Chương 3: Hồi quy bội

A comparison of the models

Model	Eq.	Slope	Elasticity
Linear	$Y = \beta_1 + \beta_2 X$	β_2	$\beta_2 \left(\frac{X}{Y} \right)$
Log-linear	$\ln Y = \beta_1 + \beta_2 \ln X$	$\beta_2 \left(\frac{Y}{X} \right)$	β_2
Log-lin	$\ln Y = \beta_1 + \beta_2 X$	$\beta_2 Y$	$\beta_2 X$
lin-log	$Y = \beta_1 + \beta_2 \ln X$	$\beta_2 \left(\frac{1}{X} \right)$	$\beta_2 \left(\frac{1}{Y} \right)$
Reci	$Y = \beta_1 + \beta_2 \; 1/X$	$-\beta_2 \frac{1}{X^2}$	$-\beta_2 \frac{1}{XY}$

3.1. Mô hình hồi quy bội

PRF: $E(Y \mid X_2, X_3) = \beta_1 + \beta_2 X_2 + \beta_3 X_3$

β_1: Hệ số chặn = giá trị trung bình của biến Y khi $X_2 = X_3 = 0$.

β_2, β_3: các hệ số hồi quy riêng.

- Giá trị biến Y ở quan sát thứ i là:

$$Y_i = E(Y|X_2, X_3) + U_i = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + U_i$$

3.2.Các giả thiết của mô hình CLRM (nhắc lại)

1. Mô hình là tuyến tính $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$
2. Kì vọng U_i bằng 0: $E(u_i \mid X_{2i}, X_{3i}) = 0$
3. Các U_i thuần nhất: $\mathrm{var}(u_i) = \sigma^2$
4. Không có sự tương quan giữa các U_i: $\mathrm{cov}(u_i u_j) = 0, i \neq j$
5. Không có quan hệ tuyến tính giữa các biến giải thích. $\lambda_1 1 + \lambda_2 X_{2i} + \lambda_3 X_{3i} \neq 0,$ $\forall \lambda_1, \lambda_2, \lambda_3 \neq (0,0,0)$

- β_2 đo lường sự thay đổi kì vọng của Y ứng với 1 đơn vị tăng lên của X_2, X_3 không đổi.
- β_3 đo lường sự thay đổi kì vọng của Y ứng với 1 đơn vị tăng lên của X_3, X_2 không đổi.

Hiện tượng đa cộng tuyến

- Giả thiết 5 bị vi phạm, ví dụ: $X_{3i} = \lambda X_{2i}$

$$
\begin{aligned}
Y_i &= \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \\
&= \beta_1 + \beta_2 X_{2i} + \beta_3 (\lambda X_{2i}) + u_i \\
&= \beta_1 + (\beta_2 + \beta_3 \lambda) X_{2i} + u_i \\
&= \alpha_1 + \alpha_2 X_{2i} + u_i
\end{aligned}
$$

- Ước lượng mô hình này không thấy được ảnh hưởng từng biến lên biến phụ thuộc.

3.3. Ước lượng các tham số trong hồi quy bội

$$\min_{\beta_1,\beta_2,\beta_3} : RSS = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (Y_i - \beta_1 - \beta_2 X_{2i} - \beta_3 X_{3i})^2$$

- Điều kiến cần

$$\sum_{i=1}^{n} (Y_i - \beta_1 - \beta_2 X_{2i} - \beta_3 X_{3i}) = 0$$

$$\sum_{i=1}^{n} (Y_i - \beta_1 - \beta_2 X_i - \beta_3 X_{3i}) X_{2i} = 0$$

$$\sum_{i=1}^{n} (Y_i - \beta_1 - \beta_2 X_i - \beta_3 X_{3i}) X_{3i} = 0$$

Giải được

$$\beta_1 = \overline{Y} - \beta_2 \overline{X}_2 - \beta_3 \overline{X}_3$$

$$\beta_2 = \frac{\sum y_i x_{2i} \sum x_{3i}^2 - \sum y_i x_{3i} \sum x_{2i} x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i} x_{3i}\right)^2}$$

$$\beta_3 = \frac{\sum y_i x_{3i} \sum x_{2i}^2 - \sum y_i x_{2i} \sum x_{3i} x_{2i}}{\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i} x_{3i}\right)^2}$$

- Rất lãng phí thời gian để nhớ kết quả này.

