

# KINH TẾ LƯỢNG

## (ECONOMETRICS)

MAI CẨM TÚ

BM Toán kinh tế - ĐH Kinh tế Quốc dân

# TÀI LIỆU

- 1. GS. TS. Nguyễn Quang Dong, TS. Nguyễn Thị Minh, *Giáo trình Kinh tế lượng*, 2012, NXB ĐH Kinh tế Quốc dân
- 2. Các tài liệu khác.

# NỘI DUNG

- Bài mở đầu
- Chương 1. Mô hình hồi quy tuyến tính hai biến
- Chương 2. Mô hình hồi quy bội
- Chương 3. Suy diễn thống kê và dự báo từ mô hình hồi quy
- Chương 4. Phân tích hồi quy với biến giả
- Chương 5. Kiểm định và lựa chọn mô hình
- Chương 6. Mô hình hồi quy với số liệu chuỗi thời gian
- Chương 7. Vấn đề tự tương quan trong mô hình hồi quy chuỗi thời gian
- Chương 8. Một số mô hình động

# MỞ ĐẦU

## 1. KINH TẾ LƯỢNG LÀ GÌ?

Có nhiều định nghĩa.

Econo: kinh tế;      Metric: đo lường

Kinh tế lượng (econometrics) là sự kết hợp giữa kinh tế học, toán học và thống kê toán nhằm lượng hóa, kiểm định và dự báo các quan hệ kinh tế. Nó là một khoa học độc lập.

❖ Mục đích của KTL:

- Thiết lập các mô hình toán học để mô tả các mối quan hệ kinh tế
- Ước lượng các tham số
- Kiểm định tính vững chắc của các giả thuyết
- Sử dụng các kết quả trên để đưa ra các dự báo, dự đoán và mô phỏng các hiện tượng kinh tế
- Đề xuất chính sách dựa trên các phân tích và dự báo

# MỞ ĐẦU

## 2. PHƯƠNG PHÁP LUẬN (8 bước)

- Nêu các giả thuyết hay giả thiết về mối quan hệ giữa các biến kinh tế
- Định dạng mô hình toán học
- Định dạng mô hình KTL
- Thu thập số liệu
- Ước lượng các tham số của mô hình
- Phân tích kết quả
- Dự báo
- Sử dụng mô hình để kiểm tra hoặc đề ra chính sách.

# MỞ ĐẦU

**Ví dụ:** Nghiên cứu tính quy luật của tiêu dùng.

1. Luận thuyết: một người sẽ tăng tiêu dùng khi thu nhập của người đó tăng lên, song không thể tăng nhiều bằng mức tăng của thu nhập (Keynes)

2.  $Y$  là tiêu dùng,  $X$  là thu nhập

Mô hình toán: 
$$Y = \beta_1 + \beta_2 X$$

Điều kiện: 
$$0 < \beta_2 < 1$$

Gọi là *Hàm tiêu dùng* của Keynes.

3. Mô hình KTL: 
$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Trong đó  $u_i$  là sai số ngẫu nhiên.

4. Thu thập số liệu.

# MỞ ĐẦU

## Ví dụ:

5. Ước lượng mô hình.

Dùng phương pháp bình phương nhỏ nhất thông thường (OLS) tìm được các ước lượng của các hệ số.

6. Phân tích kết quả

- $H_0 : \beta_2 = 0; H_1 : \beta_2 > 0.$
- $H_0 : \beta_2 = 1; H_1 : \beta_2 < 1.$
- $H_0 : Sai số ngẫu nhiên phân phối chuẩn$   
 $H_1 : Sai số ngẫu nhiên không pp chuẩn.$

....

Mục đích: kiểm chứng lại mô hình hoặc lý thuyết kinh tế.

7. Dự báo.

Dự báo giá trị của  $Y$  khi biết giá trị của  $X$

8. Kiểm soát và đề xuất chính sách.

# MỞ ĐẦU

## 3. SỐ LIỆU CHO PHÂN TÍCH KTL

### ❖ Phân loại

- Số liệu theo thời gian (time series).
- Số liệu chéo (theo không gian – cross section data).
- Số liệu hỗn hợp (pooled data).

### ❖ Nguồn gốc: Điều tra, mua, từ nguồn được phát hành như Niên giám thống kê.

### ❖ Tính chất của số liệu

- Ngẫu nhiên phi thực nghiệm.
- Phù hợp mục đích nghiên cứu.

Chú ý: đặc điểm chung của các số liệu kinh tế - xã hội là kém tin cậy.



# MỞ ĐẦU

## 4. PHÂN TÍCH HỒI QUY – *Regression Analysis*

*Phân tích hồi quy là phân tích mối liên hệ phụ thuộc giữa một biến gọi là biến phụ thuộc vào một hoặc một số biến khác gọi là (các) biến độc lập.*

- **Biến phụ thuộc** (dependent variable)

Thường được kí hiệu là  $Y$

Còn gọi là biến được giải thích, biến nội sinh

- **Biến độc lập** (independent variable(s))

Thường kí hiệu là  $X, X_2, X_3, \dots$

Còn gọi là biến giải thích, biến ngoại sinh, biến hồi quy

# MỞ ĐẦU

- **Mục đích hồi quy**

- Ước lượng trung bình biến phụ thuộc trong những điều kiện xác định của biến độc lập.
- Ước lượng các tham số.
- Kiểm định về mối quan hệ.
- Dự báo giá trị biến phụ thuộc khi biến độc lập thay đổi.

