CHƯƠNG 3
HỒI QUY TUYẾN TÍNH ĐA BIẾN

TS. Đinh Thị Thanh Bình - Khoa Kinh Tế Quốc Tế-
Đại Học Ngoại Thương- Hà Nội

Mô hình hồi quy tuyến tính đa biến

- Trong thực tế, các mối quan hệ kinh tế thường phức tạp, một số biến số kinh tế có thể chịu tác động của nhiều biến số kinh tế khác → mô hình hồi quy hai biến (hồi quy đơn) tỏ ra không thỏa đáng.
- Vì vậy cần thiết phải mở rộng mô hình hồi quy hai biến bằng cách đưa thêm nhiều biến vào mô hình → n/c hồi quy nhiều biến (hồi quy bội hay hồi quy đa biến)
- Các ý tưởng và kết quả nghiên cứu của hồi quy hai biến được khái quát cho mô hình hồi quy nhiều biến.

3.1. Các giả thiết cơ bản của mô hình

Giả thiết 1: Trong mô hình tổng thể Y có mối quan hệ với các biến X và u:

$$Y = \beta_0 + \beta_1 X + ... + \beta_k X_k + u$$

Giả thiết 2: Mẫu điều tra là mẫu ngẫu nhiên, kích cỡ n.

Giả thiết 3: X có các giá trị không đồng nhất, và các biến độc lập không có mối quan hệ tuyến tính hoàn hảo (no perfect collinearity).

Giả thiết 4: Đại lượng sai số ngẫu nhiên (nhiễu) có kỳ vọng bằng 0, tức là: **E(u/X)=0**.

Định lý 1: Ước lượng không chệch của các tham số

Với các giả thiết 1-4 trên, ta có:

$$E(\hat{\beta}_j) = \beta_j, \; j = 0,1,\ldots,k$$

Giả thiết 5: Các u_i có phương sai thuần nhất (homoscedasticity), tức là các u_i có phương sai giống nhau với bất kỳ giá trị nào của X_i

$$var\ (u_i/X_i) = E[u_i - E(u_i/X_i)]^2 = E(u_i^2/X_i) = \sigma^2$$

Định lý 3: Phương sai của các ước lượng

Với các giả thiết 1-5, ta có:

$$\mathrm{Var}(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1-R_j^2)}$$

j =1,2,….,k; $SST_j = \sum_{i=1}^{n}(X_{ij} - \overline{X}_j)^2$

R_j^2 là R^2 từ hồi qui X_j lên các biến độc lập khác

$\mathrm{var}(\beta_j)$ lớn hơn → ước lượng thiếu chính xác hơn→khoảng tin cậy lớn hơn → kiểm định giả thuyết thống kê kém chính xác hơn.

3.2. Mô hình hồi quy 3 biến

- Mô hình hồi quy tổng thể PRF

$$E(Y / X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Ý nghĩa: PRF cho biết trung bình có điều kiện của Y với điều kiện đã biết các giá trị cố định của biến X_1 và X_2.

- Y: biến phụ thuộc
- X_1 và X_2: biến độc lập
- β_0 : hệ số tự do
- β_1 , β_2 : hệ số hồi quy riêng

3.2. Mô hình hồi quy 3 biến

- Mô hình hồi quy tổng thể ngẫu nhiên:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

u_i: sai số ngẫu nhiên của tổng thể

3.3. Ước lượng các tham số

Hàm hồi quy mẫu SRF:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$$

Phần dư của mẫu ứng với quan sát thứ i

$$u_i = Y_i - \hat{Y}_i$$

Sử dụng **phương pháp bình phương nhỏ nhất OLS** để ước lượng các tham số $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$

3.4. Cách diễn giải hệ số hồi qui riêng

Ý nghĩa hệ số hồi quy riêng*: cho biết ảnh hưởng của từng biến độc lập lên giá trị trung bình của biến phụ thuộc khi các biến còn lại được giữ không đổi.*

Ví dụ: $salary = 1.29 - 0.43\,female + 0.83\,educ$

Với điều kiện là các yếu tố khác không đổi (ceteris paribus), nữ giới có thu nhập thấp hơn nam giới là 43 cent/ giờ.

