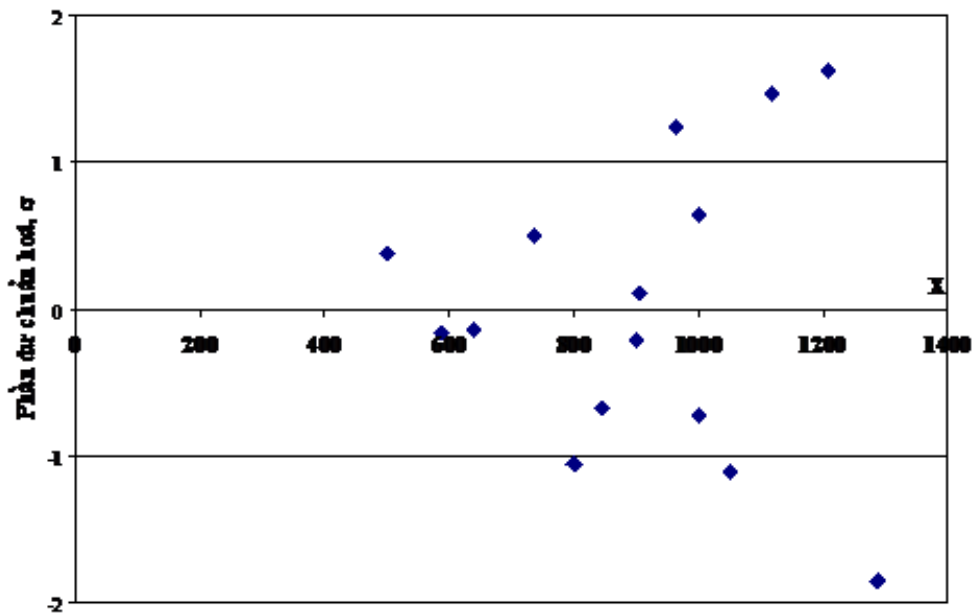
Phát hiện phương sai của sai số thay đổi.

Phương pháp đồ thị. Xét đồ thị của phần dư theo giá trị Y và X.

Giới thiệu một số vấn đề liên quan đến mô hình hồi quy



Hình 5.1. Đồ thị phân tán phần dư e_i theo \hat{Y}_i .



Hình 5.2. Đồ thị phân tán phần dư e_i theo X_i

Theo các đồ thị trên thì khi giá trị dự báo Y tăng (hoặc khi X tăng) thì phần dư có xu hướng tăng, hay mô hình có phương sai của sai số thay đổi.

Các phép thử chính thức

Xét hồi quy bội

Giới thiệu một số vấn đề liên quan đến mô hình hồi quy

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + \epsilon_i$$

(5.17)

Trong (k-1) biến độc lập trên ta trích ra (p-1) biến làm biến độc lập cho một hồi quy phụ. Trong hồi quy phụ này phần dư từ hồi quy mô hình (5.17) làm hồi quy biến phụ thuộc.

Các dạng hồi quy phụ thường sử dụng là

$$e_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \dots + \alpha_p Z_{pi} + \delta_i \quad (5.18)$$

$$|e_i| = \alpha_1 + \alpha_2 Z_{2i} + \dots + \alpha_p Z_{pi} + \delta_i \quad (5.19)$$

$$\ln(e_i^2) = \alpha_1 + \alpha_2 Z_{2i} + \dots + \alpha_p Z_{pi} + \delta_i \quad (5.20)$$

Kiểm định Breusch-Pagan căn cứ vào hồi quy phụ (5.18), kiểm định Glejser căn cứ vào (5.19) và kiểm định Harvey-Godfrey căn cứ vào (5.20).

Giả thiết không là không có phương sai không đồng nhất

$$H_0 : \beta_2 = \beta_3 = \dots = \beta_p = 0$$

H_1 : Không phải tất cả các hệ số trên đều bằng 0.

R^2 xác định từ hồi quy phụ, n là cỡ mẫu dùng để xây dựng hồi quy phụ, với cỡ mẫu lớn thì nR^2 tuân theo phân phối Chi bình phương với (p-1) bậc tự do.

Quy tắc quyết định

Nếu $\chi_{(p-1, 1-\alpha)}^2 \leq nR^2$ thì bác bỏ H_0 .

Nếu bác bỏ được H_0 thì chúng ta chấp nhận mô hình có phương sai của sai số thay đổi và thực hiện kỹ thuật ước lượng mô hình như sau:

Đối với kiểm định Breusch-Pagan

$$\hat{w}_i^2 = \hat{\alpha}_1 + \hat{\alpha}_2 Z_{2i} + \dots + \hat{\alpha}_p Z_{pi}$$

Đối với kiểm định Glejser

$$\hat{w}_i^2 = (\hat{\alpha}_1 + \hat{\alpha}_2 Z_{2i} + \dots + \hat{\alpha}_p Z_{pi})^2$$

Đối với kiểm định Harvey-Godfrey

Giới thiệu một số vấn đề liên quan đến mô hình hồi quy

$$\hat{w}_i^2 = \exp(\hat{\alpha}_1 + \hat{\alpha}_2 Z_{2i} + \dots + \hat{\alpha}_p Z_{pi})$$

Ta có $\hat{w}_i = \sqrt{\hat{w}_i^2}$. Đến đây chúng ta có thể chuyển dạng hồi quy theo OLS thông thường sang hồi quy theo bình phương tối thiểu có trọng số WLS.

