CHƯƠNG 8.
ĐỊNH DẠNG MÔ HÌNH

I.1. Các thuộc tính của một mô hình tốt

A.C Harvy đưa ra một số tiêu chuẩn sau để đánh giá một mô hình:

- Tính kiệm: Mô hình là sự biểu diễn đơn giản nhưng hoàn chỉnh của hiện tượng. Mô hình càng đơn giản càng tốt.
- Tính đồng nhất: nghĩa là với cùng 1 tập số liệu, giá trị các tham số phải thống nhất

I.1. Các thuộc tính của một mô hình tốt

- Tính vững về mặt lý thuyết: Mô hình phải có tính phù hợp về mặt lý thuyết, không được sai phạm các vấn đề kinh tế cơ bản.
- Tính thích hợp: Vì mục đích của mô hình là giải thích sự thay đổi của biến phụ thuộc do sự thay đổi của các biến độc lập nên R^2, $\overline{R}^2$ cao là một tiêu chuẩn cần thiết
- Khả năng dùng cho dự báo của mô hình: dự báo dựa trên mô hình phải phù hợp với thực tế.

I.2. Các sai lầm định dạng

a. Mô hình bỏ sót biến thích hợp

Mô hình đúng: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ (8.1)

Mô hình ước lượng : $Y_i = \alpha_1 + \alpha_2 X_{2i} + v_i$ (8.2)

$\Rightarrow$ $v_i = \beta_2 X_{3i} + u_i$

✮ <u>X_2 và X_3 có tương quan</u>

($r_{23} \neq 0$) $\Rightarrow$ $Cov(X_2, v_i) \neq 0 \Rightarrow$ (8.2) vi phạm giả thiết (5) của OLS (các biến giải thích và sai số ngẫu nhiên không có quan hệ tương quan) $\Rightarrow$ Các tham số ước lượng được bị chệch ngay cả khi kỡ mẫu lớn

I.2. Các sai lầm định dạng

- Ước lượng của α_2:

$$\hat{\alpha}_2 = \frac{\sum x_{2i} y_i}{\sum x_{2i}^2} = \frac{\sum x_{2i}(\beta_2 x_{2i} + \beta_3 x_{3i} + u_i)}{\sum x_{2i}^2}$$

$$= \frac{\beta_2 \sum x_{2i}^2 + \beta_3 \sum x_{2i} x_{3i} + \sum x_{2i} u_i}{\sum x_{2i}^2}$$

I.2. Các sai lầm định dạng

$$\hat{\alpha}_2 = \beta_2 + \beta_3 \frac{\sum x_{2i} x_{3i}}{\sum x_{2i}^2} + \frac{\sum x_{2i} u_i}{\sum x_{2i}^2}$$

$$E(\hat{\alpha}_2) = \beta_2 + \beta_3 \frac{\sum x_{2i} x_{3i}}{\sum x_{2i}^2}$$

- $\beta_3 \frac{\sum x_{2i} x_{3i}}{\sum x_{2i}^2}$ là phần chệch trong ước lượng của α

I.2. Các sai lầm định dạng

$$\beta_3 \frac{\sum x_{2i} x_{3i}}{\sum x_{2i}^2} = \beta_3 \frac{\sum x_{2i} x_{3i} /(n-1)}{\sum x_{2i}^2 /(n-1)}$$

$$= \beta_3 \frac{\overline{\text{cov}}(X_2, X_3)}{\overline{\text{var}}(X_2)}$$

I.2. Các sai lầm định dạng

Từ công thức phần chệch trên cho ta thấy, chệch sẽ không đáng kể khi:

- $\beta_3 \to 0 \Rightarrow X_3$ không nên có mặt trong mô hình
- $\overline{\text{cov}}(X_2, X_3) \to 0 \Rightarrow$ rất ít trong thực tế vì các biến số kinh tế thường có tương quan chặt chẽ.

I.2. Các sai lầm định dạng

- Kì vọng của $\hat{\alpha}_1$

$$E(\hat{\alpha}_1) = E(\bar{Y} - \hat{\alpha}_2 \bar{X}_2)$$

$$= E(\hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3 - \hat{\alpha}_2 \bar{X}_2)$$

$$= \beta_1 + (\beta_2 - \alpha_2)\bar{X}_2 + \beta_3 \bar{X}_3 \neq \beta_1$$

I.2. Các sai lầm định dạng

✯ <u>X_2 và X_3 không tương quan</u> ($r_{23} = 0$) $\Rightarrow$ ước lượng của α_2 không bị chệch (như phần trình bầy trên) nhưng ước lượng α_1 vẫn bị chệch.