3.5. Mô hình hồi quy k biến

Mô hình hồi quy tổng thể

$$E(Y / X_1,...X_k) = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki}$$

Mô hình hồi quy mẫu ngẫu nhiên:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + ... + \hat{\beta}_k X_{ki}$$

sai số của mẫu ứng với quan sát thứ I

$$u_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - ... - \hat{\beta}_k X_{ki}$$

Sử dụng **phương pháp bình phương nhỏ nhất OLS** để ước lượng các tham số.

3.6. CÁC TỔNG BÌNH PHƯƠNG ĐỘ LỆCH

- SST (Total Sum of Squares - Tổng bình phương sai số tổng cộng)

$$SST = \sum (Y_i - \overline{Y})^2$$

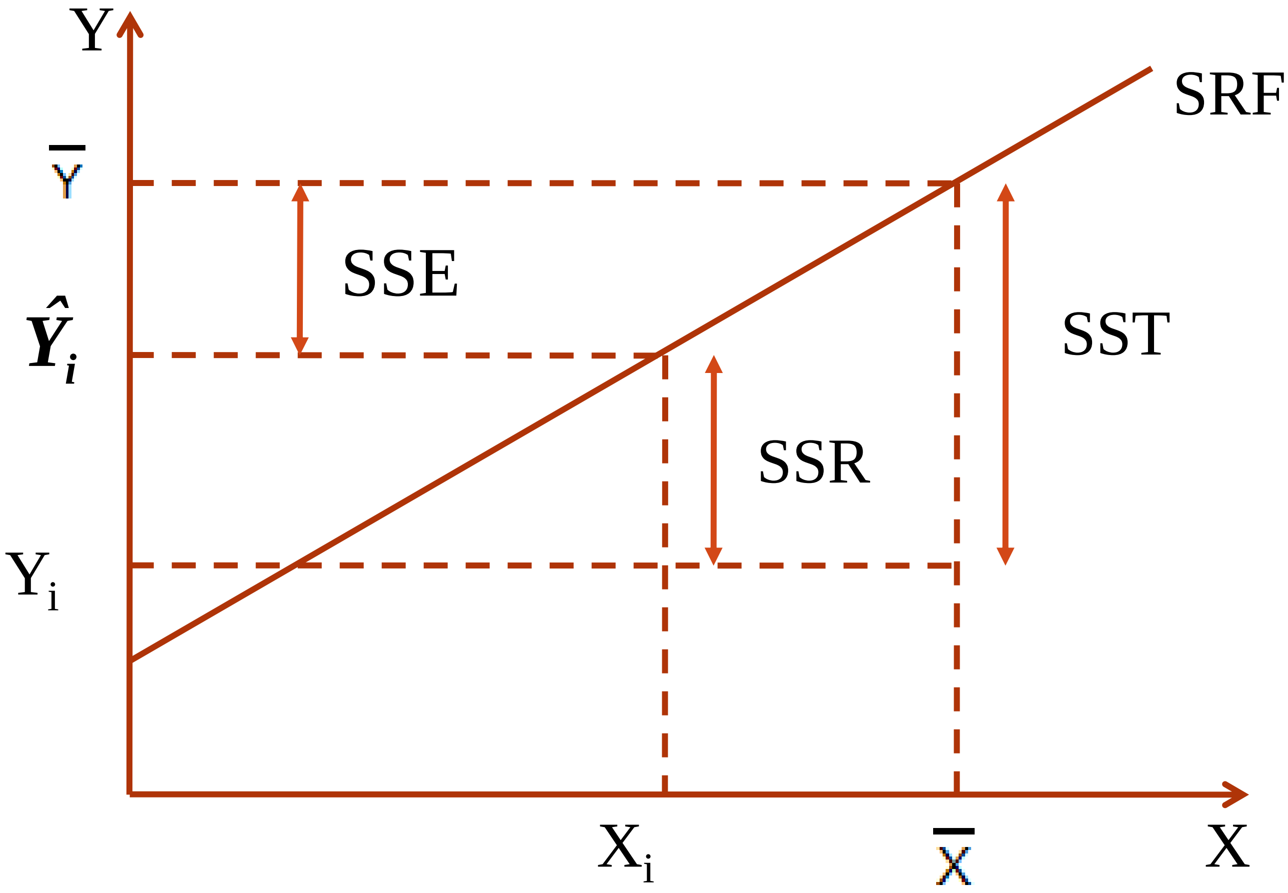
- SSE: (Explained Sum of Squares - Bình phương sai số được giải thích)