3.4. Phương sai và độ lệch chuẩn của các ƯL
A nicer expression with a simple interpretation

Let

$$x_{2i} = \hat{\delta}_1 + \hat{\delta}_2 x_{3i} + \hat{r}_{2i},$$
$$x_{3i} = \hat{\gamma}_1 + \hat{\gamma}_2 x_{2i} + \hat{r}_{3i},$$

In a multiple regression, we "partial out" the effect of the other variables

Then

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} \hat{r}_{2i} y_i}{\sum_{i=1}^{n} \hat{r}_{2i}^2}, \qquad \hat{\beta}_3 = \frac{\sum_{i=1}^{n} \hat{r}_{3i} y_i}{\sum_{i=1}^{n} \hat{r}_{3i}^2}$$

And the variances of the estimated parameters can also be written in a nice way

$$\mathrm{var}(\beta_2) = \frac{\sigma^2}{\sum \hat{r}_{2i}^2} = \frac{\sigma^2}{\sum x_{2i}^2 (1 - R_2^2)}$$

$$\mathrm{var}(\beta_3) = \frac{\sigma^2}{\sum \hat{r}_{3i}^2} = \frac{\sigma^2}{\sum x_{3i}^2 (1 - R_3^2)}$$

where R_k^2 is the R-square from the regression of X_k on the other regressors

3.5. Các tính chất của ước lượng OLS

1. SRF đi qua các điểm $(\overline{Y}, \overline{X}_2, \overline{X}_3)$
2. Trung bình $\overline{\hat{Y}} = \overline{Y}$
3. Trung bình của các phần dư bằng 0: $\sum_{i=1}^{n} e_i = 0$
4. Không có tương quan giữa $e_i, \hat{Y}_i$: $\sum_{i=1}^{n} e_i \hat{Y}_i = 0$
5. Tương quan giữa biến giải thích và phần dư bằng 0: $\sum_{i=1}^{n} e_i X_{2i} = \sum_{i=1}^{n} e_i X_{3i} = 0$

Với các giả thiết đã cho,

Các ước lượng β_2, β_3

là BLUE

3.7. Hệ số xác định R^2, ma trận tương quan, hệ số tương quan riêng phần

a. Hệ số xác định R^2

- Tương tự hồi quy đơn, chúng ta định nghĩa TSS, ESS, và RSS
- Từ đó tính R^2
- Như hồi quy đơn, R^2 đo độ thích hợp của hàm hồi quy.

$$TSS = ESS + RSS \qquad 1 = \frac{ESS}{TSS} + \frac{RSS}{TSS} \Leftrightarrow \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

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

Hệ số xác định bội đã hiệu chỉnh

- V/đ chính với R^2 là tăng khi thêm biến giải thích mới.
- Giải pháp cho v/đ này là điều chỉnh đo lường R^2 sao cho nó không phải luôn luôn tăng.
- Khi nào thì thêm biến giải thích? Ta dùng hệ số xác định bội đã hiệu chỉnh $(\bar{R}^2)$ để cân nhắc việc thêm biến giải thích mới vào mô hình.

Công thức:

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2}$$

- Xem tr902 Guarati

$$\bar{R}^2 = 1 - \frac{MSE}{MST} = 1 - \frac{\frac{1}{n-k}\sum \hat{u}_i^2}{\frac{1}{n-1}\sum y_i^2} = 1 - \frac{\hat{\sigma}^2}{s_Y^2}$$

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

- R^2 hiệu chỉnh không bị giới hạn trong khoảng 0 và 1.