# CHƯƠNG I

## MÔ HÌNH HỒI QUY

### TUYẾN TÍNH HAI BIẾN

#### 1.1. MÔ HÌNH VÀ MỘT SỐ KHÁI NIỆM

##### 1.1.1. Mô hình hồi quy

Thí dụ: Muốn đánh giá tác động của lượng phân bón lên năng suất lúa trên tổng thể các ruộng lúa ở đồng bằng sông Hồng

Lượng phân bón tăng thì năng suất tăng nên

$$NS = f(PB)$$

Thực tế không biết dạng hàm  $f(.)$

# CHƯƠNG I

Giả sử  $f(\cdot)$  có dạng tuyến tính:  $NS = \beta_1 + \beta_2 PB$

Thực tế năng suất còn phụ thuộc vào các yếu tố khác nên

$$NS = \beta_1 + \beta_2 PB + u$$

**Tổng quát:**  $Y = \beta_1 + \beta_2 X + u$

(mô hình hồi quy tổng thể - PRM)

\* Các biến số: 2 loại

+ Y: biến phụ thuộc, biến được giải thích, biến phản ứng

+ X: biến độc lập, biến giải thích, biến điều khiển

\* Sai số ngẫu nhiên (u) đại diện cho các yếu tố có tác động đến Y, ngoài X.

\* Các hệ số hồi quy:  $\beta_1, \beta_2$

# CHƯƠNG I

## 1.1.2. Hàm hồi quy tổng thể (PRF)

Với giả thiết  $E(u|X) = 0$  thì mô hình trên trở thành

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

Phương trình trên gọi là hàm hồi quy tổng thể (PRF)

### Các hệ số hồi quy

- $\beta_1$ : hệ số chặn, nó bằng giá trị trung bình của biến phụ thuộc  $Y$  khi biến độc lập  $X$  nhận giá trị bằng 0
- $\beta_2$ : hệ số góc, khi biến độc lập  $X$  tăng (giảm) 1 đơn vị thì biến phụ thuộc  $Y$  tăng (giảm)  $\beta_2$  đơn vị

# CHƯƠNG I

**Ví dụ 1.1.** Tổng thể giả định có 30 thửa ruộng

Với số liệu (giáo trình trang 26) ta thấy quan hệ giữa lượng phân bón (PB) và năng suất (NS) trung bình như sau:

$$E(NS|PB=5) = 4,3 = 1,8 + 0,5*5$$

$$E(NS|PB=6) = 4,8 = 1,8 + 0,5*6$$

$$E(NS|PB=7) = 5,3 = 1,8 + 0,5*7$$

$$E(NS|PB=8) = 5,8 = 1,8 + 0,5*8$$

$$E(NS|PB=9) = 6,3 = 1,8 + 0,5*9$$

Ta có hàm hồi quy tổng thể sau

$$E(NS|PB) = 1,8 + 0,5*PB$$

Ý nghĩa các hệ số

# CHƯƠNG I

## 1.1.3. Hàm hồi qui mẫu (SRF)

Mẫu ngẫu nhiên kích thước  $n$ :  $(Y_i, X_i)$ ,  $i = 1, 2, \dots, n$

Các hệ số ước lượng cho các hệ số hồi quy tổng thể  $\beta_1, \beta_2$  là  $\hat{\beta}_1, \hat{\beta}_2$  tương ứng. Khi đó ta có

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X$$

Hay viết chi tiết cho từng quan sát như sau

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \quad ; \quad i = 1, 2, \dots, n$$

$\hat{\beta}_1, \hat{\beta}_2$  gọi là các hệ số hồi qui mẫu

$\hat{Y}_i$  là giá trị ước lượng của giá trị thực  $Y_i$

# CHƯƠNG I

## 1.1.4. Tính tuyến tính trong mô hình hồi quy

Hàm hồi quy được hiểu là tuyến tính theo tham số, nghĩa là theo các hệ số hồi quy.

### Ví dụ 1.2.

- Mô hình hồi quy tuyến tính

$$Y = \beta_1 + \beta_2 X^2 + u$$

$$\text{Log}(Y) = \beta_1 + \beta_2 \log(X) + u$$

- Mô hình hồi quy phi tuyến

$$Y = e^{\beta_1 + \beta_2 X} + u; \quad Y = \beta_0 + \frac{1}{\beta_1 + \beta_2 X} + u$$



# CHƯƠNG I

## 1.2. PHƯƠNG PHÁP ƯỚC LƯỢNG OLS

Xét mô hình hồi quy tổng thể

$$Y = \beta_1 + \beta_2 X + u$$

Ta cần ước lượng các hệ số  $\beta_1, \beta_2$

Trên mẫu ngẫu nhiên kích thước  $n$ , tại mỗi quan sát ta có

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

Gọi  $\hat{\beta}_1, \hat{\beta}_2$  là các ước lượng cần tìm của  $\beta_1, \beta_2$  thì

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

Sai lệch giữa giá trị thực tế  $Y_i$  và giá trị ước lượng của nó gọi là phần dư (residuals) kí hiệu bởi  $e_i$

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

# CHƯƠNG I

Tổng bình phương các phần dư là

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

Phương pháp OLS chủ trương tìm  $\hat{\beta}_1, \hat{\beta}_2$  để tổng bình phương các phần dư đạt nhỏ nhất

$$\begin{cases} \frac{\partial \sum e_i^2}{\partial \hat{\beta}_1} = 0 \\ \frac{\partial \sum e_i^2}{\partial \hat{\beta}_2} = 0 \end{cases} \Rightarrow \begin{cases} \hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \\ \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \end{cases} \quad \text{với } \begin{cases} x_i = X_i - \bar{X} \\ y_i = Y_i - \bar{Y} \end{cases}$$

# CHƯƠNG I

**Ví dụ 1.3.** Sử dụng 5 quan sát đầu tiên trong bảng 1.2 (trang 34 – giáo trình) về số năm làm việc và thu nhập sau khi tốt nghiệp ngành ngân hàng

KN: kinh nghiệm (năm); TN: thu nhập (triệu đồng)

STT	KN	TN	kn	tn	kn*tn	kn^2
1	7	106.7	0.2	9.3	1.86	0.04
2	10	94.5	3.2	-2.9	-9.28	10.24
3	8	126.1	1.2	28.7	34.44	1.44
4	4	50.3	-2.8	-47.1	131.88	7.84
5	5	109.4	-1.8	12	-21.6	3.24
Tổng	34	487			<b>137.3</b>	<b>22.8</b>
Trung bình	6.8	97.4				