$$SSE = \sum (\hat{Y}_i - \overline{Y})^2$$

- SSR: (Residual Sum of Squares - Tổng bình phương các phần dư)

$$SSR = \sum_{i=1}^{n} \left(Y_i - Y_i \right)^2 = \sum u_i^2$$

3.6. CÁC TỔNG BÌNH PHƯƠNG ĐỘ LỆCH



Hình 2.3: Ý nghĩa hình học của SST, SSR và SSE

3.7. HỆ SỐ XÁC ĐỊNH R²

Ta chứng minh được: SST = SSE + SSR

$$1 = \frac{SSE}{SST} + \frac{SSR}{SST}$$

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

$$R^2 = \frac{[\sum (Y_i - \overline{Y})(Y_i - \overline{Y})]^2}{[\sum (Y_i - \overline{Y})^2][\sum (Y_i - \overline{Y})^2]}$$

3.7. TÍNH CHẤT CỦA HỆ SỐ XÁC ĐỊNH R^2

$0 \leq R^2 \leq 1$

Cho biết % sự biến động của Y được giải thích bởi các biến số X trong mô hình.

$R^2 = 1$: đường hồi quy phù hợp hoàn hảo

$R^2 = 0$: X và Y không có quan hệ

Nhược điểm: R^2 tăng khi số biến X đưa vào mô hình tăng, dù biến đưa vào không có ý nghĩa.

=>Sử dụng R^2 điều chỉnh (adjusted R^2 , $\bar{R}^2$) để quyết định đưa thêm biến vào mô hình.

3.8. Hệ số xác định điều chỉnh

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

Dùng $\bar{R}^2$ để xem xét việc đưa thêm biến vào mô hình. Biến mới đưa vào mô hình phải thỏa 2 điều kiện:

- Làm $\bar{R}^2$ tăng
- Biến mới có ý nghĩa thống kê trong mô hình mới

3.9. Các tính chất của hệ số ước lượng OLS (cont.)

- $\hat{\beta}$ được xác định một cách duy nhất với một mẫu cụ thể.
  - $\hat{\beta}$ là ngẫu nhiên. Với các mẫu khác nhau, giá trị cụ thể của chúng sẽ khác nhau.
  - Với giả thiết u có phân phối chuẩn, véc tơ $\hat{\beta}$ tuân theo quy luật chuẩn.

3.9. Các tính chất của hệ số ước lượng OLS

- Với các giả thiết của mô hình, hàm hồi quy mẫu ước lượng theo PP OLS có các tính chất tương tự như trong trường hợp hồi quy hai biến, bao gồm các tính chất sau:
  - SRF đi qua điểm ứng với các giá trị trung bình ($\bar{Y}, \bar{X}_1, \ldots, \overline{X}_k$)
  - $\bar{Y} = \overline{\hat{Y}}$
  - $\overline{\hat{u}} = 0$
  - $\hat{u}_i$ không tương quan với $X_{1i}, \ldots, X_{k,i}$, tức là cov($\hat{u}$,X) = 0
  - $\hat{u}_i$ không tương quan với $\hat{Y}_i$, tức là cov($\hat{u}$, $\hat{Y}$) = 0

3.10. Tiêu chuẩn của các ước lượng OLS- Định lý Gauss- Markov

- Với các giả thiết của phương pháp bình phương nhỏ nhất, các ước lượng bình phương nhỏ nhất thu được có tiêu chuẩn tốt nhất.
- Các tiêu chuẩn này được biết đến thông qua định lý nổi tiếng **Gauss- Markov**.

3.10. Tiêu chuẩn của các ước lượng OLS- Định lý Gauss- Markov

- Một ước lượng, ví dụ như ước lượng theo phương pháp OLS, được gọi là ước lượng tuyến tính không chệch tốt nhất (Best Linear Unbiased Estimator- BLUE) của β nếu nó thỏa mãn các tiêu chuẩn sau đây :
  - **Tuyến tính**: khi các ước lượng là hàm tuyến tính của một biến ngẫu nhiên, chẳng hạn như biến phụ thuộc Y trong mô hình hồi quy.
  - **Không chệch**: tức là giá trị trung bình của ước lượng hay chính là giá trị kỳ vọng của nó, E ($\hat{\beta}_2$), bằng với giá trị thực β_2.
  - **Có phương sai nhỏ nhất** trong lớp các ước lượng tuyến tính không chệch. Một ước lượng không chệch có phương sai nhỏ nhất được coi là một ước lượng hiệu quả.

