

KINH TẾ LƯỢNG ECONOMETRICS

HOÀNG KHẮC LỊCH
Ph.D. in Economics
 Faculty of Development Economics
 VNU University of Economics and Business
 Email: hoangkhaclich@gmail.com
 Cell phone: +84 978135777

ĐA CỘNG TUYẾN

- Bản chất và nguyên nhân
- Ước lượng các tham số
- Hậu quả
- Phát hiện đa cộng tuyến
- Khắc phục đa cộng tuyến

7/1/2014

TS. Hoàng Khắc Lịch

2

BẢN CHẤT CỦA ĐA CỘNG TUYẾN

- Trong mô hình hồi quy bội
- $$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki}$$
- Có sự phụ thuộc tuyến tính cao giữa các biến giải thích

7/1/2014

TS. Hoàng Khắc Lịch

3

BẢN CHẤT CỦA ĐA CỘNG TUYẾN

- Đa cộng tuyến hoàn hảo
Tồn tại $\lambda_2, \lambda_3, \dots, \lambda_k$ không đồng thời bằng 0 sao cho

$$\lambda_2 X_2 + \lambda_3 X_3 + \dots + \lambda_k X_k = 0$$

- Đa cộng tuyến không hoàn hảo

$$\lambda_2 X_2 + \lambda_3 X_3 + \dots + \lambda_k X_k + v_i = 0$$

với v_i là sai số ngẫu nhiên.

7/1/2014

TS. Hoàng Khắc Lịch

4

BẢN CHẤT CỦA ĐA CỘNG TUYẾN

X_2	10	15	18	24	30
X_3	50	75	90	120	150
X_4	52	75	97	129	152
V	2	0	7	9	2

$X_{3i} = 5X_{2i}$, có cộng tuyến hoàn hảo giữa X_2 và X_3 , $r_{23} = 1$

X_2 và X_4 có cộng tuyến không hoàn hảo

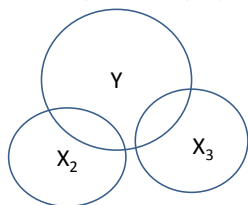
7/1/2014

TS. Hoàng Khắc Lịch

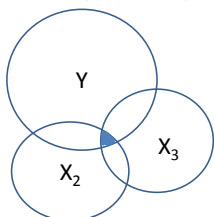
5

BẢN CHẤT CỦA ĐA CỘNG TUYẾN

Không có đa cộng tuyến



Đa cộng tuyến thấp



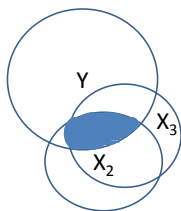
7/1/2014

TS. Hoàng Khắc Lịch

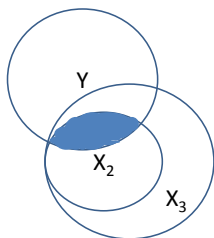
6

BẢN CHẤT CỦA ĐA CỘNG TUYẾN

Đa cộng tuyến cao



Đa cộng tuyến hoàn hảo



7/1/2014

TS. Hoàng Khắc Lịch

7

NGUYÊN NHÂN CỦA ĐA CỘNG TUYẾN

- Chọn các biến độc lập có quan hệ nhân quả hay có tương quan cao vì đồng phụ thuộc vào một điều kiện khác.
- Số quan sát nhỏ hơn số biến độc lập.
- Cách thu thập mẫu: mẫu không đặc trưng cho tổng thể
- Chọn biến X_i có độ biến thiên nhỏ.

7/1/2014

TS. Hoàng Khắc Lịch

8

ƯỚC LƯỢNG KHI CÓ ĐA CỘNG TUYẾN

1. Trường hợp có đa cộng tuyến hoàn hảo

Xét mô hình hồi qui 3 biến dưới dạng sau:

$$Y_i = \beta_2 X_{2i} + \beta_3 X_{3i} + e_i$$

giả sử $X_{3i} = \lambda X_{2i}$, mô hình được biến đổi thành:

$$Y_i = (\beta_2 + \lambda\beta_3)X_{2i} + e_i = \beta_0 X_{2i} + e_i$$

Phương pháp OLS:

$$\hat{\beta}_0 = (\hat{\beta}_2 + \lambda\hat{\beta}_3) = \frac{\sum X_{2i} Y_i}{\sum X_{2i}^2}$$

