Chương 2: MÔ HÌNH HỒI QUY BỘI

1. SỰ CẦN THIẾT CỦA MÔ HÌNH HỒI QUY BỘI
2. PHƯƠNG PHÁP ƯỚC LƯỢNG OLS
3. ĐỊNH LÝ GAUSS – MARKOV VÀ TÍNH VỮNG CỦA ƯỚC LƯỢNG OLS
4. MỘT SỐ DẠNG CỦA MÔ HÌNH HỒI QUY BỘI
5. MÔ HÌNH HỒI QUY SỬ DỤNG NGÔN NGỮ MA TRẬN

2.1.1 Mô hình 2 biến, vấn đề kỳ vọng sai số khác 0.

Nếu Cov(X, U) ≠ 0 thì E(U|X) ≠ 0 (Vi phạm GT 2).

mm

Ví dụ: Mô hình hồi quy về tác động của thu nhập (TN) lên chi tiêu (CT) của mỗi hộ gia đình là: $\mathrm{CT_i} = \beta_1 + \beta_2 \mathrm{TN_i} + \mathrm{U_i}$.

- Theo Friedman (Permanent income hypothesis), tài sản (TS) tác động đáng kể đến mức chi tiêu; nên U có chứa yếu tố TS.
- $\mathrm{Cov(TS, TN)} \neq 0$.
- Vì thế cần thiết đưa thêm biến TS vào mô hình hồi quy:

$$\mathrm{CT_i} = \beta_1 + \beta_2 \mathrm{TN_i} + \beta_3 \mathrm{TS_i} + \mathrm{U_i}.$$

2.1.1 Mô hình 2 biến, vấn đề kỳ vọng sai số khác 0.

Nếu	GT2 bị vi phạm, do đó các ước lượng OLS bị chệch.
Cov(X,U)≠0	Biến độc lập X có tương quan với sai số ngẫu nhiên được gọi là *biến độc lập nội sinh.*
thì	Giải quyết vấn đề *biến độc lập nội sinh* bằng cách đưa thêm biến quan trọng khác vào mô hình.

2.1.2 Một số ưu điểm khác

- Chất lượng dự báo tốt hơn, cung cấp các dự báo hữu ích hơn.
- Cho phép sử dụng dạng hàm phong phú hơn.
- Cho phép thực hiện các phân tích phong phú hơn

  -Có thể phân tích tác động riêng phần của biến độc lập lên biến phụ thuộc.

  -Có thể phân tích tác động đồng thời của nhiều biến độc lập lên biến phụ thuộc.

2.2.1 Mô hình và các giả thiết

Hàm hồi quy tuyến tính k biến $Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + U$

Trong đó: **Y:** biến phụ thuộc, $X_2, .., X_k$: các biến độc lập.

U: Sai số ngẫu nhiên, đại diện cho các yếu tố có tác động đến Y nhưng không đưa vào mô hình.

β_1 : là giá trị trung bình của Y khi tất cả các biến độc lập bằng 0.

β_j (j=2… k): hệ số hồi quy riêng (partial coefficient) : thể hiện tác động riêng phần của biến X_j lên giá trị trung bình của biến phụ thuộc khi các yếu tố X_s (s≠j) không đổi.

2.2.1 Mô hình và các giả thiết

Các giả thiết của mô hình hồi quy bội

GT1: Mô hình được ước lượng trên cơ sở mẫu ngẫu nhiên.

GT2: $E(U_i|_{X_{2i},\dots,X_{ki}}) = 0$ tại mỗi giá trị $(X_{2i}, \dots, X_{ki})$.

GT3: Phương sai của U_i tại mỗi giá trị $(X_{2i},\dots, X_{ki})$ đều bằng nhau: $Var(U_i|_{X_{2i},\dots,X_{ki}}) = \sigma^2 \;\forall i$

GT4: Các biến độc lập X_j không có quan hệ đa cộng tuyến hoàn hảo. Nghĩa là không tồn tại các hằng số $\lambda_2, \dots, \lambda_k$ không đồng thời bằng 0 sao cho $\lambda_2 X_2 + \dots + \lambda_k X_k = 0$.

2.2.2 Nội dung phương pháp OLS:

* Xét mô hình hồi quy tổng thể: $Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + U$

Cần ước lượng các hệ số $\beta_1, \beta_2, \ldots, \beta_k$ bằng một mẫu cụ thể kích thước n $\{(Y_i, X_{2i}, \ldots, X_{ki})\ (i=1,..n)\}$.