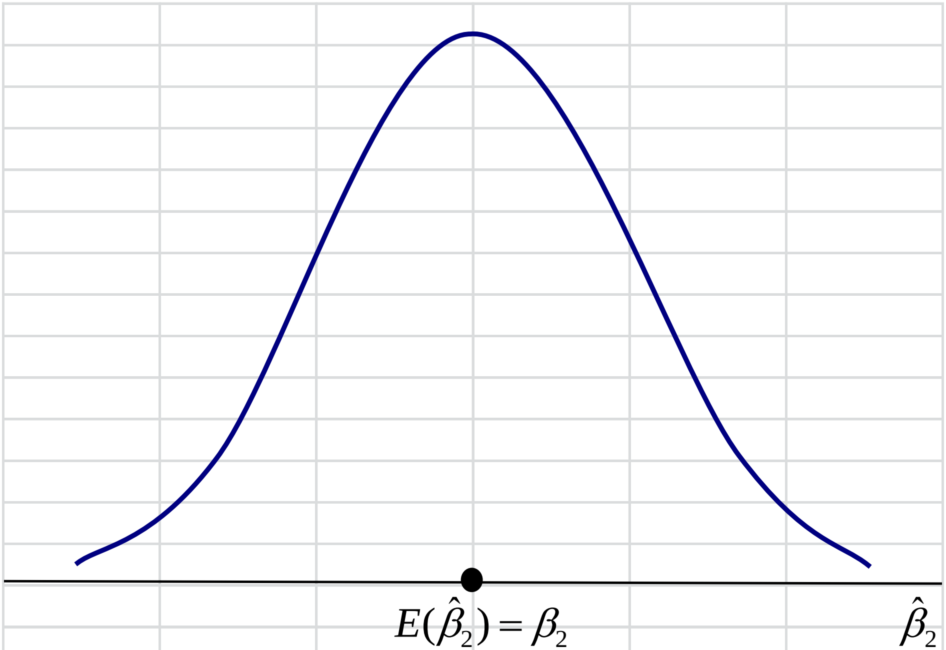
3.10. Tiêu chuẩn của các ước lượng OLS- Định lý Gauss- Markov

- Đối với mô hình hồi quy, thì các ước lượng theo phương pháp OLS được coi là các ước lượng BLUE. Đây chính là nội dung của định lý Gauss- Markov nổi tiếng, được phát biểu như sau:
- ***<u>Định lý Gauss- Markov</u>: Với các giả thiết của phương pháp bình phương nhỏ nhất, các ước lượng thu được là các ước lượng tuyến tính, không chệch và có phương sai nhỏ nhất (BLUE) trong lớp các ước lượng tuyến tính không chệch.***
- Định lý Gauss- Markov có thể được giải thích thông qua các đồ thị phân bố xác suất trong hình [3.07].

3.10. Tiêu chuẩn của các ước lượng OLS- Định lý Gauss- Markov

Hình 3.07. Phân phối mẫu của ước lượng $\hat{\beta}_2$ (OLS) và β_2^* (phương pháp khác)

- Hình 3.07 (a) mô tả phân phối mẫu của ước lượng theo phương pháp OLS. Để thuận tiện, ta giả định rằng đồ thị phân bố xác suất của là đối xứng. Đồ thị này cho ta thấy trung bình các giá trị E($\hat{\beta}_2$) bằng với giá trị thực của β_2. Trong trường hợp này, ta nói rằng $\hat{\beta}_2$ là ước lượng **không chệch** của β_2.

