

KINH TẾ LƯỢNG CƠ BẢN

BASIC ECONOMETRICS

BỔ SUNG KIẾN THỨC

Bùi Dương Hải
Bộ môn Toán kinh tế - Khoa Toán kinh tế
haiktqd@yahoo.com
www.mfe.edu.vn

1

Tài liệu

1. Vũ Thiều, Nguyễn Quang Dong, Nguyễn Khắc Minh, *Kinh tế lượng*, NXB KHK, 1996.
2. Nguyễn Quang Dong, *Bài giảng Kinh tế lượng*, 2009.
3. Bùi Dương Hải, *Hướng dẫn thực hành Kinh tế lượng với chương trình Eviews4*, 2009.
4. Website khoa Toán kinh tế: www.mfe.edu.vn

2

Nội dung

- Bài mở đầu
- Chương 1. Mô hình kinh tế lượng
- Chương 2. Ước lượng và phân tích mô hình kinh tế lượng
- Chương 3. Đánh giá về mô hình

3

BÀI MỞ ĐẦU

Khái niệm về Kinh tế lượng

- Econometrics = Econo + Metrics
- Đối tượng: các mối quan hệ, các quá trình kinh tế
- Công cụ: các mô hình kinh tế, mô hình toán, xác suất thống kê, tin học
- Công việc: xây dựng, đánh giá, phân tích mô hình
- Kết quả: bảng số, tùy thuộc mục đích sử dụng

4

Phương pháp luận

- Đặt giả thiết về vấn đề nghiên cứu
- Xây dựng mô hình
 - Mô hình lý thuyết
 - Mô hình toán học: gồm các biến (*variable*), các phương trình (*equation*), các tham số, hệ số (*parameter, coefficient*)
- Thu thập số liệu và ước lượng tham số
- Kiểm định về mối quan hệ kinh tế
- Phân tích, dự báo về các đối tượng và các mối quan hệ kinh tế

5

Nhắc lại về Xác suất thống kê

- Biến ngẫu nhiên: X
- Phân phối xác suất
- Tham số đặc trưng:
 - Kỳ vọng (*expected value*): $E(X)$
 - Phương sai (*variance*): $Var(X)$
 - Độ lệch chuẩn (*standard deviation*): σ_X
- Tổng thể (*population*) và mẫu (*sample*)
- Mẫu ngẫu nhiên và mẫu cụ thể / các quan sát (*observations*)

6

Nhắc lại về Xác suất thống kê

- Thống kê (*statistics*)
 - Trung bình mẫu
 - Phương sai mẫu
 - Độ lệch chuẩn mẫu
- Bài toán ước lượng
 - Ước lượng điểm
 - Ước lượng khoảng hay khoảng tin cậy (*confidence interval*)
- Kiểm định giả thuyết (*hypothesis testing*)

7

Chương 1

MÔ HÌNH KINH TẾ LƯỢNG Econometric Model

- 1.1. Phân tích hồi quy
 - 1.2. Mô hình hồi quy tổng thể
 - 1.3. Mô hình hồi quy mẫu
 - 1.4. Mô hình hồi quy tổng quát
 - 1.5. Một số mô hình hồi quy trong kinh tế
- Tài liệu: “*Kinh tế lượng*” chương 1 + 2 + 3

8

1.1. Phân tích hồi quy

- Phân tích mối quan hệ phụ thuộc giữa biến một biến (biến phụ thuộc) phụ thuộc vào các biến khác (biến độc lập)
- **Biến phụ thuộc** (*dependent, explained, endogenous variable*), kí hiệu là Y
- **Biến độc lập** (*independent, explanatory, exogenous variable; regressor*), kí hiệu là X, X_1, X_2, \dots
- Trường hợp đơn giản 1 biến độc lập: $Y \leftarrow X$

9

1.1. Phân tích hồi quy

- Khi $X = X_i$: là các giá trị xác định \rightarrow biến phụ thuộc Y là biến ngẫu nhiên kí hiệu: (Y / X_i)
- Đánh giá, phân tích biến ngẫu nhiên (Y / X_i) qua các đại lượng:
Trung bình: $E(Y / X_i)$ Phương sai: $Var(Y / X_i)$
- Quan hệ hàm số : $x \rightarrow ! y$
- Hệ số tương quan $\rho_{XY} \in [-1 ; 1]$

10

1.2. Mô hình hồi quy tổng thể

- Tổng thể (*population*): tất cả các phần tử chứa dấu hiệu nghiên cứu
- Phân tích mô hình xây dựng trên toàn bộ tổng thể
- Để thuận tiện: mô hình một biến độc lập
Mô hình: $X \rightarrow Y$: X giải thích cho Y , Y phụ thuộc vào X như thế nào

11

Hàm hồi quy tổng thể (PRF)

