Ước lượng các hệ số của mô hình hồi quy theo phương pháp bình phương tối thiểu

Bởi:

Phạm Trí Cao

Các giả định của mô hình hồi quy tuyến tính cổ điển

Các giả định về sai số hồi quy như sau đảm bảo cho các ước lượng hệ số hàm hồi quy tổng thể dựa trên mẫu theo phương pháp bình phương tối thiểu là ước lượng tuyến tính không chệch tốt nhất(BLUE).

Giá trị kỳ vọng bằng 0:

$$E\left[\varepsilon_i \middle| X_i\right] = 0$$

Phương sai không đổi:

$$\mathrm{var}\left[\varepsilon_i \middle| X_i\right] = E\left[\varepsilon_i^2 \middle| X_i\right] = \sigma^2$$

Không tự tương quan:

$$\mathrm{cov}\left[\varepsilon_i \varepsilon_j \middle| X_i, X_j\right] = E\left[\varepsilon_i \varepsilon_j \middle| X_i, X_j\right] = 0$$

Không tương quan với X:

$$\mathrm{cov}\left[\varepsilon_i X_j \middle| X_i, X_j\right] = E\left[\varepsilon_i X_j \middle| X_i, X_j\right] = 0$$

Có phân phối chuẩn:

$$\varepsilon_i = N(0, \sigma^2)$$

Ở chương 5 chúng ta sẽ khảo sát hậu quả khi các giả thiết trên bị vi phạm.

Phương pháp bình phương tối thiểu:

Ý tưởng của phương pháp bình phương tối thiểu là tìm $\hat{\beta}_1$ và $\hat{\beta}_2$ sao cho tổng bình phương phần dư có giá trị nhỏ nhất.

Từ hàm hồi quy (3.5)

$$e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i$$

Vậy

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right)^2$$

(3.6)

Điều kiện để (3.6) đạt cực trị là:

$$\frac{\partial\left(\sum_{i=1}^{n} e_i^2\right)}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right) = -2\sum_{i=1}^{n} e_i = 0 \; (3.7)$$

$$\frac{\partial\left(\sum_{i=1}^{n} e_i^2\right)}{\partial \hat{\beta}_2} = -2\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right)X_i = -2\sum_{i=1}^{n} e_i X_i = 0 \; (3.8)$$

Từ (3.7) và (3.8) chúng ta rút ra

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i \; (3.9)$$

$$\sum Y_i X_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2 \; (3.10)$$

Các phương trình (3.9) và (3.10) được gọi là các phương trình chuẩn. Giải hệ phương trình chuẩn ta được

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \; (3.11)$$

Thay (3.9) vào (3.8) và biến đổi đại số chúng ta có

Ước lượng các hệ số của mô hình hồi quy theo phương pháp bình phương tối thiểu

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})(X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \quad (3.12)$$

Đặt $x_i = X_i - \overline{X}$ và $y_i = Y_i - \overline{Y}$ ta nhận được

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2} \quad (3.13)$$

Tính chất của hàm hồi quy mẫu theo OLS

Tính chất của tham số ước lượng

$\hat{\beta}_1$ và $\hat{\beta}_2$ là duy nhất ứng với một mẫu xác định gồm n quan sát (Xi,Yi).

