CHƯƠNG 9

CHỌN MÔ HÌNH VÀ KIỂM ĐỊNH CHỌN MÔ HÌNH

CHỌN MÔ HÌNH

MỤC TIÊU

1. Biết cách tiếp cận để lựa chọn mô hình

2. Biết cách kiểm định việc chọn mô hình

NỘI DUNG

1 Chọn mô hình- Các sai lầm khi chọn mô hình

2 Cách tiếp cận để lựa chọn mô hình

3 Kiểm định việc chọn mô hình

1. Chọn mô hình

- Tiết kiệm
- Tính đồng nhất
- Tính thích hợp: Mô hình có R^2 càng cao càng thích hợp
- Tính bền vững về mặt lý thuyết: mô hình phải phù hợp với lý thuyết nền tảng
- Khả năng dự báo cao

2. Các sai lầm khi chọn mô hình- Hậu quả

1. Bỏ sót biến thích hợp

*i.Các tham số ước lượng sẽ bị **chệch và không vững.***

ii.Khoảng tin cậy và các kiểm định không chính xác.

iii.Dự báo dựa trên mô hình sai sẽ không đáng tin cậy.

2. Các sai lầm khi chọn mô hình- Hậu quả

2. Đưa vào mô hình những biến không phù hợp

Các ước lượng không hiệu quả, khoảng tin cậy rộng.

3. Lựa chọn mô hình không chính xác

i.Ước lượng chệch các hệ số hồi quy, dấu của hệ số hồi quy có thể sai.

ii.Có ít hệ số hồi quy ước lượng được có ý nghĩa thống kê

iii.R^2 không cao

iv.Phần dư các quan sát lớn và biểu thị sự biến thiên có tính hệ thống.

Ví dụ

- Về hàm chi phí của doanh nghiệp, dạng hàm đúng

$$Y_i = b_1 + b_2 X_i + b_3 X_i^2 + b_4 X_i^3 + u_{1i}$$

- *Bỏ sót biến quan trọng* (X_i^3)

$$Y_i = a_1 + a_2 X_i + a_3 X_i^2 + u_{2i}$$

- *Đưa biến không liên quan vào mô hình* (X_i^4)

$$Y_i = l_1 + l_2 X_i + l_3 X_i^2 + l_4 X_i^3 + l_5 X_i^4 + u_{3i}$$

- Dạng hàm sai

$$\ln Y = g_1 + g_2 X_i + g_3 X_i^2 + g_4 X_i^3 + u_{4i}$$

3. Cách tiếp cận để lưa chọn mô hình

1.Xác định số biến độc lập

Từ đơn giản đến tổng quát

Từ tổng quát đến đơn giản

2. **Kiểm định mô hình có vi phạm giả thiết**

Nếu mô hình vi phạm thì cần có biện pháp khắc phục.

3. **Chọn dạng hàm,** dựa vào

Các lý thuyết kinh tế

Các kết quả nghiên cứu thực nghiệm

4. **Sử dụng các tiêu chuẩn thông dụng** để chọn mô hình

4. Kiểm định việc chọn mô hình

a. **Kiểm định thừa biến (kiểm định Wald)**

Xét hai mô hình:

$$(U): Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_{m-1} X_{m-1} + \beta_m X_m + \beta_k X_k + U$$

$$(R): Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_{m-1} X_{m-1} + V$$

(U): mô hình không bị ràng buộc

(R): mô hình bị ràng buộc

Điều kiện ràng buộc: các hệ số hồi quy của các biến X_{m}, X_{m+1}, X_k đồng thời bằng 0

a. Kiểm định Wald

Xây dựng giả thiết để kiểm định đk ràng buộc

$H_{0:}$ $\beta_m = \ldots \beta_k = 0$

$H_{1:}$ có ít nhất một β_j khác 0

B1: Hồi quy mô hình (U) có k tham số, tính RSS_U có n-k bậc tự do

B2: Hồi quy mô hình (R) có m tham số, tính RSS_R có n-m bậc tự do

B3: Tính F

$$F = \frac{(RSS_R - RSS_U)/k - m)}{RSS_U/(n-k)} = \frac{(R^2{}_U - R^2{}_R)/(k-m)}{(1-R^2{}_U)/(n-k)}$$

a. Kiểm định Wald

B4: Tra bảng F với mức ý nghĩa α có giá trị F_{α} (k-m, n-k)