Thu được hàm hồi quy mẫu sau:  $\widehat{TN}_i = 56.45 + 6.02KN_i$

Có thể dùng excel → tool → data analysis → regression → bảng KQ

# CHƯƠNG I

- $\hat{\beta}_1 = 56.45$  nghĩa là mức lương trung bình của những người mới tốt nghiệp ngành ngân hàng (KN = 0 năm) là 56.45 triệu đồng.
- $\hat{\beta}_2 = 6.02$  nghĩa là cứ thêm 1 năm làm việc (KN tăng thêm 1 đơn vị) thì mức lương trung bình gia tăng 6,02 triệu đồng.
- Bảng so sánh giá trị ước lượng và giá trị thực tế của TN

STT	TN	$\widehat{TN}$	e
1	106.7	98.59	8.11
2	94.5	116.65	-22.15
3	126.1	104.61	21.49
4	50.3	80.53	-30.23
5	109.4	86.55	22.85

# CHƯƠNG I

## 1.3. TÍNH KHÔNG CHỆCH VÀ ĐỘ CHÍNH XÁC CỦA ƯỚC LƯỢNG OLS

### 1.3.1. Các giả thiết của phương pháp OLS

Xét mô hình hồi quy tuyến tính

$$Y = \beta_1 + \beta_2 X + u$$

**Giả thiết 1:** Mô hình được ước lượng trên cơ sở mẫu ngẫu nhiên  $\{(X_i, Y_i), i = 1, 2, \dots, n\}$

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

**Giả thiết 2:**  $E(u|X) = 0$

Khi đó ta có  $E(u_i|X_i) = 0$  với mọi  $i$

Khi giả thiết 2 thỏa mãn thì:  $E(u) = 0$

$$\text{Cov}(X, u) = 0$$

**Giả thiết 3:**  $\text{var}(u|X) = \sigma^2$

# CHƯƠNG I

## 1.3.2. Tính không chệch của các ước lượng OLS

**Định lý 1.1.** Khi giả thiết 2 thỏa mãn thì các ước lượng  $\widehat{\beta}_1, \widehat{\beta}_2$  là các ước lượng không chệch của  $\beta_1, \beta_2$

$$E(\widehat{\beta}_1) = \beta_1; E(\widehat{\beta}_2) = \beta_2$$

## 1.3.3. Độ chính xác của các ước lượng OLS

**Định lý 1.2.** Khi các giả thiết 1, 2, 3 thỏa mãn thì

$$\text{var}(\widehat{\beta}_1) = \frac{\sum X_i^2}{\sum x_i^2} \sigma^2; \quad \text{var}(\widehat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$$

Trong đó  $\sigma^2$  chưa biết nên được thay bởi UL không chệch của nó là

$$\widehat{\sigma}^2 = \frac{e_1^2 + e_2^2 + \dots + e_n^2}{n-2}$$

# CHƯƠNG I

## *Một số tính chất của hàm hồi quy mẫu*

+ Tổng các phần dư ( $e_i$ ) bằng 0.

+  $\text{cov}(X, e) = 0$

Trong đó  $e = (e_1, \dots, e_n)$ ;  $X = (X_1, \dots, X_n)$ .

+ Đường hồi quy luôn đi qua trung bình mẫu  $(\bar{X}, \bar{Y})$

+ Trung bình của các giá trị ước lượng của biến phụ thuộc bằng trung bình mẫu

$$\bar{\hat{Y}} = \bar{Y}$$

# CHƯƠNG I

## 1.4. ĐỘ PHÙ HỢP CỦA HÀM HỒI QUY MẪU – HỆ SỐ XÁC ĐỊNH $R^2$

Hệ số xác định, kí hiệu  $R^2$ , đánh giá một cách định lượng sự phù hợp của hàm hồi quy mẫu đối với số liệu mẫu.

Công thức:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}; \quad (R^2 \geq 0)$$

Trong đó:  $TSS = ESS + RSS$

$TSS = \sum (Y_i - \bar{Y})^2$  : độ dao động trong mẫu của biến phụ thuộc

$RSS = \sum e_i^2$  : tổng bình phương các phần dư

$ESS = \sum (\hat{Y}_i - \bar{\hat{Y}})^2$  : độ dao động của giá trị ước lượng



# CHƯƠNG I

**Ví dụ 1.3.** Với mẫu kích thước  $n = 5$  ta tìm được

	$df$	$SS$
Regression ( <b>ESS</b> )	1	826.811
Residual ( <b>RSS</b> )	3	2454.189
Total ( <b>TSS</b> )	4	3281

$$R^2 = 826.8 / 3281 = 0,252$$

*Ý nghĩa:* Biến số năm kinh nghiệm giải thích được 25,2% sự thay đổi của mức lương và 74,8% còn lại trong sự biến đổi của lương là do các yếu tố không đưa vào mô hình gây ra (như trình độ học vấn, giới tính, địa điểm làm việc,...).

# CHƯƠNG I

- **Ghi chú 1.** Với mô hình hồi quy 2 biến có hệ số chặn thì

$$R^2 = \frac{\left(\sum x_i y_i\right)^2}{\left(\sum x_i^2\right)\left(\sum y_i^2\right)} = r_{X,Y}^2$$

- **Ghi chú 2.** Với mô hình hồi quy 2 biến có hệ số chặn thì

$$R^2 = 0 \Leftrightarrow \widehat{\beta}_2 = 0$$

- **Ghi chú 3.** Nếu mô hình không có hệ số chặn thì các phát biểu trên về  $R^2$  không đúng nữa;  $R^2$  có thể nhận giá trị âm do đó nó không còn giữ nguyên ý nghĩa ban đầu của nó nữa

# CHƯƠNG I

## 1.5. MỘT SỐ VẤN ĐỀ BỔ SUNG

### 1.5.1. Đơn vị đo lường trong phân tích hồi quy

Khi thay đổi đơn vị đo lường của biến số thì hệ số ước lượng cũng thay đổi. Tuy nhiên sự thay đổi này không làm thay đổi kết quả phân tích hồi quy.

