

CHƯƠNG 5.

ĐA CỘNG TUYẾN

Các vấn đề cần xem xét

- Định nghĩa loại khuyết tật của mô hình (Mô hình vi phạm giả thiết nào của phương pháp OLS)
- Hậu quả của khuyết tật đối với các ước lượng OLS
- Nguyên nhân của khuyết tật
- Cách phát hiện
- Giải pháp khắc phục

Định nghĩa

$$Y_i = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k + u_i$$

- **Đa cộng tuyến hoàn hảo:**

$$\exists \lambda_1 \neq 0: \quad \lambda_2 X_{2i} + \lambda_3 X_{3i} + \dots + \lambda_k X_{ki} = 0$$

- **Đa cộng tuyến không hoàn hảo:**

$$\exists \lambda_1 \neq 0: \quad \lambda_2 X_{2i} + \lambda_3 X_{3i} + \dots + \lambda_k X_{ki} + v_i = 0$$

v_i là yếu tố ngẫu nhiên

Đa cộng tuyến hoàn hảo

- Tồn tại mối quan hệ tuyến tính giữa các biến độc lập trong mô hình
- Vi phạm giả thiết 6 của phương pháp OLS
- Nguyên nhân: Do thừa biến (ví dụ: trong mô hình bao gồm cả 2 biến giả nam và nữ)

Đa cộng tuyến hoàn hảo

$$y_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + e_i$$

$$X_{3i} = \lambda X_{2i} \Rightarrow x_{3i} = \lambda x_{2i}$$

$$y_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 (\lambda x_{2i}) + e_i$$

$$= (\hat{\beta}_2 + \lambda \hat{\beta}_3) x_{2i} + e_i$$

$$= \hat{\alpha} x_{2i} + e_i$$

$$\hat{\alpha} = (\hat{\beta}_2 + \hat{\beta}_3 \lambda) = \frac{\sum x_{2i} y_i}{\sum x_{2i}^2}$$

Hậu quả:

không có lời giải duy nhất cho các hệ số hồi qui $(\hat{\beta}_2, \hat{\beta}_3)$ mà chỉ có lời giải duy nhất cho tổ hợp của các hệ số hồi qui $(\hat{\alpha})$

Đa cộng tuyến hoàn hảo

- Cách phát hiện: Báo lỗi từ các phần mềm
- Khắc phục: Bỏ biến thừa

Đa cộng tuyến không hoàn hảo

- Tồn tại mối quan hệ tương quan chặt chẽ giữa các biến độc lập trong mô hình
- Không giả thiết nào của phương pháp OLS bị vi phạm
- Nguyên nhân: Do bản chất mối quan hệ giữa các biến số (ví dụ: tiêu dùng điện và qui mô hộ)

Đa cộng tuyến không hoàn hảo - Hậu quả

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

- Phương sai của các ước lượng OLS bị phóng đại

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

- $\Rightarrow \text{var}(\hat{\beta}_2) \text{ lớn} \Rightarrow \text{se}(\hat{\beta}_2) \text{ lớn} \Rightarrow \text{khoảng tin cậy lớn}$
$$\left[\hat{\beta}_i - 1.96 \text{se}(\hat{\beta}_i); \hat{\beta}_i + 1.96 \text{se}(\hat{\beta}_i) \right]$$

Đa cộng tuyến không hoàn hảo - Hậu quả

- Thống kê t thấp \Rightarrow tăng khả năng chấp nhận $H_0: \beta_i = 0$
- R^2 có thể rất cao \Rightarrow tăng khả năng bác bỏ $H_0: \beta_2 = \beta_3 = \dots = \beta_k = 0$
- Dấu của các ước lượng của hệ số hồi qui có thể sai
- Các ước lượng OLS và các sai số tiêu chuẩn của chúng nhạy với những thay đổi của số liệu

Đa cộng tuyến không hoàn hảo – Cách phát hiện

- R^2 cao nhưng thống kê t có ý nghĩa thấp
- Hệ số tương quan cặp giữa 2 biến giải thích cao
- Hồi qui phụ