- $X = X_i \rightarrow$ biến phụ thuộc là biến ngẫu nhiên (Y/X_i)
- \Rightarrow tồn tại quy luật phân phối xác suất
- \Rightarrow tồn tại duy nhất $E(Y/X_i)$: trung bình có điều kiện
- $X_i \rightarrow ! E(Y/X_i)$: tồn tại quan hệ hàm số
- $E(Y/X_i) = f(X_i)$ hoặc $E(Y/X) = f(X)$
Gọi là **hàm hồi quy tổng thể**
PRF: *Population Regression Function*

12

Hàm hồi quy tổng thể (PRF)

- Dạng của PRF tùy thuộc mô hình kinh tế, gồm các **hệ số** (*coefficient*) chưa biết

- Nếu hàm hồi quy tổng thể có dạng đường thẳng:

$$E(Y/X) = \beta_1 + \beta_2 X$$

- β_1 : **hệ số chặn** (*intercept*): $\beta_1 = E(Y/X = 0)$

- β_2 : **hệ số góc** (*slope coefficient*)

$$\beta_2 = \frac{dE(Y/X)}{dX}$$

13

Sai số ngẫu nhiên

- Với giá trị cá biệt $Y_i \in (Y/X_i)$

- Thông thường $Y_i \neq E(Y/X_i)$

- Đặt $u_i = Y_i - E(Y/X_i)$

u_i là sai số ngẫu nhiên (*disturbance, random error*),

$\rightarrow u_i$ có giá trị (+) và (-) hoặc 0, $E(u_i) = 0$.

- Suy ra $Y_i = E(Y/X_i) + u_i$

$$Y_i = f(X_i) + u_i$$

14

1.3. Mô hình hồi quy mẫu

- Không biết toàn bộ Tổng thể
- PRF: biết dạng, không biết giá trị tham số β_j
- Mẫu (*sample*): một bộ phận mang thông tin của tổng thể, mẫu là ngẫu nhiên hoặc mẫu cụ thể.
- $W = \{(X_i, Y_i) ; i = 1 \div n\}$ mẫu ngẫu nhiên kích thước n
- Mẫu cụ thể w có n **quan sát** (*observations*)

15

Hàm hồi quy mẫu (SRF)

- Tìm hàm số mô tả xu thế biến động của mẫu về mặt trung bình $\hat{Y}_i = \hat{f}(X_i)$

Gọi là **SRF** (*Sample Regression Function*)

- SRF có dạng giống PRF
- Nếu PRF có dạng: $E(Y / X_i) = \beta_1 + \beta_2 X_i$
Thì SRF có dạng: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$

16

Ước lượng

- $\hat{\beta}_1, \hat{\beta}_2$ là **ước lượng** ngẫu nhiên của các hệ số (*estimators*)
- Với mẫu cụ thể với các quan sát (*observation*) kết quả sẽ là con số cụ thể (*estimates*)
- \hat{Y}_i là **giá trị ước lượng** trung bình biến phụ thuộc, hay **giá trị tương hợp** (*fitted values*)

17

Phần dư

- Thông thường: $Y_i \neq \hat{Y}_i$
- Đặt: $e_i = \hat{u}_i = Y_i - \hat{Y}_i$
- Suy ra: $Y_i = \hat{Y}_i + e_i = \hat{f}(X_i) + e_i$
- Các giá trị e_i gọi là **phần dư** (*residuals*)
- Phần dư nhận giá trị (-), (+), là sai số ngẫu nhiên trong mẫu
- Phần dư là ước lượng của u_i trong mẫu
- Trong tính toán, sử dụng e_i thay cho u_i

18

Tóm tắt

- Tổng thể $E(Y / X_i) = \beta_1 + \beta_2 X_i$

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
- Mẫu $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$$

19

1.4. Mô hình tổng quát

- Y phụ thuộc k biến độc lập X_1, X_2, \dots, X_k , với $X_1 \equiv 1$

$$E(Y / X_{2i}, \dots, X_{ki}) = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} = \sum_{j=1}^k \beta_j X_{ji}$$

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k X_{ki}$$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k X_{ki} + e_i$$

20

Dạng ma trận

- Đặt: $\mathbf{X}_i = (1 \quad X_{2i} \quad X_{3i} \quad \dots \quad X_{ki})$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{pmatrix} \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \dots \\ \hat{\beta}_k \end{pmatrix} \quad \begin{cases} E(Y_i) = \mathbf{X}_i \boldsymbol{\beta} \\ Y_i = \mathbf{X}_i \boldsymbol{\beta} + u_i \\ \hat{Y}_i = \mathbf{X}_i \hat{\boldsymbol{\beta}} \\ Y_i = \mathbf{X}_i \hat{\boldsymbol{\beta}} + e_i \end{cases}$$