Nội dung phương pháp OLS:

Tìm hàm hồi quy mẫu: $\hat{Y} = \widehat{\beta_1} + \widehat{\beta_2} X_2 + \ldots + \widehat{\beta_k} X_k$

hay $Y_i = \widehat{\beta_1} + \widehat{\beta_2} X_{2i} + \ldots + \widehat{\beta_k} X_{ki} + e_i$ **sao cho** $\sum_{i=0}^{n} e_i^2$ **Min**

2.2.2 Nội dung phương pháp OLS:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n} \left(Y_i - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki}\right)^2 \to Min$$

Giải bài toán cực trị hàm k biến, tìm $\widehat{\boldsymbol{\beta}_1}$, $\widehat{\boldsymbol{\beta}_2}$,... $\widehat{\boldsymbol{\beta}_k}$, ta có hệ:

$$\begin{cases} \sum_{i=1}^{n} \left(Y_i - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki}\right) = 0 \\ \sum_{i=1}^{n} X_{2i}\left(Y_i - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki}\right) = 0 \\ \ldots\ldots\ldots \\ \sum_{i=1}^{n} X_{ki}\left(Y_i - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki}\right) = 0 \end{cases}$$

2.2.3 Độ phù hợp của hàm hồi quy

Tương tự mô hình hồi quy 2 biến, ký hiệu:

$$TSS = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

$$RSS = \sum_{i=1}^{n} e_i^2$$

$$ESS = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2$$

- Khi mô hình có hệ số chặn thì

TSS = ESS + RSS

- **Hệ số xác định của mô hình:**

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

- $0 \le R^2 \le 1$
- $R^2 = r^2(Y, \hat{Y})$

*Ý nghĩa hệ số xác định R^2:

1

- R^2 là tỷ lệ sự biến động của biến phụ thuộc Y được giải thích bởi các biến độc lập trong mô hình (theo mô hình, với số liệu mẫu).

2

- R^2 thể hiện mức độ tương quan tuyến tính giữa biến phụ thuộc với các biến độc lập.

3

- Việc đưa thêm biến số bất kỳ vào mô hình hồi quy sẽ làm gia tăng R^2, bất kể biến đó có giải thích thêm biến phụ thuộc hay không.

Hệ sốxácđịnhhiệuchỉnh (adjusted R – square)

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

- $\overline{R^2}$ có ý nghĩatươngtựnhư R^2.
- $\overline{R^2}$ cóthểnhậngiátrịâm.
- $\overline{R^2}$ Thườngđượcsửdụngthaycho R^2 khi so sánhmôhìnhhồiquycósốlượngbiếnsốkhácnhau, vàcùngbiếnphụthuộc.

2.2.4 Tính tốt nhất của ước lượng OLS

Độ chính xác của các ước lượng

Khi các GT1 – 4 thỏa mãn thì:

$$var(\hat{\beta}_j) = \frac{\sigma^2}{(1 - R_j^2)\sum_{i=1}^{n} x_{ji}^2};$$

$$se(\hat{\beta}_j) = \sqrt{\frac{\sigma^2}{(1 - R_j^2)\sum_{i=1}^{n} x_{ji}^2}}$$

Trong đó:

- R_j^2 là hệ số xác định của mô hình hồi quy biến X_j theo hệ số chặn và các biến độc lập còn lại.
- σ^2 chưa biết, có ước lượng (không chệch) là $\hat{\sigma}^2$, với:

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

2.3 ĐỊNH LÝ GAUSS – MARKOV VÀ TÍNH VỮNG CỦA ƯỚC LƯỢNG OLS

2.3.1 Tính tốt nhất của ước lượng OLS

Định lý Gauss – Markov:

Khi các GT1 – 4 thỏa mãn thì các ước lượng thu được từ phương pháp OLS là các ước lượng tuyến tính **không chệch và có phương sai nhỏ nhất** trong lớp các ước lượng tuyến tính không chệch.

2.3.2 Tính vững của ước lượng OLS

- Khi các GT 1 – 4 thỏa mãn thì các ước lượng OLS là ước lượng vững.
- Nếu các GT 1, GT3, 4 thỏa mãn và :
  - $Cov(X_j, U) = 0$ với mọi $j = 2, \ldots, k$.
  - $E(U) = 0$

thì ước lượng OLS vẫn là ước lượng vững.

2.4 MỘT SỐ DẠNG CỦA MÔ HÌNH HỒI QUY

2.4.1 Mô hình dạng log - log.

$$\ln(Y) = \beta_1 + \beta_2 \ln(X_2) + \ldots + \beta_k \ln(X_k) + U$$

Giả sử X_j thay đổi, tất cả các biến còn lại không đổi. Lấy vi phân hai vế, ta có: $\frac{dY}{Y} = \beta_j \frac{dX_j}{X_j}$

Ý nghĩa β_j : ***Khi các biến độc lập khác không đổi, X_j tăng 1% thì trung bình của Y thay đổi β_j % (j = 2… k).***

- Mô hình log – log được sử dụng để mô tả các mối quan hệ mà hệ số co giãn riêng không đổi.