- Không thể tìm được lời giải duy nhất cho $\hat{\beta}_2, \hat{\beta}_3$

7/1/2014

TS. Hoàng Khắc Lịch

9

ƯỚC LƯỢNG KHI CÓ ĐA CỘNG TUYẾN

$$\hat{\beta}_2 = \frac{\sum y_i x_{2i} \sum x_{3i}^2 - \sum y_i x_{3i} \sum x_{2i} x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i} x_{3i})^2}$$

$$\hat{\beta}_2 = \frac{\lambda \sum y_i x_{3i} \sum x_{3i}^2 - \lambda \sum y_i x_{3i} \sum x_{3i} x_{3i}}{\lambda^2 \sum x_{3i}^2 \sum x_{3i}^2 - \lambda^2 \sum x_{3i}^2 \sum x_{3i}^2} = \frac{0}{0}$$

- ☐ Các hệ số ước lượng không xác định
- ☐ Phương sai và sai số chuẩn của β_2 và β_3 là vô hạn

7/1/2014

TS. Hoàng Khắc Lịch

10

ƯỚC LƯỢNG KHI CÓ ĐA CỘNG TUYẾN

Trường hợp có đa cộng tuyến không hoàn hảo

- Đa cộng tuyến hoàn hảo thường không xảy ra trong thực tế.
- Xét mô hình hồi qui 3 biến dưới dạng sau:

$$y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + e_i$$

$$\text{Giả sử } x_{3i} = \lambda x_{2i} + v_i$$

Với $\lambda \neq 0$ và v_i là sai số ngẫu nhiên

7/1/2014

TS. Hoàng Khắc Lịch

11

ƯỚC LƯỢNG KHI CÓ ĐA CỘNG TUYẾN

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\lambda^2 \sum x_{2i}^2 + \sum v_i^2) - (\lambda^2 \sum y_i x_{2i} + \sum y_i v_i)(\lambda \sum x_{2i}^2)}{(\sum x_{2i}^2)(\lambda^2 \sum x_{2i}^2 + \sum v_i^2) - (\lambda^2 \sum x_{2i}^2)^2}$$

- ☐ Có thể ước lượng được các hệ số hồi quy nhưng sai số chuẩn rất lớn.

7/1/2014

TS. Hoàng Khắc Lịch

12

HẬU QUẢ CỦA ĐA CỘNG TUYẾN

Nếu có cộng tuyến gần hoàn hảo

1. Phương sai và hiệp phương sai của các ước lượng OLS lớn.
2. Khoảng tin cậy rộng hơn.
3. Tỷ số t "không có ý nghĩa"
4. R^2 cao nhưng tỷ số t ít có ý nghĩa

7/1/2014

TS. Hoàng Khắc Lịch

13

HẬU QUẢ CỦA ĐA CỘNG TUYẾN

5. Các ước lượng OLS và sai số chuẩn của chúng trở nên rất nhạy với những thay đổi nhỏ trong dữ liệu.
6. Dấu của các ước lượng của các hệ số hồi qui có thể sai
7. Thêm vào hay bớt đi các biến cộng tuyến với các biến khác, mô hình sẽ thay đổi về dấu hoặc thay đổi về độ lớn của các ước lượng.

7/1/2014

TS. Hoàng Khắc Lịch

14

HẬU QUẢ CỦA ĐA CỘNG TUYẾN

- Đa cộng tuyến là một hiện tượng theo mẫu, nghĩa là cho dù các biến độc lập X_i không tương quan tuyến tính trong tổng thể nhưng chúng có thể tương quan tuyến tính trong một mẫu cụ thể nào đó.
- Do đó, cỡ mẫu lớn thì hiện tượng đa cộng tuyến ít nghiêm trọng hơn cỡ mẫu nhỏ

7/1/2014

TS. Hoàng Khắc Lịch

15

CÁCH PHÁT HIỆN ĐA CỘNG TUYẾN

1. Hệ số R^2 lớn nhưng tỷ số t nhỏ
2. Tương quan cặp giữa các biến giải thích cao
3. Sử dụng mô hình hồi qui phụ
4. Sử dụng yếu tố phóng đại phương sai (VIF)

7/1/2014

TS. Hoàng Khắc Lịch

16

CÁCH PHÁT HIỆN ĐA CỘNG TUYẾN

- R^2 lớn nhưng tỷ số t nhỏ
- Tương quan cặp giữa các biến giải thích cao

$$r_{xz} = \frac{\sum (X_i - \bar{X})(Z_i - \bar{Z})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Z_i - \bar{Z})^2}}$$