21

Dạng ma trận

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \dots \\ \mathbf{X}_n \end{pmatrix} \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{pmatrix} \quad \mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix}$$

$$\begin{cases} E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta} \\ \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \\ \mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e} \end{cases} \quad \text{Với } E(\mathbf{u}) = [\mathbf{0}]$$

22

1.5. Một số mô hình trong kinh tế

- Hàm bậc nhất

$$C_i = \beta_1 + \beta_2 Y_i + u_i$$

$$Q_i^D = \beta_1 + \beta_2 P_i + u_i^D$$

$$Q_i^S = \beta_3 + \beta_4 P_i + u_i^S$$

- Hàm bậc cao

$$TC_i = \beta_1 + \beta_2 Q_i + \beta_3 Q_i^2 + \beta_4 Q_i^3 + u_i$$

$$MC_i = \beta_2 + 2\beta_3 Q_i + 3\beta_4 Q_i^2 + u_i'$$

$$AC_i = \beta_1 (1/Q_i) + \beta_2 + \beta_3 Q_i + \beta_4 Q_i^2 + u_i''$$

23

1.5. Một số mô hình trong kinh tế

- Mô hình hàm mũ

Ví dụ: Hàm sản xuất Cobb-Douglas

$$Q_i = \beta_0 K_i^{\beta_2} L_i^{\beta_3} U_i$$

$$\ln(Q_i) = \ln(\beta_0) + \beta_2 \ln(K_i) + \beta_3 \ln(L_i) + \ln(U_i)$$

$$= \beta_1 + \beta_2 \ln(K_i) + \beta_3 \ln(L_i) + u_i$$

Tổng quát

$$Y_i = \prod_{j=1}^k X_{ji}^{\beta_j} \cdot e^{u_i} \quad \ln Y_i = \sum_{j=1}^k \beta_j \ln X_{ji} + u_i$$

24

1.5. Một số mô hình trong kinh tế

- Mô hình có tính xu thế
- Số liệu theo thời gian
- T là biến xu thế thời gian: $T = 0, 1, 2, \dots$
$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 t + u_t$$
- Biến trễ
$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + u_t$$
- Mô hình tự hồi qui
$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + u_t$$

25

Chương 2. ƯỚC LƯỢNG & PHÂN TÍCH KẾT QUẢ

- 2.1. Ước lượng mô hình hai biến
- 2.2. Ước lượng mô hình tổng quát
- 2.3. Giả thiết LS – tham số của ước lượng
- 2.4. Kiểm định giả thuyết về các hệ số
- 2.5. Ước lượng khoảng các hệ số
- 2.6. Sự phù hợp của hàm hồi qui
- 2.7. Dự báo

Tài liệu: “*Kinh tế lượng*”, chương 2 + 3

26

2.1. Ước lượng mô hình hai biến

- Mô hình $Y_i = \beta_1 + \beta_2 X_i + u_i$ với β_1, β_2 chưa biết
- Với mẫu $W = \{(X_i, Y_i), i = 1 \div n\}$
$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \quad \text{hay} \quad Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$$
- Phương pháp bình phương nhỏ nhất (**Least Square**):
Tìm

$$\hat{\beta}_1, \hat{\beta}_2 \text{ sao cho: } \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n e_i^2 = RSS \rightarrow \min$$

27

2.1. Ước lượng mô hình hai biến

- Hệ phương trình

$$\begin{cases} \hat{\beta}_1 n + \hat{\beta}_2 \sum X_i = \sum Y_i \\ \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2 = \sum X_i Y_i \end{cases}$$

$$\begin{cases} \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \\ \hat{\beta}_2 = \frac{\overline{XY} - \bar{X}\bar{Y}}{\overline{X^2} - (\bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2} \end{cases} \quad \begin{matrix} x_i = X_i - \bar{X} \\ y_i = Y_i - \bar{Y} \end{matrix}$$

28

2.2. Ước lượng Mô hình tổng quát

- Mô hình ba biến

$$\begin{aligned} E(Y / X_{2i}, X_{3i}) &= \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} \\ Y_i &= \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \end{aligned}$$

- Nếu các biến độc lập có quan hệ cộng tuyến:

$$X_{3i} = \alpha_1 + \alpha_2 X_{2i} \text{ thì}$$

$$E(Y / X_{2i}, X_{3i}) = (\beta_1 + \beta_3 \alpha_1) + (\beta_2 + \beta_3 \alpha_2) X_{2i}$$

Sẽ không ước lượng được các hệ số

29

2.2. Ước lượng Mô hình tổng quát

- Mô hình k biến

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k X_{ki}$$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k X_{ki} + e_i$$

