CHƯƠNG III

Mô hình hồi quy tuyến tính đa biến

Dương Nam Hà (VNUA)

1

Nội dung chính

- I. Giới thiệu mô hình
- II. Ước lượng các tham số - Phương pháp OLS
- III. Phân tích phương sai & Kiểm định giả thuyết thống kê

Dương Nam Hà (VNUA)

2

I. Giới thiệu mô hình

- 1.1 Giới thiệu
  - $E(Y_i) = B_1 X_{1i} + B_2 X_{2i} + B_3 X_{3i} + B_4 X_{4i}$
  - $Y_i = B_1 X_{1i} + B_2 X_{2i} + B_3 X_{3i} + B_4 X_{4i} + u_i$
  - → Trường hợp tổng quát với k biến độc lập:
  - $E(Y_i) = B_1 X_{1i} + B_2 X_{2i} + ... + B_k X_{ki}$
  - $Y_i = B_1 X_{1i} + B_2 X_{2i} + ... + B_k X_{ki} + u_i$
    $= \sum_j B_j X_{ji} + u_i$
    - với $i = 1,2 ... n$; $j = 1,2,..., k$ và $X_{1i} = 1$

Dương Nam Hà (VNUA)

3

I. Giới thiệu… (tiếp)

- 1.2 Ý nghĩa các tham số
  - Hằng số (B_0)
    - Khi tất cả các $X_j = 0$
    - Hàm sx: tham số kỹ thuật
  - Tham số (hệ số) hồi quy riêng phần (B_j)
    - $\partial Y / \partial X_j = B_j$
    - Sự thay đổi của $E(Y_i)$ khi thay đổi một đơn vị của X_j (với tất cả các biến độc lập khác không đổi)
    - Chú ý: trong nhiều trường hợp, cần dựa trên mô hình ban đầu

Dương Nam Hà (VNUA)

4

I. Giới thiệu… (tiếp)

- 1.3 Các giả thiết (Assumptions):
  - (1) $Y = XB + u$
  - (2) $E(u_i) = 0$ (hay $E(Y_i) = XB$)
  - (3) $Var(u_i) = \sigma^2 I$ (constant)
    - $E(uu') = \begin{pmatrix} \sigma^2 & 0 & \ldots\ldots & 0 \\ 0 & \sigma^2 & \ldots\ldots & 0 \\ . & . & \ldots\ldots & 0 \\ 0 & 0 & \ldots\ldots & \sigma^2 \end{pmatrix} = \sigma^2 I$

Dương Nam Hà (VNUA)

5

I. Giới thiệu… (tiếp)

- (4) X là ma trận các biến phi ngẫu nhiên bậc (n x k)
- (5) $Cov(u_i, u_j) = 0$ (không có tương quan chuỗi)
- (6) $u_i \neq f(X_i)$ hay $Cov(X_i, u_i) = 0$
- (7) Không có đa cộng tuyến
- (8) $u_i \sim N(0, \sigma^2 I)$
- (9) Số quan sát n > số tham số ƯL

Dương Nam Hà (VNUA)

6

II. Ước lượng các tham số - Phương pháp OLS

- PRF: $Y_i = B_1 X_{1i} + B_2 X_{2i} + ... + B_k X_{ki} + u_i$
  hay $Y = XB + u$
- SRF: $y = Xb + e$ ($\rightarrow e = y - xb$)
- Các bước tiến hành ƯL tham số:
  - (1) Đặt các b_j là các ước lượng của B_j
  - (2) $e_i = Y_i - b_1 X_{1i} - ... - b_j X_{ji} - ... - b_k X_{ki}$
  - (3) Để $E(b_j) = B_j \Rightarrow Z = \Sigma e_i^2$ đạt min
  - (4) $\partial Z/\partial b_j = 0 \rightarrow$ hệ phương trình với các biến b_j
  - (5) Giải hệ pt trên $\rightarrow b_j$

