



Chương V – Đa cộng tuyến



Chương V – Đa cộng tuyến



- 1. Đa cộng tuyến – bản chất và nguyên nhân**
- 2. Hậu quả**
- 3. Phát hiện**
- 4. Khắc phục**



Chương V – Đa cộng tuyến



1. Đa cộng tuyến – bản chất và nguyên nhân

$$PRM : Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + U_i$$

Giả thiết OLS: Các biến độc lập không có tương quan tuyến tính với nhau.

Giả thiết bị vi phạm → khuyết tật Đa cộng tuyến

(*) Nguyên nhân:

- Mô hình có các biến không cần thiết hoặc có thông tin trùng lặp (bài tập 5.4: hồi qui **lượng bán** của 1 hãng phụ thuộc vào **giá** và **lượng bán của hãng cạnh tranh**, ...)
- Các biến KT – XH trong cùng 1 lĩnh vực thường có quan hệ chặt chẽ với nhau (hồi qui **sản lượng** phụ thuộc vào **vốn** và **lao động**, **GDP** phụ thuộc vào **IM** và **EX**, ...)



Chương V – Đa cộng tuyến



1. Đa cộng tuyến – bản chất và nguyên nhân

(*) Phân loại:

- Đa cộng tuyến hoàn hảo (*perfect multicollinearity*): quan hệ giữa các biến độc lập là quan hệ hàm số tuyến tính:

$$m_2 \times X_{2i} + m_3 \times X_{3i} + \dots + m_k \times X_{ki} = 0$$

$$\exists m_j \neq 0$$

- Đa cộng tuyến không hoàn hảo (*imperfect multicollinearity*)
Quan hệ giữa các biến độc lập là quan hệ hồi qui tuyến tính

$$m_2 \times X_{2i} + m_3 \times X_{3i} + \dots + m_k \times X_{ki} + V_i = 0$$

trong đó V_i là một sai số ngẫu nhiên



Chương V – Đa cộng tuyến



1. Đa cộng tuyến – bản chất và nguyên nhân

(*) Ước lượng OLS khi có Đa cộng tuyến:

- Khi có đa cộng tuyến hoàn hảo (*perfect multicollinearity*): không thể ước lượng được các hệ số hồi qui cũng như không xác định được SRF
- Khi có đa cộng tuyến không hoàn hảo (*imperfect multicollinearity*): vẫn có thể ước lượng được các hệ số hồi qui và xác định SRF 1 cách duy nhất, tuy nhiên sẽ dẫn đến 1 số hậu quả trong phân tích hồi qui. Đặc biệt khi mức độ cộng tuyến của các biến độc lập cao



Chương V – Đa cộng tuyến



2. Hậu quả:

- Các ước lượng vẫn là BLUE, tuy nhiên phương sai và hiệp phương sai của chúng tăng lên → các ước lượng kém chính xác
- Khoảng tin cậy của các hệ số hồi qui rộng hơn thực tế
- Các kiểm định T mất ý nghĩa, đặc biệt là T- statistic của các hệ số góc thường mất ý nghĩa thống kê → dẫn tới kết luận tồn tại các biến độc lập không cần thiết có mặt trong mô hình
- R^2 lại tăng lên đáng kể
- Ước lượng OLS và các sai số chuẩn rất nhạy với sự thay đổi nhỏ trong số liệu
- Dấu của các ước lượng có thể bị sai (ví dụ: bài tập 5.4)



Chương V – Đa cộng tuyến



2. Hậu quả:

- Nguyên nhân phương sai và hiệp phương sai của các ước lượng tăng lên: Với hàm hồi qui 3 biến:

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^n x_{2i}^2 (1 - r_{23}^2)} \quad \text{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum_{i=1}^n x_{3i}^2 (1 - r_{23}^2)}$$

$$\text{cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23} \sigma^2}{(1 - r_{23}^2) \sqrt{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{3i}^2}}$$

$$VIF = \frac{1}{(1 - r_{23}^2)} \quad (\text{variance inflating factor}) \quad TOL_j = \frac{1}{VIF_j} \quad (\text{tolerance})$$



Chương V – Đa cộng tuyến



2. Hậu quả:

- Ví dụ: Hồi qui chi tiêu hộ theo thu nhập và tài sản của hộ

Y – Chi tiêu hộ

X2 – Thu nhập hộ

X3 – Tài sản của hộ

Y, \$	X2, \$	X3, \$
70	80	810
65	100	1009
90	120	1273
95	140	1425
110	160	1633
115	180	1876
120	200	2052
140	220	2201
155	240	2435
150	260	2686



Chương V – Đa cộng tuyến



2. Hậu quả:

Dependent Variable: Y

Method: Least Squares

Sample: 1 10

Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	24.77473	6.752500	3.668972	0.0080
X2	0.941537	0.822898	1.144172	0.2902
X3	-0.042435	0.080664	-0.526062	0.6151
R-squared	0.963504	Mean dependent var		111.0000
Adjusted R-squared	0.953077	S.D. dependent var		31.42893
Log likelihood	-31.58705	F-statistic		92.40196
Durbin-Watson stat	2.890614	Prob(F-statistic)		0.000009



Chương V – Đa cộng tuyến



3. Phát hiện:

3.1. R^2 cao nhưng các tỉ số T không có ý nghĩa

$R^2 = 0,8$ trở lên \rightarrow thường cho kết luận bác bỏ H_0 khi kiểm định sự phù hợp của hàm hồi qui. Tuy nhiên các tỉ số T lại cho thấy hầu hết các hệ số hồi qui không có ý nghĩa