$$E(\hat{\alpha}_1) = \beta_1 + \beta_3 \bar{X}_3 \neq \beta_1$$

I.2. Các sai lầm định dạng

- Phương sai của yếu tố ngẫu nhiên, σ^2 , bị ước lượng chệch

$$\hat{\sigma}^2 = \frac{RSS}{df}$$

(trong hai mô hình RSS và df đều khác nhau. $RSS_1 < RSS_2$ và $df_1 < df_2$ (do (8.1) có nhiều biến giải thích hơn (8.2))

I.2. Các sai lầm định dạng

✯ $$var(\hat{\alpha}_2) = \frac{\sigma^2}{\sum x_{2i}^2}$$

là ước lượng chệch của phương sai của ước lượng đúng

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

$0 < r^2_{23} < 1 \Rightarrow var(\hat{\alpha}_2) < var(\hat{\beta}_2) \Rightarrow$ ở đây có sự đánh đổi giữa tính chệch và tính hiệu quả

I.2. Các sai lầm định dạng

- ✭Từ các hậu quả trên, khoảng tin cậy và kiểm định giả thiết không còn chính xác

  $\Rightarrow$ những kết luận sai lầm về ý nghĩa thống kê của tham số ước lượng được

- ✭ Giá trị dự báo không đáng tin cậy do mô hình sai.

I.2. Các sai lầm định dạng

b. Mô hình chứa biến số không thích hợp

- Mô hình đúng: $Y_i = \beta_1 + \beta_2 X_{2i} + u_i$ (8.3)
- Mô hình ước lượng :

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + v_i \qquad (8.4)$$

I.2. Các sai lầm định dạng

- Hậu quả của việc trong mô hình chứa biến không thích hợp như sau:

  ✯ Các tham số ước lượng từ mô hình vẫn là ước lượng không chệch và vững.

  $$E(\hat{\alpha}_1) = \beta_1 \qquad E(\hat{\alpha}_2) = \beta_2$$

  (Vì X_3 là biến không thích hợp trong mô hình nên $E(\hat{\alpha}_3) = \beta_3 = 0$)

I.2. Các sai lầm định dạng

- Phương sai của yếu tố ngẫu nhiên, σ², vẫn được ước lượng đúng
- Khoảng tin cậy và kiểm định giả thiết vẫn hợp lệ và đáng tin cậy
- Tuy nhiên, mất tính hiệu quả do ước lượng thu được của phương sai không còn là ước lượng nhỏ nhất $var(\hat{\alpha}_i) > var(\hat{\beta}_i)$

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

I.3. Dạng hàm không đúng

Hệ số thu được từ mô hình hồi quy sai sẽ không chính xác vì bị đánh giá gián tiếp thông qua thông tin khác của mô hình, do đó, sẽ cho kết luận không chính xác về ảnh hưởng của biến độc lập đến biến phụ thuộc.

II. CÁCH PHÁT HIỆN CÁC SAI LẦM ĐỊNH DẠNG

II.1. Phát hiện mô hình chứa biến không thích hợp

$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + u_t$

☞ X_5 có cần thiết trong mô hình không?

$H_0: \beta_5 = 0$

$H_1: \beta_5 \neq 0$

$\Rightarrow$ sử dụng kiểm định t

II.1. Phát hiện mô hình chứa biến không thích hợp

☞ X_4 và X_5 có cần thiết trong mô hình không?

H_0: $\beta_4 = \beta_5 = 0$

H_1: β_4, β_5 không đồng thời bằng 0

$\Rightarrow$ sử dụng kiểm định F

☞ Trường hợp không có cơ sở bác bỏ giả thiết về sự bằng 0 của các tham số ước lượng thì việc bỏ hay giữ lại các biến này cần được cân nhắc kỹ . Vì:

II.1. Phát hiện mô hình chứa biến không thích hợp

+ Khi bỏ đi một số biến số có thể dẫn đến một số giả thiết khác của mô hình không được đảm bảo.
+ Nếu trong các lý thuyết khẳng định sự có mặt của các yếu tố này trong mô hình thì dù giả thiết hệ số ước lượng của các yếu tố này có bằng 0 được "chấp nhận" cũng không nên loại bỏ các biến số này khỏi mô hình.

II.2. Phát hiện mô hình bỏ sót biến thích hợp

- Xét mô hình: $Y_i = \beta_1 + \beta_2 X_i + u_i$
- Nghi ngờ mô hình bị bỏ sót biến Z

\+ *Trường hợp 1*: có quan sát đối với biến Z

ước lượng mô hình: $Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_i + u_i$

Sau đó kiểm định giả thiết:

H_0: $\beta_3 = 0$

H_1: $\beta_3 \neq 0$

Nếu không đủ cơ sở bác bỏ giả thiết H_0 thì mô hình không bỏ sót biến Z. Còn ngược lại thì mô hình bỏ sót biến Z.