### 1.5.2. Hệ số chặn và mô hình hồi quy

- + Hệ số chặn  $\beta_1 = E(Y|X=0)$ , nó chỉ có ý nghĩa khi biến độc lập có nhận giá trị 0.
- + Trong phân tích hồi quy chúng ta không quan tâm nhiều đến ý nghĩa của hệ số chặn mà quan tâm chủ yếu đến hệ số góc.

# CHƯƠNG II

## MÔ HÌNH HỒI QUY BỘỊ

### 2.1. SỰ CẦN THIẾT CỦA MHHQ BỘỊ

#### 2.1.1. Mô hình hai biến – vấn đề về kì vọng sai số khác 0

Theo chương I, nếu giả thiết 2 [ $E(u|X)=0$ ] thỏa mãn thì các ước lượng OLS là ước lượng không chệch. Tuy nhiên với mô hình 2 biến thì trong nhiều trường hợp thực tế giả thiết này không thỏa mãn.

Chú ý: nếu giả thiết 2 thỏa mãn thì  $cov(X,u) = 0$ .

## CHƯƠNG II

**Ví dụ 2.1.** Khi xem xét tác động của thu nhập (TN) lên chi tiêu (CT) của hộ gia đình, theo chương 1, ta có thể sử dụng mô hình sau:

$$CT = \beta_1 + \beta_2 TN + u$$

Yếu tố tài sản (TS) của hộ gia đình cũng có tác động lên CT nên nó là một thành phần của  $u$ . Thực tế, một gia đình có TN cao thường sở hữu nhiều TS, nghĩa là TN và TS có quan hệ tương quan cao:

$$\text{Cov}(TN, u) \neq 0$$

Giả thiết 2 bị vi phạm, để khắc phục ta đưa thêm biến TS vào và có mô hình sau:

$$CT = \beta_1 + \beta_2 TN + \beta_3 TS + u$$

## CHƯƠNG II

**Ví dụ 2.2.** Xét quan hệ giữa vốn (K) và sản lượng (Q) của ngành dệt may

$$Q = \beta_1 + \beta_2 K + u$$

Sản lượng Q còn phụ thuộc vào các biến khác như số lao động (L). Thực tế là các doanh nghiệp nhiều máy móc thì cũng thuê nhiều lao động, nghĩa là

$$\text{Cov}(K, L) > 0$$

Mà sai số u chứa L, nên  $\text{cov}(K, u) \neq 0$ , nghĩa là giả thiết 2 bị vi phạm. Ta có thể sử dụng mô hình sau:

$$Q = \beta_1 + \beta_2 K + \beta_3 L + u$$

→ Nếu sai số ngẫu nhiên có tương quan với biến độc lập thì giả thiết 2 sẽ bị vi phạm.

## CHƯƠNG II

- Biến độc lập nội sinh là biến độc lập có tương quan với sai số ngẫu nhiên.
- Khi mô hình có biến độc lập nội sinh thì giả thiết 2 bị vi phạm, và do đó các ước lượng OLS là ước lượng chệch.
- Trong phân tích kinh tế - xã hội, hiện tượng biến độc lập nội sinh xảy ra khá phổ biến với mô hình 2 biến. Để giải quyết vấn đề này ta phải đưa thêm các biến quan trọng khác vào mô hình.
- Mô hình có nhiều hơn 1 biến độc lập gọi là mô hình hồi quy bội hay mô hình hồi quy đa biến.

# CHƯƠNG II

## 2.1.2. Một số ưu việt khác của mô hình hồi quy bội

- Mô hình hồi quy bội thường có chất lượng dự báo tốt hơn.
- Mô hình hồi quy bội cung cấp các dự báo hữu ích hơn.
- Mô hình hồi quy bội cho phép sử dụng dạng hàm phong phú hơn.
- Mô hình hồi quy bội cho phép thực hiện các phân tích phong phú hơn.



# CHƯƠNG II

## 2.2. MÔ HÌNH HỒI QUY BỘI VÀ PHƯƠNG PHÁP ƯỚC LƯỢNG OLS

### 2.2.1. Mô hình và các giả thiết

*Mô hình hồi quy tuyến tính k biến*

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + u$$

*Các giả thiết của mô hình*

**Giả thiết 1:** Việc ước lượng được dựa trên cơ sở mẫu ngẫu nhiên.

**Giả thiết 2:**  $E(u|X_{2i}, \dots, X_{ki}) = 0$

**Giả thiết 3:**  $\text{var}(u|X_{2i}, \dots, X_{ki}) = \sigma^2$

**Giả thiết 4:** Giữa các biến độc lập không có mối quan hệ đa cộng tuyến hoàn hảo.

# CHƯƠNG II

## *Ý nghĩa của các hệ số hồi quy*

Với giả thiết 2 ta có

$$E(Y|X_2, \dots, X_k) = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

- $\beta_1$  là giá trị trung bình của biến  $Y$  khi tất cả các biến độc lập trong mô hình nhận giá trị 0.
- Với giả định biến  $X_j$  liên tục

$$\beta_j = \frac{\partial E(Y | X_2, \dots, X_k)}{\partial X_j}; \quad j = 2, 3, \dots, k$$

$\beta_j$  thể hiện tác động riêng phần của biến  $X_j$  lên giá trị trung bình của biến phụ thuộc ( $Y$ ) khi các yếu tố  $X_s$  ( $s \neq j$ ) là không đổi. Do đó trong mô hình hồi quy bội các hệ số góc còn được gọi là hệ số hồi quy riêng.