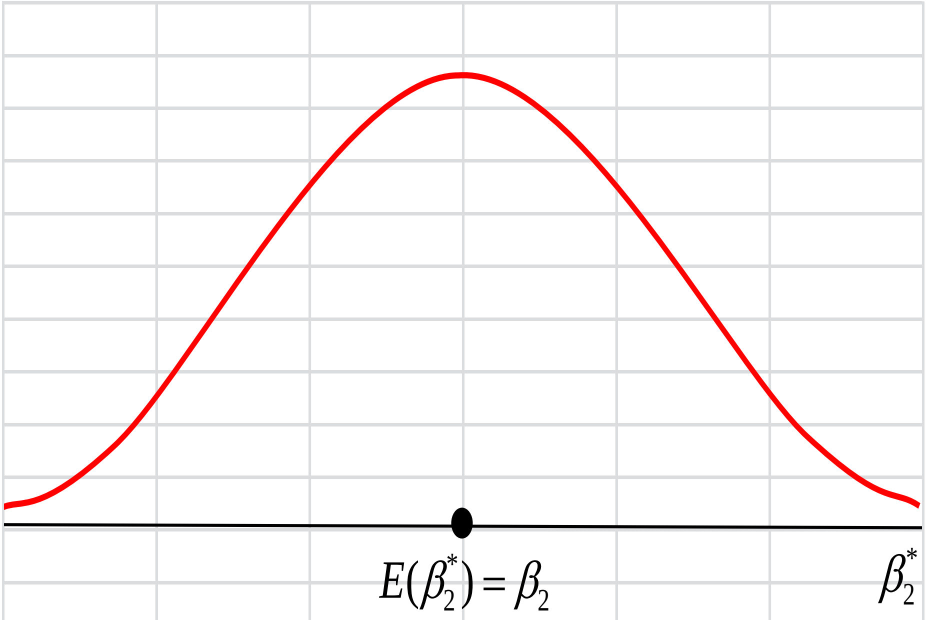


(a) Phân phối mẫu của β2

3.10. Tiêu chuẩn của các ước lượng OLS- Định lý Gauss- Markov

Hình 3.07. Phân phối mẫu của ước lượng $\hat{\beta}_2$ (OLS) và β_2^* (phương pháp khác)

- Hình 3.07 (b) biểu diễn phân phối mẫu của ƯL β_2^*, một giá trị ƯL của β_2 thu được bằng một phương pháp khác OLS. Giả định β_2^*, giống $\hat{\beta}_2$, là ƯL không chệch, nghĩa là giá trị TB của nó bằng giá trị của β_2. Ngoài ra, cũng giả định β_2^* và $\hat{\beta}_2$ đều là các ƯL tuyến tính, tức là chúng đều là hàm tuyến tính của biến phụ thuộc Y → giữa hai ƯL β_2^* và $\hat{\beta}_2$, ta chọn ước lượng nào?

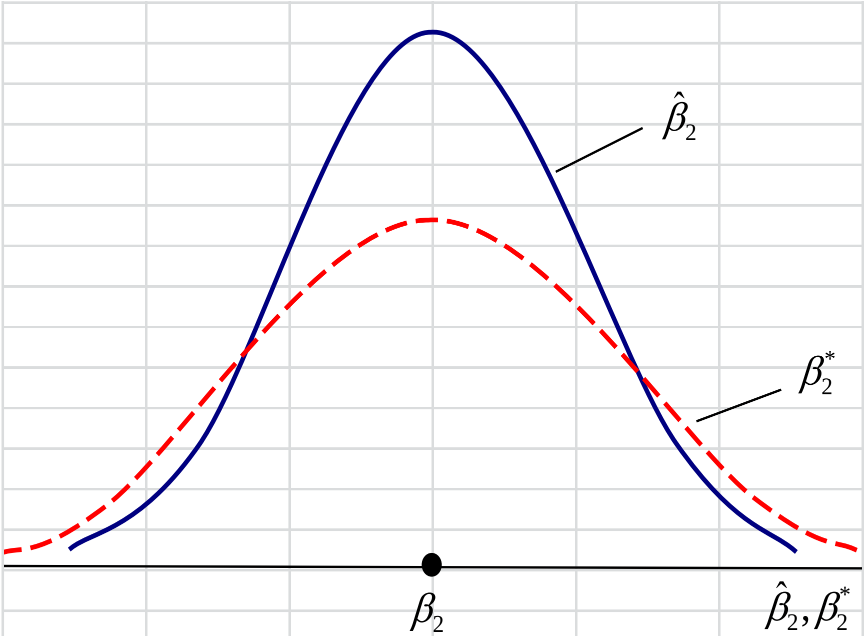


(b) Phân phối mẫu của β_2*

3.10. Tiêu chuẩn của các ước lượng OLS- Định lý Gauss- Markov

Hình 3.07. Phân phối mẫu của ước lượng $\hat{\beta}_2$ (OLS) và β_2^* (phương pháp khác)

- hình 3.07(c): cả β_2^* và $\hat{\beta}_2$ đều là các ước lượng không chệch, tuy nhiên, β_2^* phân tán rộng quanh giá trị TB hơn $\hat{\beta}_2$ .Nói cách khác, phương sai của β_2^* lớn hơn phương sai của $\hat{\beta}_2$. Bây giờ, trong hai ƯL cùng là ƯL tuyến tính, không chệch, đương nhiên ta chọn ƯL nào có **phương sai nhỏ hơn** bởi đó là ƯL có giá trị gần với giá trị của β_2 hơn →đó chính là ước lượng $\hat{\beta}_2$ vì nó thỏa mãn tiêu chuẩn BLUE.