- Phương pháp LS: tìm $\hat{\beta}_j (j=1 \div k, k \geq 2)$

$$\text{sao cho: } RSS = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n e_i^2 = \mathbf{e}'\mathbf{e} \rightarrow \min$$

30

2.2. Ước lượng Mô hình tổng quát

- Dạng hệ phương trình thông thường

$$\left\{ \begin{array}{llll} \hat{\beta}_1 n & + \hat{\beta}_2 \sum X_{2i} & + \dots + \hat{\beta}_k \sum X_{ki} & = \sum Y_i \\ \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 & & + \dots + \hat{\beta}_k \sum X_{2i} \sum X_{ki} & = \sum (X_{2i} Y_i) \\ \dots & \dots & \dots & \dots \\ \hat{\beta}_1 \sum X_{ki} + \hat{\beta}_2 \sum X_{ki} \sum X_{2i} + \dots + \hat{\beta}_k \sum X_{ki}^2 & & & = \sum (X_{ki} Y_i) \end{array} \right.$$

- Dạng ma trận $(X'X)\hat{\beta} = X'Y$

2.3. Các giả thiết LS

- **Gt 1** : Hàm hồi quy tuyến tính theo hệ số
- **Gt 2** : Các biến độc lập không ngẫu nhiên
- **Gt 3** : Trung bình sai số bằng 0: $E(u_i) = 0 \forall i$
- **Gt 4** : Phương sai sai số không đổi

$$Var(u_i) = \sigma^2 \quad \forall i$$

- **Gt 5** : Các sai số không tương quan

$$Cov(u_i; u_{i'}) = 0 \quad \forall i \neq i'$$

Suy ra: $Cov(\mathbf{u}) = \sigma^2 \mathbf{I}$ với \mathbf{I} là ma trận đơn vị

- **Gt 6** : Sai số và các biến độc lập không tương quan

2.3. Các giả thiết LS

- **Gt 7 :** Số quan sát nhiều hơn số hệ số cần ước lượng
- **Gt 8 :** Giá trị của biến độc lập có sự khác biệt đủ lớn
- **Gt 9 :** Hàm hồi quy được chỉ định đúng
- **Gt 10:** Các biến độc lập không có quan hệ cộng tuyến
- **Gt 11:** Sai số ngẫu nhiên phân phối Chuẩn

Định lý: Nếu các giả thiết được thỏa mãn thì

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

là ước lượng tuyến tính không chệch tốt nhất của β

(BLUE: *Best Linear Unbias Estimators*).

2.3. Các giả thiết LS

Khi đó

- Kỳ vọng $E(\hat{\beta}) = \beta$
- Phương sai, hiệp phương sai

$$\text{Cov}(\hat{\beta}) = \begin{pmatrix} \text{Var}(\hat{\beta}_1) & \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) & \dots & \text{Cov}(\hat{\beta}_1, \hat{\beta}_k) \\ \text{Cov}(\hat{\beta}_2, \hat{\beta}_1) & \text{Var}(\hat{\beta}_2) & \dots & \text{Cov}(\hat{\beta}_2, \hat{\beta}_k) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(\hat{\beta}_k, \hat{\beta}_1) & \text{Cov}(\hat{\beta}_k, \hat{\beta}_2) & \dots & \text{Var}(\hat{\beta}_k) \end{pmatrix} = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

- Ước lượng điểm phương sai sai số ngẫu nhiên

$$\hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k} = \frac{RSS}{n-k}$$

34

2.3. Các giả thiết LS

- Sai số chuẩn của hồi quy (*S.E. of Regression*)

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{RSS/(n-k)}$$

- Sai số chuẩn của các ước lượng hệ số

$$\text{Se}(\hat{\beta}_j) = \sqrt{\text{Var}(\hat{\beta}_j)}$$

- Trường hợp mô hình một biến độc lập, hai hệ số

$$\text{Se}(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2 \sum X_i^2}{n \sum x_i^2}} \quad \text{Se}(\hat{\beta}_2) = \sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2}}$$

35

2.4. Kiểm định các hệ số

- Sai số phân phối chuẩn: $u_i \sim N(0, \sigma^2)$
- Mức ý nghĩa α cho trước
- Các cặp giả thuyết: H_0 và H_1 : so sánh β_j (chưa biết) với số thực β_j^* cho trước
- Miền bác bỏ H_0 , tiêu chuẩn kiểm định T
- Tính T_{qs} với mẫu cụ thể, nếu T_{qs} thuộc miền bác bỏ thì bác bỏ H_0 , ngược lại thì chưa có cơ sở bác bỏ H_0