# CHƯƠNG II

## 2.2.2. Phương pháp OLS và giải thích kết quả ước lượng

### *Phương pháp OLS cho mô hình hồi quy bội*

Xét mô hình k biến

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + u$$

Hàm hồi quy mẫu

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_k$$

Tại mỗi quan sát i ta có

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

Phần dư:  $e_i = Y_i - \hat{Y}_i$

# CHƯƠNG II

Phương pháp OLS nhằm xác định các giá trị  $\widehat{\beta}_j (j = \overline{1, k})$  sao cho tổng bình phương các phần dư là bé nhất

$$\begin{aligned} \sum_{i=1}^n e_i^2 &= \sum \left( Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_{2i} - \dots - \widehat{\beta}_k X_{ki} \right)^2 \\ &= \text{Min}_{\widetilde{\beta}_1, \dots, \widetilde{\beta}_k} \sum \left( Y_i - \widetilde{\beta}_1 - \widetilde{\beta}_2 X_{2i} - \dots - \widetilde{\beta}_k X_{ki} \right)^2 \end{aligned}$$

Các giá trị  $\widehat{\beta}_j$  là nghiệm của hệ k phương trình sau:

$$\begin{cases} \sum_{i=1}^n (Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_{2i} - \dots - \widehat{\beta}_k X_{ki}) = 0 \\ \sum_{i=1}^n X_j (Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_{2i} - \dots - \widehat{\beta}_k X_{ki}) = 0; j = 2, 3, \dots, k \end{cases}$$

# CHƯƠNG II

## Giải thích kết quả ước lượng

**Ví dụ 2.3.** Sử dụng tập số liệu chitieu.wfl với 30 quan sát về chi tiêu của hộ gia đình ta có báo cáo eviews sau đây:

Dependent Variable: CT

Method: Least Squares

→ **Biến phụ thuộc**

Sample: 1 30

Included observations: 30

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	57.76804	12.54448	4.605058	0.0001
TN	0.393265	0.055562	7.077935	0.0000
TS	0.009421	0.011236	0.838464	0.4091
R-squared	0.829007			

**Biến độc lập**

C  
TN  
TS

**Các hệ số ước lượng ( $\hat{\beta}_i$ )**

## CHƯƠNG II

- Trong đó CT: chi tiêu của hộ gia đình (triệu đồng/ năm)  
TN: thu nhập từ lao động (tr. đ./ năm)  
TS: giá trị tài sản (triệu đồng)

- Từ bảng trên ta có

$$CT_i = 57.768 + 0.393 * TN_i + 0.09 * TS_i + e_i$$

(12.544) (0.056) (0.011)

- Ý nghĩa các hệ số

+  $\widehat{\beta}_1 = 57.768$  nghĩa là các hộ không có thu nhập và không có tài sản thì chi tiêu trung bình 57.768 triệu đồng/ năm. Giá trị này không có ý nghĩa vì trên thực tế hầu như không có hộ gia đình nào đồng thời có  $TN = 0$ ,  $TS = 0$ .

## CHƯƠNG II

+  $\hat{\beta}_2 = 0.393$  : khi thu nhập của hộ gia đình tăng thêm 1 triệu đồng/năm và giá trị tài sản không thay đổi thì chi tiêu trung bình của hộ gia đình tăng 0,393 triệu đồng/năm.

+  $\hat{\beta}_3 = 0.009$  : khi giá trị tài sản của hộ gia đình tăng thêm 1 triệu đồng và thu nhập không thay đổi thì chi tiêu trung bình của hộ gia đình tăng 0.009 triệu đồng/năm.

Cả 2 hệ số trên đều đánh giá tác động riêng phần của TN và TS lên CT của hộ gia đình trong điều kiện các yếu tố khác không thay đổi.

Các giá trị trong ngoặc đơn là sai số chuẩn của các hệ số ước lượng tương ứng.

# CHƯƠNG II

## 2.2.3. Độ phù hợp của hàm hồi quy

*Hệ số xác định bội  $R^2$*

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

Với  $TSS = ESS + RSS$  và tính như sau

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

$$ESS = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2; \quad RSS = \sum_{i=1}^n e_i^2$$



# CHƯƠNG II

- $R^2$  nhận giá trị trong đoạn  $[0,1]$
- $R^2$  là giá trị gắn liền với mẫu, nó đo mức độ phù hợp của mô hình với số liệu mẫu.
- Ta có thể chứng minh:  $R^2 = r^2(Y, \hat{Y})$
- ***Ý nghĩa của  $R^2$*** 
  - +  $R^2$  bằng phần trăm sự thay đổi của biến phụ thuộc được giải thích bởi các biến độc lập trong mô hình.
  - +  $R^2$  thể hiện mức độ tương quan tuyến tính giữa biến phụ thuộc với các biến độc lập.

**Ví dụ 2.3.** có  $R^2 = 0,829$  nghĩa là TN và TS giải thích được 82,9% sự thay đổi của CT.

# CHƯƠNG II

- *$R^2$  đã hiệu chỉnh*

Việc đưa thêm một biến bất kì vào mô hình nói chung sẽ làm gia tăng  $R^2$ , không kể nó có giúp giải thích thêm biến phụ thuộc hay không.

Việc đưa thêm một biến mới vào mô hình cũng tạo ra những tác động không tốt đến chất lượng của các ước lượng (chương 3).

Để tổng hòa 2 tác động trên, người ta đưa ra khái niệm  $R^2$  đã hiệu chỉnh, kí hiệu  $\bar{R}^2$

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

Giá trị này được sử dụng khi so sánh các mô hình hồi quy có số biến khác nhau

# CHƯƠNG II

## 2.2.4. Tính tốt nhất của ước lượng OLS – định lý Gauss – Markov

***Định lý Gauss – Markov:*** Khi các giả thiết 1 – 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.

# CHƯƠNG II

***Độ chính xác của ước lượng  $\hat{\beta}_j$  được đo bởi  $\text{var}(\hat{\beta}_j)$***

$$\text{var}(\hat{\beta}_j) = \frac{\sigma^2}{(1 - R_j^2) \sum x_{ji}^2}$$

Trong đó: \*  $x_{ji} = X_{ji} - \bar{X}_j$

\*  $R_j^2$  là hệ số xác định của mô hình hồi quy  $X_j$  theo các biến độc lập còn lại trong mô hình.

