CHƯƠNG 3.

MÔ HÌNH HỒI QUI BỘI

I. MÔ HÌNH HỒI QUI 3 BIẾN

I.1. Dạng mô hình

Hàm hồi qui tổng thể (PRF):

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

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

Y: biến phụ thuộc

X_2, X_3: biến giải thích

u: sai số ngẫu nhiên

i thứ tự của quan sát

I.1. Dạng mô hình

β_1 : hệ số chặn

$\beta_1 = E(Y|X_2=X_3=0)$: cho biết tác động trung bình của các biến không có trong mô hình lên biến phụ thuộc và được thể hiện bằng giá trị trung bình của Y khi $X_2 = X_3 = 0$

β_2 ,β_3 : gọi là hệ số hồi qui riêng.

$$\beta_2 = \frac{\partial E(Y|X)}{\partial X_2}$$

cho biết sự thay đổi trung bình của biến phụ thuộc Y khi X_2 thay đổi 1 đơn vị với điều kiện X_3 không đổi.

2. Các giả thiết OLS

1. Các biến giải thích là phi ngẫu nhiên
2. Kỳ vọng của sai số ngẫu nhiên u bằng 0, $E(u|X_i) = 0$

3. Phương sai của u thuần nhất (bằng nhau)

$var(u|X_i) = \sigma^2$ (với $\forall i$)

4. Không có tự tương quan giữa các yếu tố ngẫu nhiên $Cov(u_i, u_j|X_i, X_j) = 0$ (với $\forall i \neq j$)

5. u và X không tương quan với nhau

$Cov(u_i, X_i) = 0$

6. Giữa các biến X_2, X_3 không có quan hệ tuyến tính chính xác (đa cộng tuyến hoàn hảo)

7. u có phân bố chuẩn, $u \sim N(0, \sigma^2)$

3. Ước lượng các tham số của mô hình hồi qui 3 biến bằng phương pháp OLS

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + e_i$$

$$e_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}$$

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} \right)^2 \Rightarrow \min$$

3. Ước lượng các tham số của mô hình hồi qui 3 biến bằng phương pháp OLS

$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_1} = \sum_{i=1}^{n} 2\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}\right)(-1) = 0$$

$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_2} = \sum_{i=1}^{n} 2\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}\right)(-X_{2i}) = 0$$

$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_3} = \sum_{i=1}^{n} 2\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}\right)(-X_{3i}) = 0$$

3. Ước lượng các tham số của mô hình hồi qui 3 biến bằng phương pháp OLS

$$
\begin{cases}
n\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^{n} X_{2i} + \hat{\beta}_3 \sum_{i=1}^{n} X_{3i} = \sum_{i=1}^{n} Y_i \\
\hat{\beta}_1 \sum_{i=1}^{n} X_{2i} + \hat{\beta}_2 \sum_{i=1}^{n} X_{2i}^2 + \hat{\beta}_3 \sum_{i=1}^{n} X_{2i} X_{3i} = \sum_{i=1}^{n} Y_i X_{2i} \\
\hat{\beta}_1 \sum_{i=1}^{n} X_{3i} + \hat{\beta}_2 \sum_{i=1}^{n} X_{2i} X_{3i} + \hat{\beta}_3 \sum_{i=1}^{n} X_{3i}^2 = \sum_{i=1}^{n} Y_i X_{3i}
\end{cases}
$$

3. Ước lượng các tham số của mô hình hồi qui 3 biến bằng phương pháp OLS

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X}_2 - \hat{\beta}_3 \overline{X}_3$$

$$\hat{\beta}_2 = \frac{\left(\sum_{i=1}^{n} y_i x_{2i}\right)\left(\sum_{i=1}^{n} x_{3i}^2\right) - \left(\sum_{i=1}^{n} y_i x_{3i}\right)\left(\sum_{i=1}^{n} x_{2i} x_{3i}\right)}{\sum_{i=1}^{n} x_{2i}^2 \sum_{i=1}^{n} x_{3i}^2 - \left(\sum_{i=1}^{n} x_{2i} x_{3i}\right)^2}$$

$$\hat{\beta}_3 = \frac{\left(\sum_{i=1}^{n} y_i x_{3i}\right)\left(\sum_{i=1}^{n} x_{2i}^2\right) - \left(\sum_{i=1}^{n} y_i x_{2i}\right)\left(\sum_{i=1}^{n} x_{2i} x_{3i}\right)}{\sum_{i=1}^{n} x_{2i}^2 \sum_{i=1}^{n} x_{3i}^2 - \left(\sum_{i=1}^{n} x_{3i} x_{3i}\right)^2}$$