- Trong đó X, Z là 2 biến giải thích trong mô hình

7/1/2014

TS. Hoàng Khắc Lịch

17

CÁCH PHÁT HIỆN ĐA CỘNG TUYẾN

- **Sử dụng mô hình hồi qui phụ**
- Hồi qui một biến giải thích X theo các biến còn lại

$$\hat{X}_{2i} = \hat{\beta}_1 + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki}$$

Tính R^2 và F cho mỗi mô hình

$$F = \frac{R^2(n-m)}{(1-R^2)(m-1)}$$

- Lập giả thiết $H_0: R^2 = 0 \sim H_0$: không có đa cộng tuyến
- Nếu $F > F_{\alpha}(m-1, n-k)$: bác bỏ H_0 hay có đa cộng tuyến
- Nếu $F < F_{\alpha}(m-1, n-k)$: chấp nhận H_0 hay không có đa cộng tuyến

7/1/2014

TS. Hoàng Khắc Lịch

18

CÁCH PHÁT HIỆN ĐA CỘNG TUYẾN

- **Sử dụng nhân tử phóng đại phương sai (VIF)**

- Đối với hàm hồi quy 2 biến giải thích

$$VIF = \frac{1}{(1 - r_{23}^2)}$$

- Đối với trường hợp tổng quát, có (k-1) biến giải thích

$$VIF = \frac{1}{(1 - R_j^2)}$$

- ❖ R_j^2 : là giá trị R^2 trong hàm hồi quy của X_j theo (k-2) biến giải thích còn lại.
- ❖ Thông thường khi $VIF > 10$, thì biến này được coi là có cộng tuyến cao

7/1/2014

TS. Hoàng Khắc Lịch

19

CÁCH KHẮC PHỤC

- **Dùng thông tin tiên nghiệm**

- Ví dụ mô hình sản xuất Cobb-Douglas

$$\ln(Y_i) = \beta_1 + \beta_2 \ln(K_i) + \beta_3 \ln(L_i) + u_i$$

- Có thể xảy ra đa cộng tuyến do K và L cùng tăng theo quy mô sản xuất. Nếu biết hiệu suất không đổi theo quy mô tức là $\beta_2 + \beta_3 = 1$ thì

$$\ln(Y_i) = \beta_1 + \beta_2 \ln(K_i) + (1 - \beta_2) \ln(L_i) + u_i$$

$$\ln(Y_i) - \ln(L_i) = \beta_1 + \beta_2 [\ln(K_i) - \ln(L_i)] + u_i$$

$$\ln(Y_i / L_i) = \beta_1 + \beta_2 \ln(K_i / L_i) + u_i$$

→ mất đa cộng tuyến (vì đây là mô hình hồi quy đơn)

7/1/2014

TS. Hoàng Khắc Lịch

20

CÁCH KHẮC PHỤC

- **Loại trừ một biến giải thích ra khỏi mô hình**

- B1: Xem cặp biến giải thích nào có quan hệ chặt chẽ. Giả sử $X_2, X_3, \dots, X_k$ là các biến độc lập, Y là biến phụ thuộc và X_2, X_3 có tương quan chặt chẽ với nhau.
- B2: Tính R^2 đối với các hàm hồi quy: có mặt cả 2 biến; không có mặt một trong 2 biến
- B3: Loại biến mà giá trị R^2 tính được khi không có mặt biến đó là lớn hơn.

7/1/2014

TS. Hoàng Khắc Lịch

21

CÁCH KHẮC PHỤC

- Bổ sung thêm dữ liệu hoặc chọn mẫu mới

7/1/2014

TS. Hoàng Khắc Lịch

22

CÁCH KHẮC PHỤC

- Dùng sai phân cấp 1

- Có hàm hồi qui: $y_t = \alpha_1 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$

suy ra

$$y_{t-1} = \alpha_1 + \beta_1 x_{1,t-1} + \beta_2 x_{2,t-1} + u_{t-1}$$

- Trừ hai vế cho nhau, được:

$$y_t - y_{t-1} = \beta_1 (x_{1,t} - x_{1,t-1}) + \beta_2 (x_{2,t} - x_{2,t-1}) + (u_t - u_{t-1})$$

- Hay:

$$\Delta y_t = \beta_1 \Delta x_{1,t} + \beta_2 \Delta x_{2,t} + e_t$$

7/1/2014

TS. Hoàng Khắc Lịch

23