Ví dụ:  $R_2^2$  là hệ số xác định của mô hình hồi quy

$$X_2 = \alpha_1 + \alpha_2 X_3 + \dots + \alpha_{k-1} X_k + v$$

\*  $\sigma^2$  là tham số chưa biết nên khi tính toán được thay bởi ước lượng không chệch của nó là

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

# CHƯƠNG II

- Sai số của ước lượng

$$se(\hat{\beta}_j) = \sqrt{\frac{\hat{\sigma}^2}{(1-R_j^2)\sum x_{ji}^2}} = \sqrt{\frac{RSS / (n-k)}{(1-R_j^2)\sum x_{ji}^2}}; \quad j = 2, 3, \dots, k$$

- ***Các yếu tố xác định độ chính xác của ước lượng***

- Phương sai  $\sigma^2$  của sai số ngẫu nhiên (SSNN),  $\sigma^2$  càng bé thì sai số của UL càng bé, độ chính xác của UL càng cao.

- *Nhân tử phóng đại phương sai (VIF)*

VIF<sub>j</sub> càng lớn thì độ chính xác của UL càng nhỏ

$$VIF_j = \frac{1}{1-R_j^2}$$

Hay quan hệ tương quan giữa  $X_j$  với các biến độc lập còn lại càng chặt ( $R_j^2$  càng lớn) thì độ chính xác càng giảm.

- $\sum x_{ji}^2$  càng lớn thì độ chính xác của UL càng lớn

# CHƯƠNG II

## 2.3. MỘT SỐ DẠNG CỦA MÔ HÌNH HỒI QUY

### 2.3.1. Mô hình dạng log – log

Dạng mô hình

$$\text{Log}(Y) = \beta_1 + \beta_2 \log(X_2) + \dots + \beta_k \log(X_k) + u$$

Với  $j = 2, 3, \dots, k$  ta có

$$\beta_j = \frac{\partial \ln(Y)}{\partial \ln(X_j)} = \frac{\partial Y / Y}{\partial X_j / X_j} \rightarrow \frac{dY}{Y} = \beta_j \frac{dX_j}{X_j}$$

$\beta_j$  có nghĩa là khi  $X_j$  gia tăng (giảm) 1% và các yếu tố khác không đổi thì trung bình của  $Y$  gia tăng (giảm)  $\beta_j$  %.

$\beta_j$  gọi là hệ số co giãn của  $Y$  theo  $X_j$

# CHƯƠNG II

## Ví dụ 2.4. Sử dụng bộ số liệu ở VD 2.3 cho hàm log – log ta có

Dependent Variable: LOG(CT)

Method: Least Squares

Sample: 1 30

Included observations: 30

---

---

Variable	Coefficient
C	1.428828
LOG(TN)	0.660106
LOG(TS)	0.009475

---

---

R-squared 0.832970

$$\text{Log}(CT) = 1.43 + 0.66 * \text{log}(TN) + 0.009 * \text{log}(TS) + e$$

+ Khi TN tăng 1% và TS không đổi thì CT tăng 0.66%.

+ Khi TS tăng 1% và TN không đổi thì CT tăng 0.009%.

$$+ R^2 = 0.833$$

Các mô hình log – log sử dụng với các mô hình nghiên cứu về hàm cầu, hàm sản xuất dạng Cobb - Douglas.

# CHƯƠNG II

## 2.3.2. Mô hình dạng bán loga

Dạng mô hình

- Dạng log-lin:  $\text{Log}(Y) = \beta_1 + \beta_2 X_2 + u$
- Dạng lin-log:  $Y = \beta_1 + \beta_2 \ln(X_2) + u$
- Hoặc dạng hỗn hợp

$$\text{Log}(Y) = \beta_1 + \beta_2 \ln(X_2) + \beta_3 X_3 + u$$

Các mô hình dạng bán loga thường được sử dụng khi nghiên cứu quan hệ giữa tiền lương và số năm kinh nghiệm, hoặc tiền lương và số năm học ở trường, ...



# CHƯƠNG II

## *Khi nào lựa chọn dạng hàm có biến số dạng logarit?*

- Có gợi ý từ lý thuyết kinh tế về mối quan hệ giữa các biến số như dạng hàm Cobb – Douglas hay lý thuyết lượng hóa về cầu tiền.
- Thường dùng khi các biến số đều nhận giá trị dương như dân số, GDP, số lao động, ... hoặc các biến số có đuôi lệch như thu nhập, mức lương, ...
- Ưu điểm: các kết quả ước lượng không phụ thuộc vào đơn vị đo của các biến số.
- Với các biến số nhận cả giá trị âm (như lợi nhuận công ty, lợi tức cổ phiếu, ...) thì không lấy log trực tiếp được.

# CHƯƠNG II

## 2.3.3. Mô hình dạng đa thức

Các dạng hàm có chứa các số mũ bậc cao của biến độc lập.

Quan hệ giữa tuổi (Age) và tiền lương (W) của người lao động

$$W = \beta_1 + \beta_2 \text{Age} + \beta_3 \text{Age}^2 + u$$

Mô hình tuyến tính với các tham số nên ta dùng OLS để ƯL

Tác động biên của tuổi lên mức lương là thay đổi

$$\frac{\partial E(W / \text{Age})}{\partial \text{Age}} = \beta_2 + 2\beta_3 \text{Age}$$

- VD khác (giáo trình/105)

Thường dùng để nghiên cứu hàm chi phí hoặc tiền lương.

# CHƯƠNG II

## **2.3.4. Mô hình phi tuyến**

Với các mô hình phi tuyến, phương pháp OLS không phù hợp nữa.

# CHƯƠNG II

## 2.4. TÍNH VỮNG CỦA ƯỚC LƯỢNG OLS

**Định lý 2.4.** Khi các giả thiết 1 – 4 thỏa mãn thì các ước lượng OLS là ước lượng vững.

**Định lý 2.5.** Nếu trong định lý 2.4, giả thiết 2 được thay bởi giả thiết sau:

(i)  $\text{cov}(X_j, u) = 0$  với  $j = 2, 3, \dots, k$ .