Quy tắc quyết định

- Nếu F≥ Fα (k-m, n-k): bác bỏ H_0, tức mô hình (U) không thừa biến
- Nếu F< Fα (k-m, n-k): chấp nhận H_0

Nếu dùng kết quả p-value thì quy tắc quyết định như sau:

- Nếu p ≤ α : Bác bỏ H_0
- Nếu p > α: Chấp nhận H_0

b. Kiểm định bỏ sót biến giải thích

Dùng kiểm định Reset của Ramsey:

Bước 1: Dùng OLS để ước lượng mô hình

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

Từ đó tính $\hat{Y}_i$ và R^2_{old}

Bước 2: dùng OLS để ước lượng mô hình

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 \hat{Y}^2 + \beta_4 \hat{Y}^3 + \ldots + v_i$$

Tính R^2_{new}

Kiểm định giả thiết H_0: $\beta_3 = \beta_4 = \ldots = \beta_k = 0$

Bước 3: Tính

$$F = \frac{(R^2_{new} - R^2_{old})/m}{(1 - R^2_{new})/(n-k)}$$

n: số quan sát

k: số tham số trong mô hình mới

m: số biến đưa thêm vào

b. Kiểm định bỏ sót biến giải thích

Bước 4: Quy tắc quyết định

- Nếu $F > F_{\alpha}(m,n-k)$: Bác bỏ H_0, tức các hệ số $\beta_3, \beta_4, \ldots \beta_k$ không đồng thời bằng 0, mô hình cũ đã bỏ sót biến
- Nếu $F < F_{\alpha}(m,n-k)$: Chấp nhận H_0

Nếu dùng kết quả p-value thì quy tắc quyết định như sau:

- Nếu $p \leq \alpha$: Bác bỏ H_0
- Nếu $p > \alpha$: Chấp nhận H_0

c. Kiểm định giả thiết phân phối chuẩn của u_i

Dùng kiểm định χ^2, hay kiểm định Jarque-Bera

Kiểm định giả thiết H_0: u_i có phân phối chuẩn

$$JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right]$$

$$S = \frac{\sum (u_i - \bar{u})^3}{n.SE_u^3} \qquad K = \frac{\sum (u_i - \bar{u})^4}{n.SE_u^4}$$

Nếu JB > χ^2, Bác bỏ H_0, ngược lại, chấp nhận H_0

5.Tiêu chuẩn lựa chọn mô hình

- R^2,
- R^2 điều chỉnh,
- Giá trị của hàm hợp lý log-likelihood (L),
- Tiêu chuẩn thông tin Akaike (AIC),
- Tiêu chuẩn thông tin Schwarz (SIC)

Tiêu chuẩn R^2

- R^2 đo lường % biến động của Y được giải thích bởi các X_i trong mô hình.
- R^2 càng gần 1, mô hình càng phù hợp.
- Lưu ý:
  - R^2 chỉ đo lường sự phù hợp trong mẫu
  - Khi so sánh R^2 giữa các mô hình khác nhau, các biến phụ thuộc phải giống nhau.
  - R^2 không giảm khi tăng thêm biến độc lập.

Tiêu chuẩn R^2 điều chỉnh ($\bar{R}^2$)

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

- $\bar{R}^2 \leq R^2$. $\bar{R}^2$ chỉ tăng khi giá trị tuyệt đối của giá trị t của biến được thêm vào mô hình lớn hơn 1.
- R^2 là tiêu chuẩn tốt hơn R^2.
- Các biến phụ thuộc cũng phải giống nhau.

Giá trị của hàm hợp lý log-likelihood (L)

$$L = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum U_i^2$$

- Giá trị L càng lớn chứng tỏ mô hình càng phù hợp

Tiêu chuẩn thông tin Akaike (AIC)

$$AIC = \left(\frac{RSS}{n}\right)e^{2k/n}$$

hay

$$\ln AIC = \left(\frac{2k}{n}\right) + \ln\left(\frac{RSS}{n}\right)$$

- Trong đó k là số biến được ước lượng (gồm cả hệ số tự do) và n là cỡ mẫu.
- Giá trị AIC càng nhỏ chứng tỏ mô hình càng phù hợp.

Tiêu chuẩn thông tin Schwarz (SC)

$$SC = \left(\frac{RSS}{n}\right) n^{k/n}$$

hay

$$\ln SC = \frac{k}{n}\ln n + \left(\frac{RSS}{n}\right)$$

- SC khắt khe hơn AIC.
- SC càng nhỏ, mô hình càng tốt.