II.2. Phát hiện mô hình bỏ sót biến thích hợp

+ *Trường hợp 2*: chưa có quan sát đối với biến Z

Khi đó, người ta sẽ tìm một biến đại diện cho Z, ví dụ Z', và ước lượng lại mô hình có biến Z':

$Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_i' + u_i$

Từ đó, có kết luận về việc bỏ sót hay không bỏ sót biến Z.

II.2. Phát hiện mô hình bỏ sót biến thích hợp

a. Kiểm định Ramsey

Mô hình: $Y_i = \beta_1 + \beta_2 X_i + u_i$ (R)

Bước 1: ước lượng mô hình (R) $\Rightarrow \hat{Y}$ và R^2_R

$\Rightarrow \quad \hat{Y}^2, \hat{Y}^3 \ldots$

Bước 2: Ước lượng mô hình:

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

$\Rightarrow R^2_{UR}$

II.2. Phát hiện mô hình bỏ sót biến thích hợp

- Bước 3: Kiểm định giả thiết :

  H_0: $\beta_3 = \beta_4 = ... = 0$ (không bỏ sót biến)

  H_1 : $\beta_i \neq 0$ (i=3,4,...) (bỏ sót biến)

$$F = \frac{\left(R_{UR}^2 - R_R^2\right)/m}{\left(1 - R_{UR}^2\right)/(n-k)}$$

m: số biến giải thích mới được đưa vào mô hình

k: số các hệ số của mô hình UR

II.2. Phát hiện mô hình bỏ sót biến thích hợp

b. Kiểm định bằng nhân tử Lagrange (LM)

- Bước 1: Ước lượng mô hình xuất phát

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

$\Rightarrow$ phần dư $e_i, \hat{Y}_i$

- Bước 2: Ước lượng mô hình sau:

$$e_i = \beta_1 + \beta_2 X_i + \beta_3 \hat{Y}^2 + \beta_4 \hat{Y}^3 + \ldots + v_i$$

$\Rightarrow R^2$

II.2. Phát hiện mô hình bỏ sót biến thích hợp

Với n khá lớn nR^2 có phân bố xấp xỉ $\chi^2(m)$

m là số các biến số , $\hat{Y}^2$,...,$\hat{Y}^{m+1}$

Bước 3 kiểm định:

H_0: $\beta_3 = \beta_4 = ... = 0$ $\Leftrightarrow$ mô hình đúng

H_1 : $\beta_i \neq 0$ (i=3,4,...) $\Leftrightarrow$ mô hình sai

II.2. Phát hiện mô hình bỏ sót biến thích hợp

c. Kiểm định Durbin-Watson d

Mô hình: $Y_i = \beta_1 + \beta_2 X_i + u_i$

- Bước 1: Ước lượng mô hình gốc nhận được phần dư tương ứng.
- Bước 2: Nếu ta nghi ngờ đã bỏ sót biến Z thì sắp xếp lại e_i theo giá trị tăng dần của biến Z, trong trường hợp biến Z không có thì sắp xếp e_i theo 1 trong các biến độc lập nào đó.

II.2. Phát hiện mô hình bỏ sót biến thích hợp

- Bước 3: Tính thống kê d:

$$d = \frac{\sum_{i=2}^{n} \left(e_i - e_{i-1}\right)^2}{\sum_{i=2}^{n} e_i^2}$$

II.2. Phát hiện mô hình bỏ sót biến thích hợp

- Bước 4: Kiểm định giả thiết

  H_0: Dạng hàm đúng

  H_1 : Dạng hàm sai

  Dựa vào bảng Durbin-Watson và mức ý nghĩa, ta tìm được các giá trị d_L và d_U tương ứng. Kết luận về dạng hàm sai được khẳng định nếu giá trị d cho thấy tồn tại dấu hiệu của tự tương quan.

II.2. Phát hiện mô hình bỏ sót biến thích hợp

* Chú ý: Các tiêu chuẩn Ramsey và LM cũng được sử dụng để kiểm định giả thiết

H_0: Dạng hàm đúng

H_1 : Dạng hàm sai

II.3. Kiểm định về tính phân bố chuẩn của u

- Giả thiết:

  H_0: u có phân bố chuẩn

  H_1 : u không có phân bố chuẩn

  Do chưa biết $u_i \Rightarrow$ sử dụng e_i

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

II.3. Kiểm định về tính phân bố chuẩn của U

Trong đó: S là hệ số bất đối xứng

K là hệ số nhọn

n là kích thước mẫu

Với n khá lớn JB có phân bố xấp xỉ $\chi^2(2)$

Nếu JB > $\chi^2(2)$: Bác bỏ H_0

Nếu JB < $\chi^2(2)$: Không đủ cơ sơ bác bỏ H_0

II.3. Kiểm định về tính phân bố chuẩn của U

Trường hợp tổng quát,

$$S = \frac{E(X-\mu)^3}{\sigma^3}$$

$$K = \frac{E(X-\mu)^4}{\left[E(X-\mu)^2\right]^2}$$