Mô hình : $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + u_i$

Hồi qui phụ:

$$X_i = \alpha_1 + \alpha_2 X_2 + \dots + \alpha_{i-1} X_{i-1} + \alpha_{i+1} X_{i+1} + \dots + v$$

Đa cộng tuyến không hoàn hảo – Cách phát hiện

H_0 : Không có đa cộng tuyến

$$\Leftrightarrow \alpha_2 = \alpha_3 = \dots = \alpha_{i-1} = \alpha_{i+1} = \dots = 0$$

H_1 : Có đa cộng tuyến $\Leftrightarrow \exists \alpha_i \neq 0$

$$F_i = \frac{R_i^2}{1 - R_i^2} \cdot \frac{n - k}{k - 1} \quad F \sim F_{\alpha}(k-1, n-k)$$

*(R_i^2 là hệ số xác định thu được từ hồi qui phụ)
(k là số hệ số trong mô hình hồi qui phụ)*



Đa cộng tuyến không hoàn hảo – Cách phát hiện

- Tiêu chuẩn Theil

$$m = R^2 - \sum_{i=2}^k (R^2 - R_{-i}^2)$$

R_{-i}^2 : là R^2 thu được từ mô hình hồi qui ban đầu sau khi đã bỏ biến X_i

$R^2 - R_{-i}^2$ được gọi là mức độ đóng góp của X_i đối với R^2 .

- Nếu có đa cộng tuyến : $m \neq 0$
- Nếu không có đa cộng tuyến : $m = 0$

Tuy nhiên chỉ số này không xác định được mức độ nghiêm trọng của đa cộng tuyến. 12

Đa cộng tuyến không hoàn hảo – Cách phát hiện

- Hệ số phóng đại phương sai (VIF _variance-inflating factor)

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

VIF > 10 là dấu hiệu của đa cộng tuyến nhưng không nhất thiết đúng

(R_i^2 là hệ số xác định thu được từ hồi qui phụ)

Đa cộng tuyến không hoàn hảo – Cách phát hiện

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

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

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



Đa cộng tuyến không hoàn hảo – Giải pháp

- Không làm gì (Do nothing)
- Các phương pháp khắc phục
 - ☆ Dùng thông tin tiên nghiệm
 - ☆ Gộp số liệu chéo (cross-sectional data) và số liệu theo thời gian (time series)
 - ☆ Thu thập thêm số liệu hoặc lấy thêm mẫu mới

Đa cộng tuyến không hoàn hảo – Giải pháp

☆ Bỏ bớt biến

Dựa vào giá trị của R^2 và \bar{R}^2

Phương pháp này không được khuyến khích vì nó có thể dẫn tới các ước lượng thu được trong mô hình bị chệch.

Đa cộng tuyến không hoàn hảo – Giải pháp

☆ Thực hiện phép biến đổi với các biến số

1. Sử dụng sai phân cấp 1 (Áp dụng với chuỗi thời gian)

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

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

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

Đa cộng tuyến không hoàn hảo – Giải pháp

- ***Sử dụng sai phân cấp 1***

X_2 và X_3 có thể tương quan cao nhưng không có lý do tiên nghiệm nào chắc chắn rằng sai phân của chúng cũng tương quan cao.

Một số vấn đề của phương trình sai phân cấp 1: (1) v_t có thể có tự tương quan \Rightarrow nghiêm trọng hơn; (2) mất đi một quan sát do tiến hành sai phân sẽ làm giảm một bậc tự do của mô hình \Rightarrow cân nhắc đối với các mẫu nhỏ. (3) không thể áp dụng cho số liệu chéo (không có yếu tố thời gian).

Đa cộng tuyến không hoàn hảo – Giải pháp

2. Biến đổi số liệu về dạng phân số (theo đầu người, trên một đơn vị)
3. Hồi qui đa thức (Biến độc lập có thể ở dạng bậc 2, bậc 3...)