(c) Phân phối mẫu của β2 và β2*

3.11. So sánh ước lượng của hồi qui đa biến và đơn biến

(1) $$Y = \beta_0 + \beta_1 X_1$$

(2) $$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

$$\beta_1 = \beta_1 + \beta_2 \delta_1 \quad \textcolor{red}{[3.1]}$$

- δ_1 là hệ số độ dốc từ hồi qui của X_{i2} lên X_{i1}
- β_1 & β_1 bằng nhau khi:

$$\beta_2 = 0 \qquad \text{hoặc}$$

$$\beta_2 \# 0 \quad \text{và} \quad \delta_1 = 0$$

3.12. Kỳ vọng toán trong mô hình không xác định

- **Mô hình không xác định:**

  - Đưa biến không liên quan vào mô hình (overspesifying the model)

  - Không đưa biến liên quan vào mô hình (underspecifying the model)

Nhắc lại

Định lý 1: Ước lượng không chệch của các tham số

Với các giả thiết 1-4 trên, ta có:

$$E(\beta_1) = \beta_1, và\, E(\beta_2) = \beta_2$$

Nghĩa là, $\beta_1 và\, \beta_2$ là ước lượng không chệch của $\beta_1 và\, \beta_2$

Định lý 2: Ước lượng không chệch của phương sai sai số của tổng thể:

Với các giả thiết trên, ta có:

$$E(\sigma^2) = \sigma^2$$

3.12.1. Bao gồm biến không liên quan vào mô hình

- Biến độc lập được đưa vào mô hình ngay dù nó không có ảnh hưởng đến Y ở tổng thể (Hệ số β ở tổng thể = 0)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

- X_3 không có ảnh hưởng đến Y, $\beta_3 = 0$

$$E(Y / X_1, X_2, X_3) = E(Y / X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

- Vì chúng ta ko biết $\beta_3 = 0$ ➔ đưa X_3 vào phtr SRF
  - Không ảnh hưởng đến tính không chệch của $\beta_1 \& \beta_2$
  - Ảnh hưởng đến phương sai của các ước lượng (giải thích ở phần sau)

3.12.2. Không bao gồm biến liên quan vào mô hình

- Mô hình đầy đủ PRF:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

- Khi ước lượng không đưa X_2 vào mô hình mặc dù nó có ý nghĩa thống kê.

$$Y = \beta_0 + \beta_1 X_1$$

- [3.1] ➔ $$\beta_1 = \beta_1 + \beta_2 \delta_1$$

3.12.2. Không bao gồm biến liên quan vào mô hình

$$E(\beta_1) = E(\beta_1 + \beta_2\delta_1) = E(\beta_1) + E(\beta_2)\delta_1$$
$$= \beta_1 + \beta_2\delta_1$$

- Phần chệch của β_1 là:

$$Bias(\beta_1) = E(\beta_1) - \beta_1 = \beta_2\delta_1$$

- β_1 là ước lượng không chệch của β_1 khi:

  $\beta_2 = 0$ ➔ loại vì từ đầu giả định X_2 có ý nghĩa

  hoặc $\delta_1 = 0$ ngay dù $\beta_2 \# 0$ → X_1 & X_2 ko có mối quan hệ

3.12.2. Không bao gồm biến liên quan vào mô hình

- Như vậy, nếu $X_1 \& X_2$ có mối liên hệ với nhau và X_2 có ảnh hưởng đến Y , chúng ta ko đưa X_2 vào mô hình, khi đó ta có ước lượng chệch của β_1

- **Khi phần chệch (bias) tồn tại:**

	$Corr(X_1, X_2) > 0$	$Corr(X_1, X_2) < 0$
$\beta_2 > 0$	Phần chệch dương	Phần chệch âm
$\beta_2 < 0$	Phần chệch âm	Phần chệch dương