Dương Nam Hà (VNUA)

7

II. Ước lượng… (tiếp)

- Trong đó, tổng bình phương các sai số (e):
  - $e'e = (Y - Xb)'(Y - Xb)$ (') ma trận chuyển vị
  - $= Y'Y - b'X'Y - Y'Xb + b'X'Xb$
  - $= Y'Y - 2b'X'Y + b'X'Xb$
- Điều kiện cần: $\partial(e'e)/\partial b = -2X'Y + 2X'Xb = 0$
  - $\rightarrow X'Xb = X'Y$
  - $\rightarrow b = (X'X)^{-1}(X'Y)$
- Điều kiện đủ: $\partial^2(e'e)/\partial b \partial b = 2X'X > 0$

Dương Nam Hà (VNUA)

8

II. Ước lượng… (tiếp)

- ***Ước lượng HQTT với 2 biến độc lập***
  - Công thức của các Tham số HQ sau khi biến đổi:

$$b_1 = \overline{Y} - b_2\,\overline{X_2} - b_3\,\overline{X_3}$$

$$b_2 = \frac{S_{33}S_{2y} - S_{23}S_{3y}}{S_{22}S_{33} - S_{23}^2}$$

$$b_3 = \frac{S_{22}S_{3y} - S_{23}S_{2y}}{S_{22}S_{33} - S_{23}^2}$$

II. Ước lượng… (tiếp)

- ***Ước lượng HQTT với 2 biến độc lập***
  -

$$S_{22} = \sum_i X_{2i}^2 - n\,\overline{X_2}^2$$

$$S_{23} = \sum_i X_{2i}X_{3i} - n\,\overline{X_2}\,\overline{X_3}$$

$$S_{33} = \sum_i X_{3i}^2 - n\,\overline{X_3}^2$$

$$S_{2y} = \sum_i Y_iX_{2i} - n\,\overline{X_2}\,\overline{Y}$$

$$S_{3y} = \sum_i Y_iX_{3i} - n\,\overline{X_3}\,\overline{Y}$$

$$S_{yy} = \sum_i Y_i^2 - n\,\overline{Y}^2 .$$

II. Ước lượng… (tiếp)

- ***Ước lượng HQTT với 2 biến độc lập***
  - Tính toán các giá trị bình quân của Y, X_2 và X_3;
  - Tính toán tổng bình phương và tổng các tích của các đại lượng sau:
    $\sum_i X^2_{2i}$; $\sum_i X^2_{3i}$; $\sum_i X_{2i} X_{3i}$; $\sum_i Y_i X_{2i}$ và $\sum_i Y_i X_{3i}$;
  - Xác định các giá trị: S_{22}; S_{23}; S_{33}; S_{2y}; S_{3y}; và S_{yy};
  - Ước lượng các tham số b_2 và b_3 thông qua hệ phương trình tuyến tính và các đẳng thức;
  - Thay thế các giá trị của b_2 và b_3 vào phương trình để tìm b_1

Dương Nam Hà (VNUA)

11

II. Ước lượng… (tiếp)

- Ví dụ:
  - Q = Số lít bia được tiêu thụ
  - PB = Giá bia ($)
  - PL = Giá các loại đồ uống có cồn khác ($)
  - PR = Giá các loại thức ăn và dịch vụ còn lại (một chỉ số)
  - I = Thu nhập ($)
  - → $\mathbf{Q = B_0 + B_1PB + B_2PL + B_3PR + B_4I + u}$
  - beer.xls

Dương Nam Hà (VNUA)

12

III. Phân tích phương sai & Kiểm định giả thuyết thống kê

- 3.1 Độ chặt chẽ của MH
  - TSS = ESS + RSS
  - HSXĐ: R^2 = ESS/TSS = 1 - RSS/TSS

$$R^2 = \frac{\text{Phần biến động giải thích được bằng mô hình hồi quy}}{\text{Tổng biến động của biến phụ thuộc}}$$

$$R^2 = \frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y - \bar{Y})^2} = 1 - \frac{\sum(Y - \hat{Y})^2}{\sum(Y - \bar{Y})^2}$$