Tự tương quan (tương quan chuỗi)

Trong mô hình hồi quy tuyến tính cổ điển chúng ta giả định không có tương quan giữa các phần dư hay $E(\epsilon_i \epsilon_j) = 0$ với mọi i, j .

Trong thực tế đối với dữ liệu chuỗi thời gian, giả định này hay bị vi phạm. Một lý do nôm na là biến số kinh tế có một quán tính (sức ỳ) nhất định. Ví dụ sự tăng cầu một loại hàng hóa của năm nay sẽ làm tăng lượng cung nội địa của hàng hoá đó vào năm sau, đây là tác động trễ của biến độc lập hay biến phụ thuộc thời kỳ t chịu tác động của biến độc lập ở thời kỳ $t-1$.

Đôi khi nền kinh tế lại phản ứng quá nhạy với sự thay đổi. Ví dụ giá mía cao ở năm nay sẽ làm cho nông dân đổ xô trồng mía, sản lượng mía năm sau tăng vọt làm giảm giá mía ở năm sau, đây là tác động trễ của biến phụ thuộc hay giá trị biến phụ thuộc thời kỳ t chịu ảnh hưởng của giá trị biến phụ thuộc thời kỳ $t-1$.

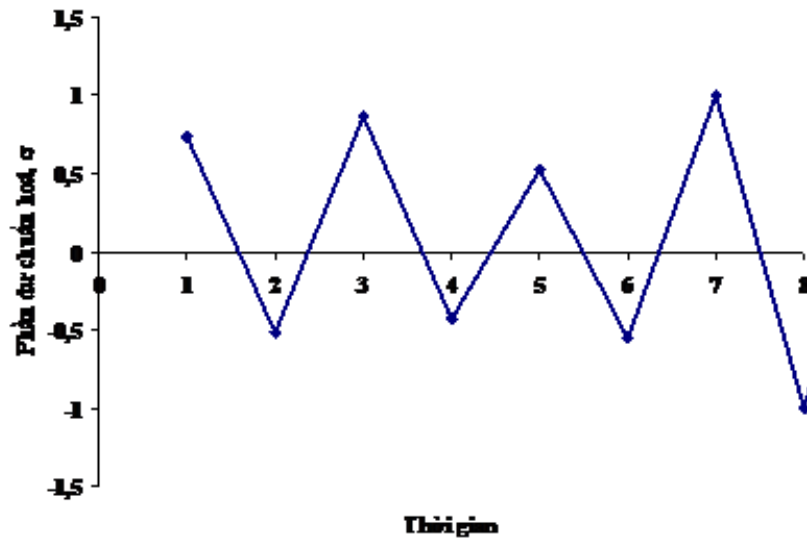
Hiện tượng tự tương quan làm cho $E(\epsilon_i \epsilon_j) \neq 0$ và gây ra các hậu quả sau

Ước lượng theo OLS không chệch nhưng không hiệu quả

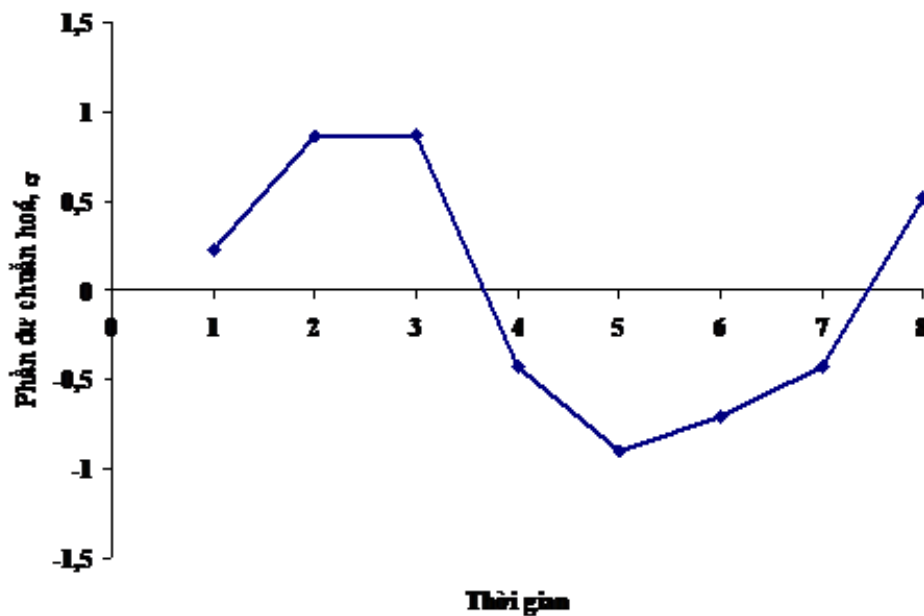
Các trị thống kê tính theo OLS không hữu ích trong việc nhận định mô hình.