3.13. Phương sai trong mô hình không xác định

Định lý 3: Phương sai của các ước lượng

Với các giả thiết trên, ta có:

$$\mathrm{Var}(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1-R_j^2)}$$

j =1,2,….,k; $SST_j = \sum_{i=1}^{n}(X_{\mathrm{ij}} - \overline{X}_j)^2$

R_j^2 là R^2 từ hồi qui X_j lên các biến độc lập khác

$\mathrm{var}(\beta_j)$ lớn hơn → ước lượng thiếu chính xác hơn → khoảng tin cậy lớn hơn → kiểm định giả thuyết thống kê kém chx hơn

3.13.1. Các yếu tố ảnh hưởng đến phương sai của ước lượng OLS

1. **Phương sai sai số, σ^2:**

- σ^2 càng lớn → $\text{var}(\beta_j)$ càng lớn
- σ^2 lớn hơn nghĩa là việc phân bố của các biến không quan sát được ảnh hưởng đến Y càng rộng hơn → "nhiễu" (noise) hơn trong phương trình → khó ước lượng hơn ảnh hưởng từng phần của từng biến X đến Y.
- σ^2 là giá trị ko biết, thuộc về tổng thể
- Muốn giảm σ^2 → đưa thêm nhiều biến X vào phương trình

3.13.1. Các yếu tố ảnh hưởng đến phương sai của ước lượng OLS

2. **Sự biến động ở X_j, SST_j :**

- SST_j càng lớn thì $\mathrm{var}(\hat{\beta}_j)$ càng nhỏ

→ Khi các yếu tố khác giống nhau, khi ước lượng $\hat{\beta}_j$ chúng ta muốn có càng nhiều biến động ở X_j

→ **tăng kích cỡ mẫu.**

3.13.1. Các yếu tố ảnh hưởng đến phương sai của ước lượng OLS

3. Mối quan hệ tuyến tính giữa các biến độc lập, R_j^2

- R_j^2 phản ánh % sự biến động của X_j được giải thích bởi các biến độc lập khác.
- R_j^2 càng lớn → mối quan hệ tuyến tính giữa X_j và các biến X khác càng lớn → thì $\mathrm{var}(\beta_j)$ càng lớn
- Nếu $R_j^2 = 0$ →là trường hợp tốt nhất để ước lượng β_j nhưng điều này hiếm khi xảy ra.
- Nếu $R_j^2 = 1$ → vi phạm giả thiết 8 về cộng tuyến hoàn hảo.
- R_j^2 → 1 → Sự tương quan lớn (không phải hoàn hảo) giữa 2 hay nhiều biến độc lập thì $\mathrm{var}(\beta_j) \rightarrow \infty$

3.13.1. Các yếu tố ảnh hưởng đến phương sai của ước lượng OLS

- Sự tương quan lớn (không phải hoàn hảo) giữa 2 hay nhiều biến độc lập → **đa cộng tuyến**
- Đa cộng tuyến **không** vi phạm bất kỳ giả thiết nào.

3.13.2. Đưa biến ko liên quan vào mô hình:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

- X_2 không có ảnh hưởng đến Y, $\beta_2 = 0$

➔ Đưa biến không liên quan vào mô hình không ảnh hưởng gì đến tính không chệch của ước lượng:

$$E(\beta_1) = \beta_1$$

$$E(\beta_2) = \beta_2 = 0$$

3.13.3. Không đưa biến liên quan vào mô hình:

- PRF: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$
- Khi ước lượng không đưa X_2 vào mô hình mặc dù nó có ý nghĩa thống kê: $Y = \beta_0 + \beta_1 X_1$
- Biasness: $\Longrightarrow E(\beta_1) = \beta_1 + \beta_2 \delta_1$

$$E(\beta_1) = \beta_1 \iff \begin{array}{ll} (1) & \beta_2 = 0 \\ (2) & \beta_2 \# 0; Corr(X_1, X_2) = 0 \end{array}$$

$$E(\beta_1) = \beta_1 + bias \iff \beta_2 \# 0; Corr(X_1, X_2) \neq 0$$

Một số kết luận:

• Nếu không chệch là tính chất duy nhất để đo chất lượng của ước lương → β_1 được ưa thích hơn β_1 . Nghĩa là ta cứ cho biến X_2 ngay dù nó có liên quan hoặc không liên quan gì đến Y.