36

2.4. Kiểm định các hệ số

Tiêu chuẩn	Cặp giả thuyết	Miền bác bỏ H_0
$T = \frac{\hat{\beta}_j - \beta_j^*}{Se(\hat{\beta}_j)}$	$\begin{cases} H_0 : \beta_j = \beta_j^* \\ H_1 : \beta_j \neq \beta_j^* \end{cases}$	$ T > t_{\alpha/2}^{(n-k)}$
	$\begin{cases} H_0 : \beta_j = \beta_j^* \\ H_1 : \beta_j > \beta_j^* \end{cases}$	$T > t_{\alpha}^{(n-k)}$
	$\begin{cases} H_0 : \beta_j = \beta_j^* \\ H_1 : \beta_j < \beta_j^* \end{cases}$	$T < -t_{\alpha}^{(n-k)}$

37

2.4. Kiểm định các hệ số

- Cặp giả thuyết cơ bản

$$\begin{cases} H_0 : \beta_j = 0 & \text{Hệ số không có ý nghĩa thống kê} \\ H_1 : \beta_j \neq 0 & \text{Hệ số có ý nghĩa thống kê} \end{cases}$$
- Giá trị “Prob.” hay “P-value”: mức xác suất thấp nhất để bác bỏ H_0 ứng với một mẫu
- Nếu P-value < mức ý nghĩa α thì bác bỏ H_0 .
- Nếu P-value > α thì chưa có cơ sở bác bỏ H_0 .

38

2.5. Ước lượng khoảng các hệ số

- Độ tin cậy $(1 - \alpha)$ cho trước
- Ước lượng bằng khoảng tin cậy: đối xứng, tối đa, tối thiểu cho từng hệ số hồi quy

$$\begin{aligned} \hat{\beta}_j - Se(\hat{\beta}_j)t_{\alpha/2}^{(n-k)} &< \beta_j < \hat{\beta}_j + Se(\hat{\beta}_j)t_{\alpha/2}^{(n-k)} \\ \beta_j &< \hat{\beta}_j + Se(\hat{\beta}_j)t_{\alpha}^{(n-k)} \\ \hat{\beta}_j - Se(\hat{\beta}_j)t_{\alpha}^{(n-k)} &< \beta_j \end{aligned}$$

39

Mở rộng cho hai hệ số

- Kiểm định hai hệ số

$$\begin{cases} H_0 : \beta_i \pm \beta_j = \beta^* \\ H_1 : \beta_i \pm \beta_j \neq \beta^* \end{cases} \quad T = \frac{(\hat{\beta}_i \pm \hat{\beta}_j) - \beta^*}{Se(\hat{\beta}_i \pm \hat{\beta}_j)}$$

- Ước lượng hai hệ số

$$(\hat{\beta}_i \pm \hat{\beta}_j) - Se(\hat{\beta}_i \pm \hat{\beta}_j)t_{\alpha/2}^{(n-k)} < \beta_i \pm \beta_j < (\hat{\beta}_i \pm \hat{\beta}_j) + Se(\hat{\beta}_i \pm \hat{\beta}_j)t_{\alpha/2}^{(n-k)}$$

$$\begin{aligned} \text{Với } Se(\hat{\beta}_i \pm \hat{\beta}_j) &= \sqrt{Var(\hat{\beta}_i \pm \hat{\beta}_j)} \\ &= \sqrt{Var(\hat{\beta}_i) + Var(\hat{\beta}_j) \pm 2Cov(\hat{\beta}_i, \hat{\beta}_j)} \end{aligned}$$

40

2.6. Sự phù hợp của hàm hồi qui

Đo sự biến động của biến phụ thuộc

$$\left. \begin{aligned} y_i &= Y_i - \bar{Y} \\ \hat{y}_i &= \hat{Y}_i - \bar{Y} \\ e_i &= Y_i - \hat{Y} \end{aligned} \right\} \Rightarrow y_i = \hat{y}_i + e_i$$

$$\begin{aligned} \text{Chứng minh được } \sum_{i=1}^n y_i^2 &= \sum_{i=1}^n \hat{y}_i^2 + \sum_{i=1}^n e_i^2 \\ TSS &= ESS + RSS \end{aligned}$$

41

Đo độ biến động của biến phụ thuộc

- TSS (Total Sum of Squares)**: đo tổng biến động tổng hợp của biến phụ thuộc
- ESS (Explained Sum of Squares)**: tổng biến động của biến phụ thuộc được giải thích bởi mô hình – các biến độc lập.
- RSS (Residual Sum of Squares)**: tổng biến động của biến phụ thuộc được giải thích bởi các yếu tố nằm ngoài mô hình – yếu tố ngẫu nhiên.