Chúng ta có thể phát hiện hiện tượng tự tương quan bằng cách quan sát đồ thị phần dư của mô hình trên dữ liệu chuỗi thời gian.

Giới thiệu một số vấn đề liên quan đến mô hình hồi quy



Hình 5.3. Tương quan chuỗi nghịch



Hình 5.4. Tương quan chuỗi thuận

Chúng ta sẽ tiếp tục làm việc với dữ liệu chuỗi và xử lý hiện tượng tự tương quan ở phần sau của giáo trình liên quan đến các mô hình dự báo.

Lựa chọn mô hình

Một yếu tố quan trọng đầu tiên để chọn đúng mô hình hồi quy là chọn đúng dạng hàm. Để chọn đúng dạng hàm chúng ta phải hiểu ý nghĩa và mối quan hệ kinh tế của các biến

Giới thiệu một số vấn đề liên quan đến mô hình hồi quy

số. Ý nghĩa của một số loại hàm thông dụng đã được trình bày ở mục 3.8.2 chương 3. Ở phần này chúng ta xét hậu quả của một số dạng xây dựng mô hình sai và chiến lược xây dựng mô hình kinh tế lượng. Chúng ta cũng không đi sâu vào chứng minh các kết quả.

Thiếu biến có liên quan và chứa biến không liên quan.

Xét hai hồi quy sau

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_K X_{Ki} + \epsilon_i \quad (5.21)$$

và

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_K X_{Ki} + \beta_{(K+1)} X_{K+1,i} + \dots + \beta_{(K+L)} X_{K+L,i} + \epsilon_i \quad (5.22)$$

Mô hình (5.21) có các trị thông kê tương ứng có ký hiệu R và mô hình (5.22) có các trị thống kê tương ứng có ký hiệu U.

Có hai trường hợp xảy ra:

Trường hợp 1: Nếu mô hình (5.22) là đúng nhưng chúng ta chọn mô hình (5.21) nghĩa là chúng ta bỏ sót L biến quan trọng (X_{K+1}, \dots, X_{K+L}). Hậu quả là ước lượng các hệ số cho K-1 biến độc lập còn lại bị chệch, mô hình kém tính giải thích cho cả mục tiêu dự báo và phân tích chính sách.

Trường hợp 2: Nếu mô hình (5.21) là đúng nhưng chúng ta chọn mô hình (5.22), nghĩa là chúng ta đưa vào mô hình các biến không liên quan. Hậu quả là ước lượng hệ số cho các biến quan trọng vẫn không chệch nhưng không hiệu quả.

Kiểm định so sánh mô hình (5.21) và (5.22) - Kiểm định Wald

Chúng ta muốn kiểm định xem L biến (X_{K+1}, \dots, X_{K+L}) có đáng được đưa vào mô hình hay không.

H_0 :

$$\beta_{K+1} = \beta_{K+2} = \dots = \beta_{K+L} = 0$$

Trị thống kê

Giới thiệu một số vấn đề liên quan đến mô hình hồi quy

$$\frac{(RSS_R - RSS_U)/L}{RSS_U/(n - K - L)} \sim F^* \sim F_{(L, n-K-L)}$$

Quy tắc quyết định: Nếu

$$F^* > F_{(L, n-K-L), 1-\alpha}$$

thì ta bác bỏ H_0 hay chấp nhận L biến $(X_{K+1}, \dots, X_{K+L})$ xứng đáng được đưa vào mô hình.

Hai chiến lược xây dựng mô hình

Có hai chiến lược xây dựng mô hình kinh tế lượng là:

Xây dựng mô hình từ đơn giản đến tổng quát: chứa tất cả các biến có liên quan trong mô hình và loại bỏ dần những biến ít ý nghĩa thống kê nhất cho đến khi nhận được mô hình “tốt nhất”.

Xây dựng mô hình tổng quát đến đơn giản : Xuất phát từ biến độc lập có quan hệ kinh tế trực tiếp nhất với biến phụ thuộc, tiếp tục bổ sung biến mới cho đến khi nhận được mô hình “tốt nhất”.

Mỗi cách làm đều có những ưu và nhược điểm. Hiện nay với công cụ máy vi tính, người ta không còn ngại tính toán trên mô hình lớn và nhiều nhà kinh tế lượng cho rằng xây dựng mô hình từ tổng quát đến đơn giản thì hiệu quả hơn từ đơn giản đến tổng quát. Nét chung của cả hai chiến lược này là ở từng bước đều phải thực hiện kiểm định Wald.