3. Ước lượng các tham số của mô hình hồi qui 3 biến bằng phương pháp OLS

$$y_i = Y_i - \overline{Y}$$

$$x_{2i} = X_{2i} - \overline{X}_2 \qquad x_{3i} = X_{3i} - \overline{X}_3$$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

- $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ được gọi là các ước lượng bình phương nhỏ nhất

4. Phương sai và độ lệch chuẩn của các ước lượng bình phương nhỏ nhất

$$\operatorname{var}\left(\hat{\beta}_2\right) = \frac{\sum_{i=1}^{n} x_{3i}^2}{\sum_{i=1}^{n} x_{2i}^2 \sum_{i=1}^{n} x_{3i}^2 - \left(\sum_{i=1}^{n} x_{2i} x_{3i}\right)^2} \sigma^2$$

$$= \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)}$$

3. Ước lượng các tham số của mô hình hồi qui 3 biến bằng phương pháp OLS

Không có σ^2 nên sử dụng $\hat{\sigma}^2$ thay thế

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-3}$$

II. Mô hình hồi qui k biến tổng quát

1. Mô hình hồi qui k biến

- PRF: $E(Y|X_2, X_3, ..., X_k) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki}$
- Giá trị cá biệt: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + u_i$

u_i : yêú tố ngẫu nhiên

β_1: Hệ số chặn

β_2 , $\beta 3$,..., βk: các hệ số hồi qui (hệ số góc)

β_k cho biết sự thay đổi trung bình của biến phụ thuộc, Y, khi X_k thay đổi 1 đơn vị, các biến độc lập khác không đổi.

Mô hình hồi qui k biến

Biểu diễn hàm hồi qui dưới dạng ma trận

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \quad U = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \vdots & & & & \\ 1 & X_{2n} & X_{3n} & \cdots & X_{kn} \end{bmatrix}$$

$$\Rightarrow Y = X\beta + u$$

2. Ước lượng các tham số bằng phương pháp OLS

- e = Y – X$\hat{\beta}$
- Phương pháp OLS ước lượng giá trị của các tham số $\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$ sao cho:

$$RSS = \sum_{i=1}^{n} e_i^2 = e'e \Rightarrow \min$$

2. Ước lượng các tham số bằng phương pháp OLS

$$RSS = \left(Y - X\hat{\beta}\right)' \left(Y - X\hat{\beta}\right)$$

$$= \mathrm{Y}'\mathrm{Y} - \hat{\beta}'\mathrm{X}'\mathrm{Y} - \mathrm{Y}'\mathrm{X}\hat{\beta} + \hat{\beta}'\mathrm{X}'\mathrm{X}\hat{\beta}$$

$$= \mathrm{Y}'\mathrm{Y} - 2\hat{\beta}'\mathrm{X}'\mathrm{Y} + \hat{\beta}'\mathrm{X}'\mathrm{X}\hat{\beta}$$

$$\frac{\partial RSS}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0$$

$$\left(\mathrm{X}'\mathrm{X}\right)^{-1} \neq 0 \quad \Longrightarrow \quad \boxed{\hat{\beta} = \left(\mathrm{X}'\mathrm{X}\right)^{-1}\mathrm{X}'\mathrm{Y}}$$

$\hat{\beta}$ là ma trận hệ số ước lượng OLS

3. Giả thiết OLS trong mô hình k biến tổng quát

1. $X_2, X_3, ..., X_k$ là các biến xác định, hay ma trận X xác định

2. $$E(u) = E\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_u \end{bmatrix} = \begin{bmatrix} E(u_1) \\ E(u_2) \\ \vdots \\ E(u_u) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

3. Giả thiết OLS trong mô hình k biến tổng quát

3. var(u) = E(u'u) = $\sigma^2 I$
4. $\operatorname{cov}\left(u_i, u_j\right) = 0 \qquad \forall i \neq j$
5. cov (X, u) = 0
6. Không có đa cộng tuyến giữa các biến độc lập hay hạng của ma trận X bằng k
7. u_i phân bố chuẩn. u~N(0, σ^2I)