Một số tính chất:

- R^2 luôn tăng khi thêm biến giải thích.
- R^2 cực đại tương đương RSS cực tiểu.
- Nếu k > 1, $\bar{R}^2 \le R^2 \le 1$.
- $\bar{R}^2 \ge 0$, nhưng $\bar{R}^2$ có thể âm. Như vậy khi $\bar{R}^2$ còn tăng thì ta còn phải đưa thêm biến mới. $\bar{R}^2$ còn có thể tăng khi mà hệ số của biến mới trong hàm hồi quy khác không.

b. Ma trận tương quan

- Xét mô hình $Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + U_i$
- Kí hiệu r_{tj} là hệ số tương quan giữa biến t và biến thứ j. Nếu t=1 thì r_{1j} là hệ số tương quan giữa biến Y và biến X_j.

$$r_{1j}^2 = \frac{\left[\sum_{i=1}^{n} y_i x_{ij}\right]^2}{\sum_{i=1}^{n} y_i^2 \sum_{i=1}^{n} x_{ji}^2}; \qquad r_{tj}^2 = \frac{\left[\sum_{i=1}^{n} x_{ti} x_{ji}\right]^2}{\sum_{i=1}^{n} x_{ti}^2 \sum_{i=1}^{n} x_{ji}^2}$$

- Dễ thấy $r_{tj}=r_{jt}$; $r_{jj}=1$;Ma trận hệ số tương quan

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ r_{21} & r_{22} & \cdots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \cdots & r_{kk} \end{bmatrix} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1k} \\ r_{21} & 1 & \cdots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \cdots & 1 \end{bmatrix}$$

C. Hệ số tương quan riêng phần

- Xét mô hình $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$
- $R_{12,3}$ là hệ số tương quan giữa Y và X_2 trong khi X_3 không đổi (bậc nhất - sau dấu phẩy có 1 số hạng).
- $R_{13,2}$ là hệ số tương quan giữa Y và X_3 trong khi X_2 không đổi.
- $R_{23,1}$ là hệ số tương quan giữa X_2 và X_3 trong khi Y không đổi.

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}}$$

$$r_{23.1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{(1 - r_{12}^2)(1 - r_{13}^2)}}$$

3.8. Khoảng tin cậy và kiểm định giả thiết về các hệ số hồi quy riêng - Kiểm định T

- Với giả thiết yếu tố ngẫu nhiên phân bố chuẩn, KTC và kiểm định giả thiết về các hệ số hồi quy riêng hoàn toàn như phần trình bày ở hồi quy đơn. (page 257 Guarati)

$$t = \frac{\hat{\beta}_i - \beta_i}{Se(\hat{\beta}_i)} \sim T(n-3)$$

- df = n-3

$$\hat{\beta}_i - t_{\alpha/2} Se(\hat{\beta}_i) < \beta_i < \hat{\beta}_i + t_{\alpha/2} Se(\hat{\beta}_i), \quad i = \overline{1,3}$$

- Kđ giả thiết tương tự hàm hai biến

(với df = n-3).

3.9. Phân tích phương sai (ANOVA) - Kđ F

Source of variation	SS	df	MS (or MSS)
Model	ESS	$k-1$	ESS/($k-1$)
Residual	RSS	$n-k$	RSS/($n-k$)
Total	TSS	$n-1$	TSS/($n-1$)

Kiểm định F

- Chúng ta có thể áp dụng phân tích phương sai để kiểm định giả thiết:
- H_0: $\beta_2 = \ldots = \beta_k = 0$
- H_1: tồn tại ít nhất một hệ số riêng $\beta_i \neq 0$.

$$F = \frac{\mathrm{ESS}/\mathrm{df}}{\mathrm{RSS}/\mathrm{df}} = \frac{\mathrm{ESS}/(k-1)}{\mathrm{RSS}/(n-k)}$$

- Nếu $F > F_{\alpha}(k-1,\, n-k)$ thì bác bỏ H_0.

Quan hệ giữa R^2 và thống kê F

- Thống kê F của các tham số có thể biểu diễn như hàm của R^2.

$$F = \frac{ESS/(k-1)}{RSS/(n-k)}$$

$$= \frac{ESS/(k-1)}{[TSS-ESS]/(n-k)}$$

$$= \frac{[ESS/TSS]/(k-1)}{[1-ESS/TSS]/(n-k)}$$

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

- Như vây kđ F cũng là Kđ mức ý nghĩa R^2.