42

Hệ số xác định R^2

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \in [0; 1]$$

- Ý nghĩa: R^2 cho biết tỷ lệ (%) sự biến động của biến phụ thuộc được giải thích bởi sự biến động của tất cả các biến độc lập có trong mô hình
- Hệ số xác định điều chỉnh (*Adjusted R-squared*)

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

$$\bar{R}^2 < R^2$$

43

Kiểm định sự phù hợp

$$\begin{cases} H_0 : R^2 = 0 \\ H_1 : R^2 > 0 \end{cases} \quad \text{hay} \quad \begin{cases} H_0 : \beta_2 = \dots = \beta_k = 0 \\ H_1 : \exists \beta_j \neq 0 : (j \neq 1) \end{cases}$$

$$F_{qs} = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{R^2}{1-R^2} \times \frac{n-k}{k-1}$$

- Miền bác bỏ H_0 $F > F_{\alpha}^{(k-1; n-k)}$
- Nếu bác bỏ H_0 : hàm hồi quy là **phù hợp**: ít nhất một biến độc lập có giải thích cho biến phụ thuộc

44

Kiểm định thu hẹp hồi quy

- Nghi ngờ m biến độc lập X_{k-m+1}, \dots, X_k không giải thích cho Y

$$\begin{cases} H_0 : \beta_{k-m+1} = \beta_{k-m+2} = \dots = \beta_k = 0 \\ H_1 : \exists \beta_j \neq 0 : (j = k-m+1 \div k) \end{cases}$$

$$E(Y / X_2, \dots, X_{k-m}, \dots, X_k) = \beta_1 + \dots + \dots + \beta_k X_k \quad (L)$$

$$E(Y / X_2, \dots, X_{k-m}) = \beta_1 + \dots + \beta_{k-m} X_{k-m} \quad (N)$$

$$F = \frac{RSS_N - RSS_L}{RSS_L} \times \frac{n-k}{m} = \frac{R_L^2 - R_N^2}{1 - R_L^2} \times \frac{n-k}{m}$$

45

Kiểm định thu hẹp hồi quy

- Miền bác bỏ $H_0: F > F_{\alpha}^{(m, n-k)}$
- $m = 1: F_{qs} = (T_{qs}^*)^2$ của hệ số tương ứng
- $m = k - 1: F_{qs}$ kiểm định thu hẹp chính là F_{qs} kiểm định sự phù hợp.
- Kiểm định thu hẹp hồi quy còn dùng cho mở rộng hồi quy, kiểm định quan hệ tuyến tính giữa các hệ số (kiểm định điều kiện ràng buộc tuyến tính giữa các hệ số)

46

2.6. Dự báo

$$\mathbf{X}^0 = \begin{pmatrix} 1 & X_2^0 & X_3^0 & \dots & X_k^0 \end{pmatrix}$$

- Dự báo, ước lượng khoảng trung bình biến phụ thuộc với độ tin cậy cho trước
- Ước lượng điểm: $\hat{Y}_0 = \mathbf{X}^0 \hat{\boldsymbol{\beta}} = \hat{\beta}_1 + \hat{\beta}_2 X_2^0 + \dots + \hat{\beta}_k X_k^0$
- Ước lượng khoảng:

$$\hat{Y}_0 - Se(\hat{Y}_0) t_{\alpha/2}^{(n-k)} < E(Y / \mathbf{X}^0) < \hat{Y}_0 + Se(\hat{Y}_0) t_{\alpha/2}^{(n-k)}$$

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

47

Chương 3. ĐÁNH GIÁ MÔ HÌNH (Diagnostic Tests)

- 3.1. Đa cộng tuyến (*Multicollinearity*)
- 3.2. Phương sai sai số thay đổi (*Heteroskedasticity*)
- 3.3. Tự tương quan (*Autocorrelation*)
- 3.4. Định dạng phương trình (*Equation specification*)

Tài liệu: “Kinh tế lượng”, chương 5 + 6 + 7 + 8

48

Cơ sở đánh giá

- **Định lý Gauss-Markov**: Nếu mô hình hồi qui thỏa mãn các giả thiết của phương pháp LS thì các ước lượng LS là tuyến tính, không chệch, tốt nhất
- **BLUE** (*Best Linear Unbias Estimators*)
- Các giả thiết không được thỏa mãn: các ước lượng không tốt, kết quả không đáng tin cậy, cần phải khắc phục

49

3.1. Đa cộng tuyến

- Mô hình nhiều biến độc lập
 $E(Y/X_{2i}, \dots, X_{ki}) = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} \quad (k \geq 3)$
- Các biến độc lập có quan hệ cộng tuyến: quan hệ bậc nhất với nhau: mô hình có hiện tượng Đa cộng tuyến (*multicollinearity*)