4. Ma trận hiệp phương sai của $\hat{\beta}$

$$\mathrm{cov}(\hat{\beta}) = \begin{bmatrix} \mathrm{var}(\hat{\beta}_1) & \mathrm{cov}(\hat{\beta}_1, \hat{\beta}_2) & \cdots & \mathrm{cov}(\hat{\beta}_1, \hat{\beta}_k) \\ \mathrm{cov}(\hat{\beta}_2, \hat{\beta}_1) & \mathrm{var}(\hat{\beta}_2) & \cdots & \mathrm{cov}(\hat{\beta}_2, \hat{\beta}_k) \\ \vdots & & & \\ \mathrm{cov}(\hat{\beta}_k, \hat{\beta}_1) & \mathrm{cov}(\hat{\beta}_k, \hat{\beta}_2) & \cdots & \mathrm{var}(\hat{\beta}_k) \end{bmatrix}$$

$$\boxed{\mathrm{cov}(\hat{\beta}) = [X'X]^{-1}\sigma^2}$$

4. Ma trận hiệp phương sai của $\hat{\beta}$

- σ^2 không quan sát được nên $\hat{\sigma}^2$ được sử dụng thay thế

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-k} = \frac{RSS}{n-k}$$

với k là số hệ số có trong hàm hồi qui

III. Phân tích các hệ số của mô hình

1. Khoảng tin cậy và kiểm định giả thiết các hệ số hồi qui - Kiểm định T

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

   $\hat{\beta}_i$ phân phối chuẩn với kì vọng β_i, độ lệch chuẩn $\sigma_{\hat{\beta}_i}$

$$\Rightarrow t_i = \frac{\hat{\beta}_i - \beta_i}{se\left(\hat{\beta}_i\right)} \sim T(n-k)$$

a. Khoảng tin cậy

- Khoảng tin cậy đối xứng:

$$\left(\hat{\beta}_i - se\left(\hat{\beta}_i\right).t_{\alpha/2}^{\ (n-k)} < \beta_i < \hat{\beta}_i + se\left(\hat{\beta}_i\right).t_{\alpha/2}^{\ (n-k)}\right)$$

- Khoảng tin cậy tối đa (phía trái):

$$\beta_i < \hat{\beta}_i + se\left(\hat{\beta}_i\right).t_{\alpha}^{\ (n-k)}$$

- Khoảng tin cậy tối thiểu (phía phải):

$$\beta_i > \hat{\beta}_i - se\left(\hat{\beta}_i\right)t_{\alpha}^{\ (n-k)}$$

b. Kiểm định giả thiết

- $H_0 : \beta_i = \beta_i^*$ (với i=1,2,...,k)

Loại kiểm định	Giả thiết H_1	Miền bác bỏ
Hai phía	$\beta_i \neq \beta_i^*$	$\|ti\| > t_{\alpha/2}^{(n-k)}$
Phía phải	$\beta_i > \beta_i^*$	$t_i > t_{\alpha}^{(n-k)}$
Phía trái	$\beta_i < \beta_i^*$	$t_i <- t_{\alpha}^{(n-k)}$

2. Kiểm định sự phù hợp của hàm hồi qui

- Hệ số xác đinh bội, R^2

$$R^2 = \frac{ESS}{TSS}$$

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

$$R^2 = \frac{\hat{\beta}_2 \sum_{i=1}^{n} y_i x_{2i} + \hat{\beta}_3 \sum_{i=1}^{n} y_i x_{3i} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} y_i x_{ki}}{\sum_{i=1}^{n} y_i^2}$$

2. Kiểm định sự phù hợp của hàm hồi qui (tiếp)

- Hệ số xác đinh bội, R^2

$$RSS = \sum e_i^2 = e'e = Y'Y - \hat{\beta}'X'Y$$

$$TSS = \sum y_i^2 = \sum Y_i^2 + 2\sum Y_i\bar{Y} + n\bar{Y}^2 = Y'Y - n\bar{Y}^2$$

$$ESS = TSS - RSS = \hat{\beta}'X'Y - n\bar{Y}^2$$

$$R^2 = \frac{\hat{\beta}'X'Y - n\bar{Y}^2}{Y'Y - n\bar{Y}^2}$$

2. Kiểm định sự phù hợp của hàm hồi qui (tiếp)

- R = $\sqrt{R^2}$ được gọi là hệ số tương quan bội, đo sự kết hợp tuyến tính đồng thời của biến phụ thuộc Y với tất cả các biến giải thích X
- R^2 được gọi là hệ số xác định bội, cho ta biết % sự thay đổi của biến phụ thuộc được giải thích bởi các biến độc lập trong mô hình
- R^2 tăng theo số biến độc lập được đưa vào mô hình, ngay cả khi biến độc lập được đưa thêm vào không giải thích sự biến động của Y