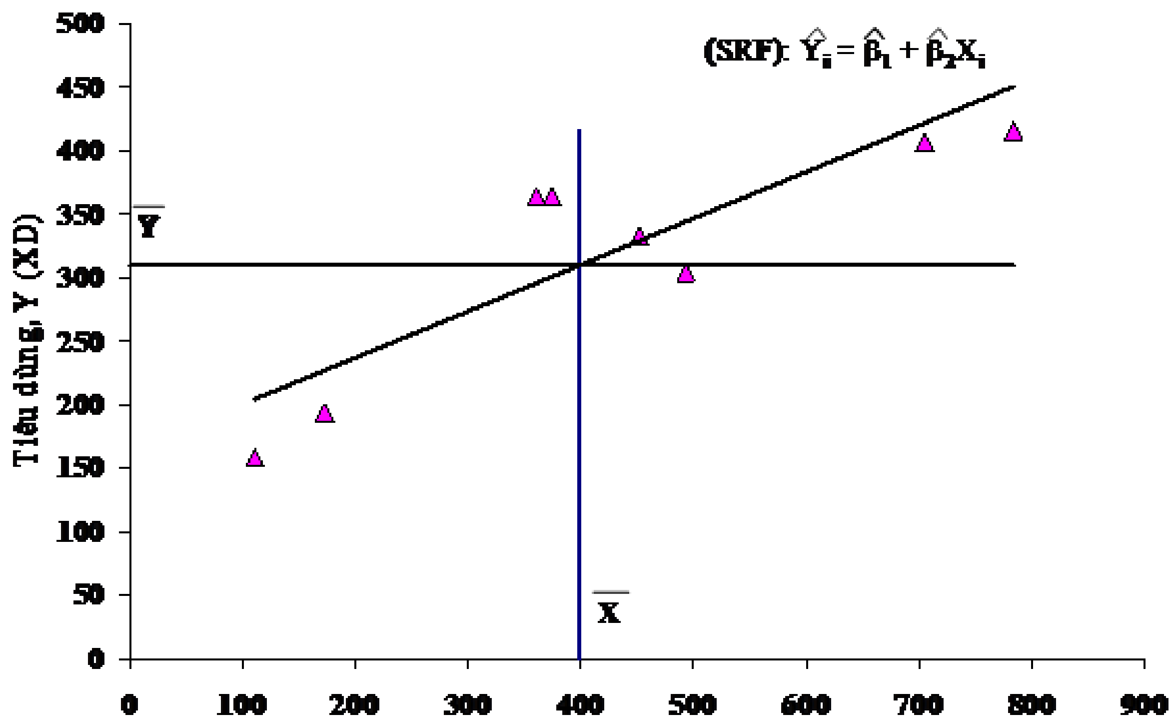
$\hat{\beta}_1$ và $\hat{\beta}_2$ là các ước lượng điểm của $\hat{\beta}_1$ và $\hat{\beta}_2$. Giá trị của $\hat{\beta}_1$ và $\hat{\beta}_2$ thay đổi theo mẫu dùng để ước lượng.

Tính chất của hàm hồi quy mẫu

Phần chứng minh các tính chất ở phần này có thể tìm đọc ở Gujarati, Basic Econometrics,3rd Edition, p56-59.

Hàm hồi quy mẫu đi qua giá trị trung bình của dữ liệu

Thật vậy, từ (3.11) ta có $\overline{Y} = \hat{\beta}_1 - \hat{\beta}_2\overline{X}$



Thu nhập X (XD)

Hình 3.4. Đường hồi quy mẫu đi qua giá trị trung bình của dữ liệu

Giá trị trung bình của ước lượng bằng giá trị trung bình của quan sát đối với biến phụ thuộc: $E\left(\hat{Y}\right) = \bar{Y}$.

Giá trị trung bình của phần dư bằng 0: $E(e_i) = 0$

Các phần dư e_i và Y_i không tương quan với nhau:

$$\sum_{i=1}^{n} e_i Y_i = 0$$

Các phần dư e_i và X_i không tương quan với nhau:

$$\sum_{i=1}^{n} e_i X_i = 0$$

Phân phối của $\hat{\beta}_1$ và $\hat{\beta}_2$

Có thể tính toán chứng minh các biểu thức này dựa vào các định nghĩa và định lý về kỳ vọng và phương sai. Tham khảoVũ Thiếu và đồng sự, Kinh tế lượng, PL chương 2, trang 61.

Ước lượng $\hat{\beta}_1 \hat{\beta}_2$

Kỳ vọng $E\left(\hat{\beta}_1\right) = \beta_1 E\left(\hat{\beta}_2\right) = \beta_2$

Phương sai

$$\operatorname{var}\left(\hat{\beta}_1\right) = \frac{\sum_{i=1}^{n} X_i^2}{n\sum_{i=1}^{n} x_i^2}\sigma^2$$

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

Sai số chuẩn

$$\sigma_{\hat{\beta}_1} = \sqrt{\frac{\sum_{i=1}^{n} X_i^2}{n\sum_{i=1}^{n} x_i^2}}\sigma \; \sigma_{\hat{\beta}_2} = \frac{\sigma}{\sqrt{\sum_{i=1}^{n} x_i^2}}$$

Phân phối

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sum_{i=1}^{n} X_i^2}{n\sum_{i=1}^{n} x_i^2}\sigma^2\right) \hat{\beta}_2 \sim N\left(\beta_2, \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}\right)$$

Hiệp phương sai của hai hệ số ước lượng

$$\operatorname{cov}\left(\hat{\beta}_1, \hat{\beta}_2\right) = -\bar{X}\operatorname{var}\left(\hat{\beta}_2\right) = -\bar{X}\left(\frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}\right)$$

Trong các biểu thức trên $\sigma^2 = \operatorname{var}(\varepsilon_i)$ với giả định $\varepsilon_i \sim N(0,\sigma^2)$