(ii)  $E(u) = 0$

Thì ước lượng OLS vẫn là ước lượng vững.

Tính vững của ước lượng phản ánh chất lượng của ước lượng khi mẫu lớn.

# CHƯƠNG II

## 2.5. MÔ HÌNH HQ SỬ DỤNG NGÔN NGỮ MA TRẬN

Xét mô hình

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i ; \quad i = 1, 2, \dots, n.$$

Đặt

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix}_{nx1} ; \mathbf{X} = \begin{pmatrix} 1 & X_{21} & \dots & X_{k1} \\ 1 & X_{22} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots \\ 1 & X_{2n} & \dots & X_{kn} \end{pmatrix}_{nxk} ; \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{pmatrix}_{kx1} ; \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{pmatrix}_{nx1}$$

Ta có dạng ma trận như sau

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

Hàm hồi quy mẫu tương ứng là  $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$

## CHƯƠNG II

Phần dư là  $e = Y - \hat{Y} = Y - X\hat{\beta}$

Kí hiệu  $e' = (e_1, e_2, \dots, e_n)$

Do giả thiết 4 nên tồn tại ma trận nghịch đảo  $(X'X)^{-1}$ . Khi đó  $e.e'$  đạt cực tiểu tại

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

Sai số của ước lượng là

$$\text{var}(\hat{\beta} | X) = \sigma^2 (X'X)^{-1}$$

# CHƯƠNG III

## SUY DIỄN THỐNG KÊ VÀ DỰ BÁO TỪ MÔ HÌNH HỒI QUY

### 3.1. QUY LUẬT PHÂN PHỐI XÁC SUẤT CỦA MỘT SỐ THỐNG KÊ MẪU

Xét MHHQ tuyến tính k biến

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + u$$

Hàm hồi quy mẫu là

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k X_{ki}; \quad i = 1, 2, \dots, n$$

**Giả thiết 5:** SSNN phân phối chuẩn

Kết hợp các giả thiết ta có  $u_i \sim N(0; \sigma^2)$

# CHƯƠNG III

- **Định lý 3.1.** Khi các giả thiết 1 – 5 thỏa mãn thì ta có

$$\widehat{\beta}_j \sim N(\beta_j; \text{var}(\widehat{\beta}_j)); \quad j = 1, 2, \dots, k$$

- **Định lý 3.2.** Khi các giả thiết 1 – 5 thỏa mãn thì ta có

$$t = \frac{\widehat{\beta}_j - \beta_j}{\text{se}(\widehat{\beta}_j)} \sim T(n - k); \quad j = 1, 2, \dots, k$$

- Tương tự

$$t = \frac{a\widehat{\beta}_j + b\widehat{\beta}_s - (a\beta_j + b\beta_s)}{\text{se}(a\widehat{\beta}_j + b\widehat{\beta}_s)} \sim T(n - k)$$



# CHƯƠNG III

## 3.2. BÀI TOÁN XÂY DỰNG KHOẢNG TIN CẬY CHO CÁC HỆ SỐ HỒI QUY

### 3.2.1. Khoảng tin cậy cho 1 hệ số hồi quy: đánh giá tác động khi 1 biến độc lập thay đổi

Với độ tin cậy  $(1-\alpha)$ , các khoảng tin cậy đối xứng, tối đa, tối thiểu lần lượt là

$$\widehat{\beta}_j - t_{\alpha/2}^{(n-k)} se(\widehat{\beta}_j) < \beta_j < \widehat{\beta}_j + t_{\alpha/2}^{(n-k)} se(\widehat{\beta}_j); \quad j = 1, 2, \dots, k$$

$$-\infty < \beta_j < \widehat{\beta}_j + t_{\alpha}^{(n-k)} se(\widehat{\beta}_j)$$

$$\widehat{\beta}_j - t_{\alpha}^{(n-k)} se(\widehat{\beta}_j) < \beta_j < +\infty$$

Với độ tin cậy  $(1-\alpha)*100\%$  khi  $X_j$  tăng 1 đơn vị và các biến khác không đổi thì giá trị trung bình của biến phụ thuộc thay đổi trong khoảng nào

# CHƯƠNG III

**Ví dụ 3.1.** Sử dụng ví dụ 2.3 ta có mô hình sau

$$CT = 57.768 + 0.393*TN + 0.009*TS + e$$

(12.544) (0.056) (0.011)

Khoảng tin cậy (đối xứng) của các hệ số là

$$57.768 - t_{0,025}^{(30-3)} * 12.544 < \beta_1 < 57.768 + t_{0,025}^{(30-3)} * 12.544$$

(2.052) → 32.028 <  $\beta_1$  < 83.508

$$0.393 - t_{0,025}^{(30-3)} * 0.056 < \beta_2 < 0.393 + t_{0,025}^{(30-3)} * 0.056 \rightarrow 0.278 < \beta_2 < 0.508$$

$$0.009 - t_{0,025}^{(30-3)} * 0.011 < \beta_3 < 0.009 + t_{0,025}^{(30-3)} * 0.011 \rightarrow -0.014 < \beta_3 < 0.032$$

# CHƯƠNG III

## 3.2.2. Khoảng tin cậy cho biểu thức của hai hệ số hồi quy: đánh giá tác động khi hai biến độc lập cùng thay đổi

Với  $a, b$  là hằng số và độ tin cậy  $(1-\alpha)$  ta có

$$(a\hat{\beta}_2 + b\hat{\beta}_3) - t_{\alpha/2}^{(n-k)} se(a\hat{\beta}_2 + b\hat{\beta}_3) < a\beta_2 + b\beta_3 < (a\hat{\beta}_2 + b\hat{\beta}_3) + t_{\alpha/2}^{(n-k)} se(a\hat{\beta}_2 + b\hat{\beta}_3)$$

$$se(a\hat{\beta}_2 + b\hat{\beta}_3) = \sqrt{a^2 \text{var}(\hat{\beta}_2) + b^2 \text{var}(\hat{\beta}_3) + 2ab \text{cov}(\hat{\beta}_2, \hat{\beta}_3)}$$