2. Kiểm định sự phù hợp của hàm hồi qui (tiếp)

- Cân nhắc việc đưa một biến giải thích mới vào mô hình, người ta sử dụng hệ số xác định bội đã hiệu chỉnh, $\bar{R}^2$

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

- k$\uparrow$ $\Rightarrow$ $\bar{R}^2 \uparrow$ (do $R^2 \uparrow$)
  $\Rightarrow$ $\bar{R}^2 \downarrow$ (theo công thức)

$\Rightarrow$nếu đưa thêm 1 biến vào mô hình mà làm $\bar{R}^2$ giảm thì không nên đưa thêm biến đó vào mô hình

2. Kiểm định sự phù hợp của hàm hồi qui (tiếp)

- ***Tính chất của*** $\bar{R}^2$
- Nếu k>1, $\bar{R}^2 \leq R^2 \leq 1$, có nghĩa là nếu tăng số biến giải thích thì $\bar{R}^2$ có thể tăng nhưng luôn tăng chậm hơn R^2.
- R^2 không âm nhưng $\bar{R}^2$ có thể âm
- $\bar{R}^2$ là một trong hai tiêu chuẩn (kỹ thuật) để xét có nên đưa thêm biến vào mô hình hay không.

  + $\bar{R}^2$ tăng

  + Hệ số ứng với biến mới đưa vào có ý nghĩa về mặt thống kê

Kiểm định sự phù hợp của hàm hồi qui

- H_0: $\beta_2 = \beta_3 = ... = \beta_k = 0 \quad \Leftrightarrow R^2 = 0$

  $\Leftrightarrow$ hồi qui không phù hợp
- H_1: $\exists \beta_i \neq 0 \quad \Leftrightarrow R^2 > 0$

  $\Leftrightarrow$ hàm hồi qui phù hợp

$$\boxed{F = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{ESS}{TSS - ESS} \cdot \frac{n-k}{k-1} = \frac{R^2}{1-R^2} \cdot \frac{n-k}{k-1}}$$

- $F \sim F_{\alpha}(k-1, n-k)$

III.3. Kiểm định sự thu hẹp hồi qui

$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + U_i$ (UR)

H_0: $\beta_2 = \beta_3 = 0 = \ldots = \beta_{r+1} = 0$

H_1: $\exists \beta_i \neq 0$, (i = 2, 3, .., r+1)

Giả thiết H_0 cho rằng r hệ số bằng 0 $\Rightarrow$ bao gồm r ràng buộc

$Y_i = \beta_1 + \beta_{r+2} X_{r+2i} + ... + \beta_k X_{ki} + U_i$ (R)

(R): hàm hồi qui thu hẹp (r hệ số)

3. Kiểm định sự thu hẹp hồi qui (tiếp)

- Các bước kiểm định

+ Hồi qui (UR) và (R), thu được tổng bình phương các phần dư, RSS_{UR} và RSS_R

RSS_{UR} có $df_{UR} = n – k$

RSS_R có $df_R = [n – (k – r)]$

$$F = \frac{RSS_R - RSS_{UR}}{RSS_{UR}} \cdot \frac{n-k}{r} \sim F(r, n-k)$$

3. Kiểm định sự thu hẹp hồi qui (tiếp)

- Do (UR) và (R) có cùng biến phụ thuộc $\Rightarrow$ TSS_{UR}=TSS_R = TSS

$$F = \frac{\left(TSS - ESS_R\right) - \left(TSS - ESS_{UR}\right)}{TSS - ESS_{UR}} \cdot \frac{n-k}{r}$$

$$F = \frac{R_{UR}^2 - R_R^2}{1 - R_{UR}^2} \cdot \frac{n-k}{r} \sim F(r, n-k)$$

V.1. Dự báo giá trị trung bình

- $\hat{Y}_0 = X_0'\hat{\beta}$ là ước lượng của E(Y|X$_0$) với kì vọng là $X_0'\beta$ và phương sai là:

$$\boxed{\operatorname{var}\left(\hat{Y}_0\right) = \operatorname{var}\left(X_0'\hat{\beta}\right) = X_0'\operatorname{cov}\left(\hat{\beta}\right)X_0}$$

$$= X_0'\sigma^2\left(X'X\right)^{-1}X_0 = \sigma^2 X_0'\left(X'X\right)^{-1}X_0$$