2.4.2 Mô hình dạng log - lin.

$$Ln(Y) = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + U \quad (*)$$

Ý nghĩa β_j : ***Khi các biến độc lập khác không đổi, X_j tăng 1 đơn vị thì trung bình của Y thay đổi $100\beta_j$ % ($j = 2\ldots k$).***

Ví dụ: Hàm hồi quy thể hiện quan hệ giữa thu nhập (TN – đơn vị: triệu đồng/tháng) và trình độ học vấn (HV – đơn vị: số năm đi học):

$$Ln(TN) = 1,08 + 0,24.HV + e.$$

Hãy giải thích ý nghĩa của các hệ số ước lượng.

2.4.2 Mô hình dạng bán lin - log.

$$Y = \beta_1 + \beta_2 \ln(X_2) + \ldots + \beta_k \ln(X_k) + U \quad (**)$$

Ý nghĩa β_j : ***Khi các biến độc lập khác không đổi, X_j tăng 1 % thì trung bình của Y thay đổi (β_j /100) đơn vị (j = 2… k).***

Ví dụ: Mô hình hồi quy thể hiện quan hệ giữa số giờ mà người lao động muốn làm (L) và mức trả cho một giờ lao động (TL-Đơn vị USD) là: L = 2 + 33Ln(TL) + e

Hãy giải thích ý nghĩa của các hệ số hồi quy.

Sử dụng dạng hàm có biến số dạng logarit khi	
	Có lý thuyết kinh tế về mối quan hệ giữa các biến số.
	Các biến nhận giá trị dương như dân số, GDP, tổng tài sản, giá cổ phiếu, số lao động…
	Các biến số có phân phối đuôi lệch như tiền lương, thu nhập …

Mục đích:

- Phân tích sự thay đổi của các biến số dưới dạng thay đổi tương đối.

- Làm cho phân phối của sai số ngẫu nhiên gần hơn với phân phối chuẩn.

Ví dụ:

Hãy cho biết ý nghĩa các hệ số ước lượng của các mô hình hồi quy sau đây.

a) $\hat{Y} = 10 + 12,4X_2 - 3,6Ln(X_3)$

b) $\hat{Y} = -0,3 + 0,2Ln(X_2) - 1,3Ln(X_3)$.

c) $Ln(\hat{Y}) = 2 - 0,9X_2 + 4,6.Ln(X_3)$.

2.4.3 Mô hình dạng đa thức.

$$Y = \beta_1 + \beta_2 X + \beta_3 X^2 \ldots + \beta_k X^{k-1} + U \quad (**)$$

Ví dụ: Mô hình hồi quy thể hiện quan hệ giữa tiền lương (W) và số tuổi lao động (Age: số năm đi làm) là:

$$W = \beta_1 + \beta_2 Age + \beta_3 Age^2 + U$$

Việc đưa thêm thành phần Age^2 vào mô hình nhằm thể hiện tác động của quy luật giá trị cận biên của năng suất lao động theo số năm đi làm giảm dần khi tuổi lao động vượt quá một ngưỡng nào đó.

Hãy phân tích tác động của mức tuổi đến tiền lương.

2.5 MÔ HÌNH HỒI QUY SỬ DỤNG NGÔN NGỮ MA TRẬN

Xét mô hình hồi quy tuyến tính k biến

$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + U$.

Với mẫu ngẫu nhiên kích thước n, có thể viết

$$\begin{cases} Y_1 = \beta_1 + \beta_2 X_{21} + \ldots + \beta_k X_{k1} + U_1 \\ Y_2 = \beta_1 + \beta_2 X_{22} + \ldots + \beta_k X_{k2} + U_2 \\ \ldots \\ Y_n = \beta_1 + \beta_2 X_{2n} + \ldots + \beta_k X_{kn} + U_n \end{cases}$$

Dạng ma trận của hệ phương trình này: **Y = X.β + U**, với

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_n \end{pmatrix}_{n \times 1} ; X = \begin{pmatrix} 1 & X_{21} & \ldots & X_{k1} \\ 1 & X_{22} & \ldots & X_{k2} \\ \ldots & \ldots & & \\ 1 & X_{2n} & \ldots & X_{kn} \end{pmatrix}_{n \times k} ; \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \ldots \\ \beta_k \end{pmatrix}_{k \times 1} ; U = \begin{pmatrix} U_1 \\ U_2 \\ \ldots \\ U_n \end{pmatrix}_{n \times 1}$$

Phương pháp OLS cho kết quả ước lượng:

$$\beta = (X^T X)^{-1} X^T Y$$

Lưu ý: GT 4 tương đương với điều kiện ma trận ($X^T X$) khả đảo.

- Ma trận hiệp phương sai giữa các hệ số ước lượng:

$$Var(\hat{\beta} \mid X) = \sigma^2 (X^T X)^{-1}$$

- Hệ số xác định của mô hình:

$$TSS = Y^T.Y - n(\overline{Y})^2.$$

$$ESS = \hat{\beta} X^T Y - n(\overline{Y})^2$$

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