3.10. Hồi quy có điều kiện ràng buộc - Kiểm định F

- Kđ F được dùng để kđ sự hạn chế tổng quát hơn.
- Cho mô hình $Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + U_i$
- Kđ cặp giả thiết

H_0: $\beta_{k-m+1} = \beta_{k-m+2} = \ldots = \beta_k = 0$

H_1: $\beta_{k-m+1}, \beta_{k-m+2}, \ldots, \beta_k$ không đồng thời bằng 0.

$$F = \frac{(ESS_{UR} - ESS_R)/m}{RSS_{UR}/(n-k)} = \frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(n-k)} \sim F(m,n-k)$$

- $F > F_\alpha(m,n-k)$: H_0 bị bác bỏ.

- Thủ tục kđ tổng quát:
  1. ƯL mô hình không có ràng buộc.
  2. ƯL mô hình với ràng buộc.
  3. Tính toán thống kê, kết luận.

- Nếu giả thiết ràng buộc không làm thay đổi biến phụ thuộc trong 2 mô hình, ta có thể dùng công thức rút gọn sau:

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

Ràng buộc trên hàm Cobb-Douglas

The Cobb-Douglas function in log form is

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

Constant returns to scale implies

$$\beta_2 + \beta_3 = 1 \Leftrightarrow \beta_2 = 1 - \beta_3,$$

$$\ln Y_i = \beta_1 + (1 - \beta_3) \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

$$\ln(Y_i / X_{2i}) = \beta_1 + \beta_3 \ln(X_{3i} / X_{2i}) + u_i$$

This is a linear restriction

(Biến phụ thuộc thay đổi)

- Cách 2: $H_0: \beta_2 + \beta_3 = 1$

  $H_1: \beta_2 + \beta_3 \neq 1$

$$t = \frac{(\hat{\beta}_2 + \hat{\beta}_3) - (\beta_2 + \beta_3)}{se(\hat{\beta}_2 + \hat{\beta}_3)}$$

$$= \frac{(\hat{\beta}_2 + \hat{\beta}_3) - 1}{\sqrt{\mathrm{var}(\hat{\beta}_2) + \mathrm{var}(\hat{\beta}_3) + 2\,\mathrm{cov}(\hat{\beta}_2 \hat{\beta}_3)}}$$

Nếu $|t| > t_{\alpha/2}$ (n-k) thì bác bỏ H_0.

3.11. Dự báo:

- Xét mô hình $Y_i = \beta_1+\beta_2X_{2i}+\ldots+ \beta_kX_{ki}+U_i$
- Cho $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}; \quad U = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & X_{21} & \cdots & X_{k1} \\ 1 & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & \cdots & X_{kn} \end{bmatrix}$

  $Y = X\beta + U$
- Dự báo giá trị trung bình $E(Y|X^0)$ tại

$$X^0 = \begin{bmatrix} 1 \\ X_2^0 \\ X_3^0 \\ \vdots \\ X^0 \end{bmatrix};$$

Dự báo giá trị trung bình

$$\hat{Y} = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k = X'\beta$$

$$(\hat{Y}_0 \mid X^0) = X^{0\prime}\beta$$

$$Var(\hat{Y}_0 \mid X^0) = \sigma^2 X^{0\prime}(X'X)^{-1}X^0$$

$$Se(\hat{Y}_0 \mid X^0) = \sqrt{\sigma^2 X^{0\prime}(X'X)^{-1}X^0}$$

$$\hat{Y}_0 - t_{\alpha/2} Se(\hat{Y}_0 \mid X^0) \leq E(Y \mid X^0) \leq \hat{Y}_0 + t_{\alpha/2} Se(\hat{Y}_0 \mid X^0)$$

(df = n-k)

- Dự báo giá trị cá biệt:

$$Y_i = X'\beta + e_i \Rightarrow \operatorname{var}(Y_0 \mid X^0) = \operatorname{var}(X'\beta + e_i) = \operatorname{var}(X'\beta) + \sigma^2$$

$$Var(Y_0 \mid X^0) = \sigma^2[1 + X^{0'}(X'X)^{-1}X^0)$$

$$Se(Y_0 \mid X^0) = \sqrt{\sigma^2[1 + X^{0'}(X'X)^{-1}X^0)}$$

$$\hat{Y}_0 - t_{\alpha/2}Se(Y_0 \mid X^0) \le (Y_0 \mid X^0) \le \hat{Y}_0 + t_{\alpha/2}Se(Y_0 \mid X^0)$$