50

Phân loại Đa cộng tuyến

- Đa cộng tuyến hoàn hảo:

$$\lambda_1 + \lambda_2 X_{2i} + \dots + \lambda_k X_{ki} = 0 \quad \forall i,$$

hay $X_{ki} = \alpha_1 + \alpha_2 X_{2i} + \dots + \alpha_{k-1} X_{k-1}$

- Đa cộng tuyến không hoàn hảo:

$$\lambda_1 + \lambda_2 X_{2i} + \dots + \lambda_k X_{ki} + w_i = 0 \quad \forall i,$$

hay $X_{ki} = \alpha_1 + \alpha_2 X_{2i} + \dots + \alpha_{k-1} X_{k-1} + v_i$

(với $\alpha_j = \lambda_j / \lambda_k$; w_i là sai số ngẫu nhiên, $v_i = w_i / \lambda_k$)

51

Nguyên nhân – Hậu quả

- Đa cộng tuyến hoàn hảo do đặt sai mô hình: ít khi xảy ra → không giải được
- Đa cộng tuyến không hoàn hảo thường xảy ra: do bản chất kinh tế xã hội, thu thập và xử lý số liệu
- Đa cộng tuyến không hoàn hảo → vẫn giải được nhưng kết quả không tốt nhất
- Đa cộng tuyến gần hoàn hảo → các kiểm định T, F mâu thuẫn nhau, ước lượng sai về dấu

52

Kiểm định phát hiện

- Nghi ngờ X_k phụ thuộc vào các biến độc lập khác
- Dùng hồi quy phụ (*auxiliary regression*)

$$X_{ki} = \sum_{j \neq k} \alpha_j X_{ji} + v_i \quad (*)$$

$$\begin{cases} H_0 : R_*^2 = 0 \\ H_1 : R_*^2 > 0 \end{cases} \Leftrightarrow \begin{cases} H_0 : \forall \alpha_j \equiv 0 \\ H_1 : \exists \alpha_j \neq 0 \end{cases} \quad (j \neq 1)$$

H_0 : mô hình gốc không có đa cộng tuyến

H_1 : mô hình gốc có đa cộng tuyến

$$F_{qs} = \frac{R_*^2}{1 - R_*^2} \times \frac{n - k_*}{k_* - 1}$$

53

Khắc phục

- Cách khắc phục đơn giản nhất: bỏ bớt biến độc lập
- Có thể đổi dạng mô hình
- Sử dụng thông tin tiên nghiệm

54

3.2. Phương sai sai số thay đổi

- Mô hình: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ (1)
- Giả thiết: $Var(u_i) \equiv \sigma^2 \forall i$ thì mô hình (1) có
Phương sai sai số không đổi (đồng đều)
Homoscedasticity
- Nếu: $Var(u_i) \neq Var(u_j)$ với $i \neq j$ thì mô hình (1) có
Phương sai sai số thay đổi (không đồng đều)
Heteroscedasticity

55

Nguyên nhân – Hậu quả

- Bản chất kinh tế xã hội: sự dao động của biến phụ thuộc trong những điều kiện khác nhau không giống nhau
 - Quá trình thu thập số liệu không chính xác
 - Xử lý, làm tròn số liệu
- Các ước lượng là không chệch nhưng không tốt nhất

56

Kiểm định phát hiện

- Vì $\sigma_i^2 = Var(u_i) = E(u_i^2)$ là chưa biết, thay bởi $E(e_i^2)$
- Kiểm định theo biến độc lập X_2 :
$$E(e_i^2) = \alpha_1 + \alpha_2 X_{2i} \quad E(e_i^2) = \alpha_1 + \alpha_2 X_{2i}^2$$
$$E(e_i^2) = \alpha_1 + \alpha_2 \sqrt{X_{2i}} \quad E(e_i^2) = \alpha_1 + \alpha_2 (1/X_{2i})$$
- Kiểm định Park: $E(\ln e_i^2) = \alpha_1 + \alpha_2 \ln X_{2i}$
- Kiểm định dựa trên biến phụ thuộc:
$$E(e_i^2) = \alpha_1 + \alpha_2 \hat{Y}_i^2$$
- Nếu $\alpha_2 \neq 0$ hay $R^2_{phụ} \neq 0$ thì mô hình gốc có phương sai sai số thay đổi

57

Kiểm định White

- Trường hợp có tích chéo (*cross items*)

$$E(e_i^2) = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} \quad (*)$$

H_0 : MH đầu có PSSS không đổi

H_1 : MH đầu có PSSS thay đổi

- Kiểm định F
- Kiểm định χ^2 : $\chi^2_{qs} = n.R_*^2$
 Nếu $\chi^2_{qs} > \chi^2_{\alpha}(k_* - 1)$: bác bỏ H_0

58

Khắc phục

- Khi biết $\sigma_i^2 = Var(u_i)$, chia phương trình hồi quy gốc ban đầu cho $Se(u_i) = \sigma_i$

$$\frac{Y_i}{\sigma_i} = \beta_1 \frac{1}{\sigma_i} + \beta_2 \frac{X_{2i}}{\sigma_i} + \beta_3 \frac{X_{3i}}{\sigma_i} + \frac{u_i}{\sigma_i}$$