  - $0 \leq R^2 \leq 1$

Dương Nam Hà (VNUA)

13

III. Phân tích phương sai… (tiếp)

- Chú ý:
  - $R^2 < 0$: MH ƯL không có hằng số
  - R^2 thấp thì đường hồi quy phản ánh Y_i kém (có thể do nhiều nguyên nhân)
  - R^2 nhạy cảm với số lượng biến độc lập trong MH (→chủ quan của người nghiên cứu!!!)
- ➔Hệ số xác định hiệu chỉnh $\overline{R^2}$

$$\overline{R^2} = 1 - \frac{\sum(Y - \hat{Y})^2/(n-k)}{\sum(Y - \bar{Y})^2/(n-1)} = 1 - \frac{(n-1)}{(n-k)}\frac{\sum e^2}{\sum(Y - \bar{Y})^2}$$

  - $0 \leq \overline{R^2} \leq 1$

Dương Nam Hà (VNUA)

14

III. Phân tích phương sai… (tiếp)

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

- Khi k = 1 thì $\overline{R^2}$ = R^2
- Khi k > 1, thì $\overline{R^2}$ < R^2
- Có thể xảy ra $\overline{R^2}$ < 0
- $\overline{R^2}$ đánh giá độ chặt chẽ của mô hình tốt hơn R^2

Dương Nam Hà (VNUA)

15

III. Phân tích phương sai… (tiếp)

- 3.2 KĐ giả thuyết TK
  - 3.2.1 KĐ tham số
  - H_0: $B_j = 0$
  - H_1: $B_j \neq 0$ (hay H_A)
  - Sử dụng kiểm định t: $t_{kđ} = \frac{b_j - B_j}{Se(b_j)} \sim t_{n-k}$
  - Nếu $B_j = 0$ thì:
  - $-t_{(n-k,\,\alpha/2)} \leq b_j/Se(b_j) \leq +t_{(n-k,\,\alpha/2)}$ hay
  - $-t_{(n-k,\,\alpha/2)} \leq t_{kđ} \leq +t_{(n-k,\,\alpha/2)}$
  - Nguyên tắc:
    - $|t_{kđ}| < t_c$ → chấp nhận H_0, bác bỏ H_1
    - $|t_{kđ}| > t_c$ → bác bỏ H_0 hay chấp nhận H_1

Dương Nam Hà (VNUA)

16

III. Phân tích phương sai… (tiếp)

Nguyên tắc kiểm định giả thuyết về b_j

Loại giả thuyết	Giả thuyết không, H_0	Giả thuyết đối, H_1	Miền bác bỏ
Hai phía	$b_j = B_j$	$b_j \neq B_j$	$\lvert t_{kđ} \rvert > t_{\alpha/2(n-k)}$
Phía phải	$b_j \leq B_j$	$b_j > B_j$	$t_{kđ} > t_{\alpha\,(n-k)}$
Phía trái	$b_j \geq B_j$	$b_j < B_j$	$t_{kđ} < -t_{\alpha\,(n-k)}$

III. Phân tích phương sai… (tiếp)

- ƯL khoảng tin cậy

$$t_{kđ} = \frac{b_j - B_j}{Se(b_j)} \sim t_{n-k}$$

  - Lựa chọn t_c tra bảng ($t_{n-k,\,\alpha/2}$) sao cho:
    - $P(-t_c \leq t_{kđ} \leq t_c) = 1 - \alpha$
    - → $P\left(-t_c \leq \frac{b_j - B_j}{Se(b_j)} \leq t_c\right) = 1 - \alpha$
    - → $P(-t_c\, Se(b_j) + b_j \leq B_j \leq b_j + t_c\, Se(b_j)) = 1 - \alpha$
    - ➔ KTC: $b_j - t_c\, Se(b_j) \leq B_j \leq b_j + t_c\, Se(b_j)$
      - với mức ý nghĩa thống kê α