3.2. Hệ số tương quan giữa các biến độc lập

Trường hợp hồi qui chính chỉ có 2 biến độc lập, ta có thể dùng hệ số tương quan giữa các biến độc lập này để kết luận về hiện tượng đa cộng tuyến ($> 0,8$)

3.3. Nhân tử phóng đại phương sai

$VIF > 10$ hoặc TOL càng gần với 0 thì mức độ cộng tuyến giữa các biến độc lập càng cao \rightarrow ĐCT nghiêm trọng



Chương V – Đa cộng tuyến



3. Phát hiện:

3.4. Hồi qui phụ (auxiliary regressions):

Hồi qui của 1 biến độc lập theo các biến độc lập còn lại trong mô hình

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

Hồi qui phụ:

$$X_{2i} = m_1 + m_2 X_{3i} + \dots + m_{k-1} X_{ki} + V_i \quad (2)$$

$$X_{2i} = m_1 + m_2 X_{3i} + V_i \quad (3)$$

Về kỹ thuật, có thể chọn bất cứ biến độc lập nào để đóng vai trò biến phụ thuộc trong hồi qui phụ, tuy nhiên trên thực tế, vấn đề này khá nhạy cảm và còn phụ thuộc vào kinh nghiệm của người nghiên cứu.



Chương V – Đa cộng tuyến



3. Phát hiện:

3.4. Hồi qui phụ (auxiliary regressions):

(*) Sử dụng (2) để kiểm tra (1)

$$X_{2i} = m_1 + m_2 X_{3i} + \dots + m_{k-1} X_{ki} + V_i \quad (2)$$

$$\begin{cases} H_0: R_2^2 = 0 \\ H_1: R_2^2 \neq 0 \end{cases} \rightarrow \begin{cases} H_0: (1) \text{ không có Đa cộng tuyến} \\ H_1: (1) \text{ có Đa cộng tuyến} \end{cases}$$

Tiêu chuẩn kiểm định:

$$F_{qs} = \frac{\frac{R_2^2}{(k-2)}}{\frac{(1-R_2^2)}{(n-k+1)}}$$

Miền bác bỏ H_0 : $W_\alpha = F : F > F_\alpha^{(k-2, n-k+1)}$



Chương V – Đa cộng tuyến



3. Phát hiện:

3.4. Hồi qui phụ (auxiliary regressions):

(*) Sử dụng (3) để kiểm tra (1)

$$X_{2i} = m_1 + m_2 X_{3i} + V_i \quad (3)$$

$$\begin{cases} H_0: R_3^2 = 0 \\ H_1: R_3^2 \neq 0 \end{cases} \Leftrightarrow \begin{cases} m_2 = 0 \\ m_2 \neq 0 \end{cases} \rightarrow \begin{cases} H_0: (1) \text{ không có Đa cộng tuyến} \\ H_1: (1) \text{ có Đa cộng tuyến} \end{cases}$$

$$F_{qs} = \frac{\frac{R_3^2}{(2-1)}}{\frac{(1-R_3^2)}{(n-2)}}$$

$$T_{qs} = \frac{\hat{m}_2}{S.E(\hat{m}_2)}$$

$$W_\alpha = F : F > F_{\alpha}^{(1, n-2)}$$

$$W_\alpha = T : |T| > T_{\alpha/2}^{(n-2)}$$



Chương V – Đa cộng tuyến



3. Phát hiện:

3.4. Hồi qui phụ (auxiliary regressions):

Một cách kiểm tra khác từ hồi qui phụ mà không cần sử dụng các kiểm định nói trên:

R^2 của hồi qui phụ $>$ R^2 của hồi qui chính. Khi đó có thể kết luận hồi qui chính có hiện tượng đa cộng tuyến nghiêm trọng. (qui tắc Lawrence R. Klien – *Introduction to Econometrics* - 1962).



Chương V – Đa cộng tuyến



4. Khắc phục:

4.1. Sử dụng thông tin tiên nghiệm (a priori information)

$$\ln(Y_i) = \beta_1 + \beta_2 \ln(K_i) + \beta_3 \ln(L_i) + U_i \quad (1)$$

Với thông tin cho trước: quá trình sản xuất có hiệu quả không đổi theo qui mô, $\beta_2 + \beta_3 = 1 \rightarrow \beta_3 = 1 - \beta_2$

Mô hình (1) trở thành:

$$\ln\left(\frac{Y_i}{L_i}\right) = \beta_1 + \beta_2 \ln\left(\frac{K_i}{L_i}\right) + U_i \quad (*)$$

4.2. Bỏ bớt biến độc lập là nguyên nhân gây ra đa cộng tuyến

4.3. Thu thập thêm các quan sát mới



Chương V – Đa cộng tuyến



4. Khắc phục:

4.4. Sử dụng sai phân cấp 1

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i \quad (1)$$

(1) Được biến đổi thành:

$$Y_i - Y_{i-1} = \beta_2 (X_{2i} - X_{2i-1}) + \beta_3 (X_{3i} - X_{3i-1}) + V_i (*)$$

4.5. Sử dụng hồi qui đa thức

(1) Được biến đổi thành:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}^2 + U_i \quad (1)$$