Khi đó $Var(u_i/\sigma_i) = 1$ không đổi

- Nếu chưa biết σ_i , hồi quy phụ e_i^2 theo đại lượng nào thì chia cho căn bậc hai của đại lượng đó

59

3.3. Tự tương quan

- Xét mô hình gốc: $Y_i = \beta_1 + \beta_2 X_i + u_i$ (1)
- Giả thiết LS: $Cov(u_p, u_{i-p}) = 0 \quad \forall p \neq 0$
- Nếu $Cov(u_p, u_{i-p}) \neq 0$
 → Tự tương quan bậc p (*Autocorrelation order p^{th}*)
- Tự tương quan bậc 1: $u_i = \rho u_{i-1} + \varepsilon_i$
 ($-1 \leq \rho \leq 1$, ε_i thỏa mãn các giả thiết LS)
 - $1 < \rho < 0$ tự tương quan âm
 $\rho = 0$ không có tự tương quan
 $0 < \rho < 1$ tự tương quan dương

60

Nguyên nhân – Hậu quả

- Hiện tượng quán tính của các chuỗi thời gian
- Hiện tượng mạng nhện trong kinh tế
- Quá trình xử lý, nội ngoại suy số liệu
- Mô hình thiếu biến hoặc dạng hàm sai

→ Các ước lượng không còn tốt nhất

61

Kiểm định Durbin - Watson

- Dùng e_i thay thế cho u_i trong kiểm định
- Kiểm định Durbin-Watson (DW)

$$DW = d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \cong 2(1 - \hat{\rho})$$

$$\hat{\rho} = \frac{\sum_{i=2}^n e_i e_{i-1}}{\sum_{i=1}^n e_i^2} \quad \text{là ước lượng cho } \rho$$

62

Kiểm định Durbin Watson

- Với $n, k' = k - 1, \alpha = 5\%$ cho trước, tra bảng
- d_L và d_U là các giá trị tới hạn

Tự tương quan dương $\rho > 0$	Không có kết luận	Không có tự tương quan $\rho = 0$	Không có kết luận	Tự tương quan âm $\rho < 0$
0	d_L	d_U	$4 - d_U$	$4 - d_L$
				4

63

Kiểm định Breusch-Godfrey

- Hồi qui phụ:

$$e_i = [\alpha_1 + \alpha_2 X_i] + \rho_1 e_{i-1} + \dots + \rho_p e_{i-p} + v_i \quad (*)$$

$$e_i = [\alpha_1 + \alpha_2 X_i] + v_i \quad (**)$$

Kiểm định χ^2 : $\chi^2_{qs} = (n-p)R^2_*$

Nếu $\chi^2_{qs} > \chi^2_{\alpha}(p)$: bác bỏ H_0

- Kiểm định F:

$$F_{qs} = \frac{R^2_* - R^2_{**}}{1 - R^2_*} \times \frac{n - k_*}{p}$$

64

Khắc phục

- Khắc phục tự tương quan bậc 1:
- Sử dụng phương trình sai phân có dạng

$$Y_i - \rho Y_{i-1} = \beta_1 (1 - \rho) + \beta_2 (X_i - \rho X_{i-1}) + \varepsilon_i \quad (2)$$

Trong đó dùng ước lượng của ρ từ thống kê DW hoặc từ các hồi quy phụ.

- Có thể thêm biến, đổi dạng mô hình

65

3.4. Định dạng mô hình

- Mô hình gốc: $Y_i = \beta_1 + \beta_2 X_i + u_i \quad (1)$
- Kiểm định Ramsey RESET

$$Y_i = [\beta_1 + \beta_2 X_i] + \alpha_1 \hat{Y}_i^2 + \dots + \alpha_m \hat{Y}_i^{m+1} + u_i$$

$$\begin{cases} H_0 : \alpha_1 = \dots = \alpha_m = 0 \\ H_1 : \exists \alpha_j \neq 0, j = 1 \div m \end{cases} \quad F_{qs} = \frac{R^2_{(2)} - R^2_{(1)}}{1 - R^2_{(2)}} \times \frac{n - k_{(2)}}{m}$$

H_0 : Mô hình gốc không thiếu biến, dạng hàm đúng

H_1 : Mô hình gốc thiếu biến, dạng hàm sai

66

Tổng kết

- Nếu mô hình không có khuyết tật, các ước lượng là tốt nhất, ước lượng khoảng, kiểm định là đáng tin cậy, kết quả là tốt cho phân tích
- Phân tích sự tác động của các biến độc lập đến sự biến động của biến phụ thuộc thông qua các hệ số hồi qui và hệ số xác định R^2