III. Phân tích phương sai… (tiếp)

- 3.2.2 KĐ F
  - H_0: $B_2 = B_3 = ... = B_j = 0$ (hay $R^2 = 0$)
  - H_1: có ít nhất 1 hệ số $B_j \neq 0$ (hay $R^2 \neq 0$)
  - Sử dụng kiểm định F:

$$F_{kđ} = \frac{\sum(\hat{Y}-\bar{Y})^2/(k-1)}{\sum(Y-\hat{Y})^2/(n-k)} = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \sim F^{\lambda}_{k-1,n-k}$$

    - Trong đó: λ là mức độ ý nghĩa thống kê
  - Nguyên tắc:
    - $F_{kđ} < F^{\lambda}_{k-1,n-k}$ → Chấp nhận H_0
    - $F_{kđ} > F^{\lambda}_{k-1,n-k}$ → Bác bỏ H_0, chấp nhận H_1

Dương Nam Hà (VNUA)

19

III. Phân tích phương sai… (tiếp)

Nguồn biến động	Bậc tự do (d.f)	Tổng bình phương (SS)	Tổng bình phương bình quân, MSS = SS/d.f	Giá trị $F_{kđ}$
Do đường HQ (ESS)	k - 1	$\sum(\hat{Y}-\bar{Y})^2$	$MSE = \frac{\sum(\hat{Y}-\bar{Y})^2}{(k-1)}$	$\frac{MSE}{MSR}$
Do sai số (RSS)	n - k	$\sum(Y-\hat{Y})^2 = \sum e^2$	$MSR = \frac{\sum(Y-\hat{Y})^2}{(n-k)}$	
Tổng (TSS)	n - 1	$\sum(Y-\bar{Y})^2$		

Dương Nam Hà (VNUA)

20

III. Phân tích phương sai… (tiếp)

- Trình bày kết quả ước lượng
  - Ví dụ: KQ ƯL lượng tiêu dùng bia (hàm cầu) từ VD trước như sau:
    - $\hat{Y} = 82,16 – 23,74PB – 4,08PL + 12,92PR + 0,002 I$
    - Se (17,96) (5,43) (3,89) (4,16) (0,001)
    - $t_{kđ}$ (4,57) (-4,37) (-1,05) (3,10) (2,57)
    - $R^2 = 0,82$; $F_{kđ} = 28,89$
    - n = 30

III. Phân tích phương sai… (tiếp)

- Đánh giá kết quả phân tích HQ
  - Dấu của các tham số ƯL so với lý thuyết (kì vọng ban đầu)
  - Xem xét mức ý nghĩa thống kê của th/số ƯL
  - MHHQ giải thích biến động của Y tốt ntn?
  - Liệu MH(ƯL) có thỏa mãn các giả thiết CLRM
    - Chú ý: liệu u_i có tuân theo PP chuẩn? (đồ thị, JB test)
    - Kiểm định t và F yêu cầu sai số phải tuân theo PP chuẩn, nếu không thì quá trình KĐ sẽ không thỏa mãn với các mẫu có hạn (hoặc nhỏ)

III. Phân tích phương sai… (tiếp)

- Jarque Bera test:

$$JB = \frac{n}{6}\left[S^2 + \frac{1}{4}(K-3)^2\right] \sim \chi_2^2$$

- Trong đó: $S = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{3/2}}$

  $K = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{2}},$

- Giả thuyết: H_0: JB = 0 (tức S = 0; K = 3)
  H_1: JB ≠ 0
- JB > χ_2^2 → Bác bỏ H_0

Dương Nam Hà (VNUA)

23