•Tuy nhiên nếu xét cả tính chất phương sai nhỏ nhất, sự lựa chọn sẽ thay đổi. Cụ thể:

$$\mathrm{Var}(\beta_1) = \sigma^2 / [SST_1(1 - R_1^2)]$$

$$\mathrm{Var}(\beta_1) = \sigma^2 / SST_1$$

•Nếu $Cov(X_1, X_2) \# 0$ ➔ $Var(\beta_1) < \mathrm{Var}(\beta_1)$

❖ Nếu $Cov(X_1, X_2) \# 0$ (**vấn đề đa cộng tuyến)**

TH1. Khi $\beta_2 \# 0$, β_1 là ước lượng chệch, β_1 là ước lượng không chệch, và $Var(\beta_1) < Var(\beta_1)$

TH2. Khi $\beta_2 = 0$, β_1 và β_1 là ước lượng không chệch, và $Var(\beta_1) < Var(\beta_1)$

(TH2) Như vậy, khi X_2 không ảnh hưởng đến Y, khi đưa vào mô hình sẽ làm trầm trọng hơn vấn đề của đa cộng tuyến, dẫn đến việc ước lượng β_1 kém hiệu quả.

➔ Phương sai của ước lượng cao hơn là cái giá của việc đưa vào mô hình biến không liên quan.

(TH1) Khi $\beta_2 \# 0$, không đưa X_2 vào mô hình → ước lượng chệch

- Có 2 lý do để đưa X_2 vào mô hình:
  - phần ước lượng chệch ở β_1 không giảm khi kích cỡ mẫu tăng
  - $Var(\beta_1) \& Var(\beta_1) \to 0$ khi kích cỡ mẫu tăng; nghĩa là vấn đề đa cộng tuyến khi cho X_2 vào mô hình trở nên ít quan trọng hơn khi kích cỡ mẫu tăng.
  - Với mẫu đủ lớn, β_1 được ưa thích hơn , nghĩa là nên đưa X_2 vào mô hình khi ta biết nó có liên quan đến Y.

❖Nếu

$$Corr(X_1, X_2) = 0$$

$$và\ \beta_2 \# 0$$

➔

$$E(\beta_1) = \beta_1$$

$$Var(\beta_1) = \mathrm{Var}(\beta_1)$$

3.14. Đa cộng tuyến (multiconlinearity)

- Đa cộng tuyến xảy ra khi X_j có quan hệ tuyến tính "mạnh" với các biến X khác → $R_j^2 \to 1$
- Nếu các yếu tố khác như nhau, khi ước lượng β_j, tốt hơn nếu $Corr(X_1, X_2)$ thấp hơn.
- R_j^2 càng lớn → $Var(\beta_j)$ càng lớn → ước lượng ko hiệu quả.

<u>Giải pháp:</u>

- Tăng kích cỡ mẫu → SST_j tăng
- Với một mẫu cố định, bỏ một số biến X ra khỏi mô hình → dẫn đến ước lượng chệch

3.14. Đa cộng tuyến (multiconlinearity)

Tuy nhiên nếu:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

- X_1 không liên quan đến X_2 & X_3, nhưng X_2 & X_3 lại liên quan đến nhau
- Khi đó $E(\beta_1)$ & $Var(\beta_1)$ không bị ảnh hưởng gì.
- Nếu mối quan tâm của chúng là β_1 → không cần phải quan tâm đến mối quan hệ của X_2 & X_3

3.15. Độ chính xác của các ước lượng OLS

Phương sai (var) của các ước lượng:

$$\mathrm{Var}(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1-R_j^2)}$$

độ lệch chuẩn của các ước lượng (sd):

$$sd(\hat{\beta}_j) = \frac{\sigma}{[SST_j(1-R_j^2)]^{1/2}}$$

3.15. Độ chính xác của các ước lượng OLS

<u>Sai số chuẩn của hồi qui</u> (σ)

$$\sigma^2 = (\sum_{i=1}^{n} u_i^2)/(n-k-1)$$

(n-k-1): số bậc tự do; n: số quan sát; k: số biến độc lập

<u>Sai số của ước lượng:</u>

$$se(\hat{\beta}_j) = \frac{\sigma}{[SST_j(1-R_j^2)]^{1/2}}$$