- σ^2 không biết, sử dụng $\hat{\sigma}^2$ thay thế

$$\boxed{\Rightarrow se(\hat{Y}_0) = \hat{\sigma}\sqrt{X_0'\left(X'X\right)^{-1}X_0}}$$

V.1. Dự báo giá trị trung bình (tiếp)

Với mức ý nghĩa α, giá trị trung bình của biến phụ thuộc tương ứng với vecto các biến độc lập X_0 được dự báo nằm trong khoảng:

$$\left[\hat{Y}_0 - t_{\alpha/2}^{(n-k)} se\left(\hat{Y}_0\right), \hat{Y}_0 + t_{\alpha/2}^{(n-k)} se\left(\hat{Y}_0\right)\right]$$

V.2. Dự báo giá trị cá biệt

- $\hat{Y}_0 = X_0'\hat{\beta}$ là ước lượng của Y_0 với kì vọng là $X_0'\beta$ và phương sai là:

$$\operatorname{var}\left(Y_0 - \hat{Y}_0\right) = \operatorname{var}\left(X_0'\hat{\beta}\right) + \sigma^2$$

$$= \sigma^2\left[1 + X_0'\left(X'X\right)^{-1}X_0\right]$$

- σ^2 không biết, sử dụng $\hat{\sigma}^2$ thay thế

$$\boxed{\Rightarrow se(Y_0 - \hat{Y}_0) = \hat{\sigma}\sqrt{1 + X_0'\left(X'X\right)^{-1}X_0}}$$

V.2. Dự báo giá trị cá biệt (tiếp)

- Với mức ý nghĩa α, giá trị cá biệt của biến phụ thuộc tương ứng với vecto các biến độc lập X_0 được dự báo nằm trong khoảng

$$\left[\hat{Y}_0 - t_{\alpha/2}{}^{(n-k)} se\left(Y_0 - \hat{Y}_0\right), \hat{Y}_0 + t_{\alpha/2}{}^{(n-k)} se\left(Y_0 - \hat{Y}_0\right)\right]$$

Một số dạng hàm hồi qui

- Hàm Cobb-Douglas (log - log)

$$Y = \beta_1 X_2^{\beta_2} X_3^{\beta_3} \ldots X_k^{\beta_k}$$

$$\Leftrightarrow \ln Y = \ln \beta_1 + \beta_2 \ln X_2 + \beta_3 \ln X_3 \ldots + \beta_k \ln X_k$$

$$\varepsilon_{X_i}^{Y} = \frac{\partial Y}{\partial X_i} \frac{X_i}{Y} = \beta_i$$

βi cho biết khi **Xi tăng 1%** thì trung bình của **Y thay đổi bao nhiêu %** trong điều kiện các yếu tố khác không đổi.

Một số dạng hàm hồi qui

- Hàm log – lin

VD: $Y_t = Y_0(1 + r)^t$

$\ln(Y_t) = \ln(Y_0) + t*\ln(1 + r)$

$\Leftrightarrow \quad \ln(Y_t) = \beta_1 + X_t*\beta_2$

$$\beta_2 = \frac{\dfrac{dY}{Y}}{dX}$$

βi cho biết khi **Xi tăng 1 đơn vị** thì trung bình của **Y thay đổi bao nhiêu** % trong điều kiện các yếu tố khác không đổi.

Một số dạng hàm hồi qui

- Hàm lin – log

VD: $Y = \beta_1 + \beta_2 \ln(X)$

$$\beta_2 = \frac{dY}{\frac{dX}{X}}$$

βi cho biết khi **Xi tăng 1 %** thì trung bình của **Y thay đổi bao nhiêu đơn vị** trong điều kiện các yếu tố khác không đổi.