**Ví dụ 3.1.** Nếu tài sản tăng thêm 10 triệu đồng và thu nhập giảm 1 triệu/năm thì chi tiêu trung bình một năm thay đổi  $(-\beta_2 + 10\beta_3)$  đơn vị. Với

$$\text{cov}(\hat{\beta}_2, \hat{\beta}_3) = -0,0005; \quad \text{var}(\hat{\beta}_2) = 0.003; \quad \text{var}(\hat{\beta}_3) = 0.00013$$

$$se(-\hat{\beta}_2 + 10\hat{\beta}_3) = 0,1612; \quad -\hat{\beta}_2 + 10\hat{\beta}_3 = -0.303$$

$$t_{0.025}^{(30-3)} = 2.052; \quad -0.63 < -\beta_2 + 10\beta_3 < 0.03$$

Nghĩa là chi tiêu trung bình giảm.

# CHƯƠNG III

## 3.2.3. Ý nghĩa của khoảng tin cậy

Nếu lấy nhiều lần các mẫu một cách ngẫu nhiên từ cùng một tổng thể thì có khoảng  $(1-\alpha)*100\%$  số khoảng tin cậy được xây dựng từ các mẫu này là có chứa giá trị  $\beta_j$

## 3.2.4. Các yếu tố ảnh hưởng đến độ dài khoảng tin cậy

- Số bậc tự do  $(n-k)$  càng bé thì khoảng tin cậy càng rộng. Khi kích thước mẫu  $(n)$  nhỏ thì việc đưa thêm biến vào mô hình (tăng  $k$ ) làm cho KTC rộng. Nếu mẫu lớn thì thêm biến không ảnh hưởng đáng kể.
- Mọi tương quan tuyến tính giữa biến  $X_j$  và các biến độc lập còn lại trong mô hình (đo bởi  $R_j^2$ )

# CHƯƠNG III

## 3.3. BÀI TOÁN KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ VỀ HỆ SỐ HỒI QUY

### 3.3.1. Kiểm định giả thuyết về một hệ số hồi quy

Câu hỏi liệu biến  $X_j$  nào đó trong mô hình có tác động đến biến  $Y$  hay không? Ta kiểm định cặp giả thuyết:

$$H_0 : \beta_j = 0; \quad H_1 : \beta_j \neq 0$$

Tiêu chuẩn kiểm định: 
$$t = \frac{\widehat{\beta}_j - 0}{se(\widehat{\beta}_j)}$$

Nếu  $H_0$  đúng thì  $t \sim T(n-k)$

Miền bác bỏ:  $W_\alpha = \{t: |t| > t_{\alpha/2}(n-k)\}$

Nếu  $|t_{qs}| > t_{\alpha/2}(n-k)$  thì bác bỏ  $H_0$ , thừa nhận  $H_1$ .

# CHƯƠNG III

**Ví dụ 3.2.** Sử dụng ví dụ 2.3 ta có bảng sau

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	57.76804	12.54448	4.605058	0.0001
TN	0.393265	0.055562	7.077935	0.0000
TS	0.009421	0.011236	0.838464	0.4091

VD: Kiểm định cặp giả thuyết với mức ý nghĩa 5%

$$H_0 : \beta_2 = 0; \quad H_1 : \beta_2 \neq 0$$

Từ bảng kết quả trên ta có  $t_{qs} = 0.3933 / 0.0556 = 7.078$

Miền bác bỏ:  $W_\alpha = \{t : |t| > t_{0.025, (30-3)} = 2.052\}$

$t_{qs} \in W_\alpha$  nên bác bỏ  $H_0$ , nghĩa là  $\beta_2 \neq 0$  hay biến TN có tác động đến CT. Lúc đó ta nói  $\hat{\beta}_2$  có ý nghĩa thống kê.

# CHƯƠNG III

- Ta có thể sử dụng phương pháp kiểm định P\_value để kết luận về cặp giả thuyết trên. Trên báo cáo Eviews, giá trị P\_value của cặp giả thuyết  $H_0 : \beta_j = 0$ ;  $H_1 : \beta_j \neq 0$  được tính sẵn tại cột **Prob.**

Nếu **Prob.**  $< \alpha$  thì bác bỏ  $H_0$ , thừa nhận  $H_1$

Nếu **Prob.**  $\geq \alpha$  thì chưa có cơ sở bác bỏ  $H_0$

**Ví dụ 3.2.** Từ báo cáo ta thấy với mức ý nghĩa  $\alpha = 0.05$  thì các số  $\hat{\beta}_1 = 57,768$ ;  $\hat{\beta}_2 = 0,393$  đều có ý nghĩa thống kê, số  $\hat{\beta}_3 = 0,009$  không có ý nghĩa thống kê.

- Với trường hợp muốn kiểm định giả thuyết về độ lớn của  $\beta_j$  ta sử dụng bảng sau:

Chú ý: giả thuyết gốc  $H_0$  luôn chứa dấu ‘=’

# CHƯƠNG III

$H_0$	$H_1$	Tiêu chuẩn KĐ	Bác bỏ $H_0$ nếu
$\beta_j = \beta^*$	$\beta_j \neq \beta^*$	$t = \frac{\widehat{\beta}_j - \beta^*}{se(\widehat{\beta}_j)}$	$ t_{qs}  > t_{\alpha/2}(n-k)$
$\beta_j \geq \beta^*$	$\beta_j < \beta^*$		$t_{qs} < -t_{\alpha}(n-k)$
$\beta_j \leq \beta^*$	$\beta_j > \beta^*$		$t_{qs} > t_{\alpha}(n-k)$

**Ví dụ 3.2.** Có thể cho rằng khi TN của hộ gia đình tăng thêm 1 triệu và TS không đổi thì chi tiêu trung bình tăng ít hơn 0,5 triệu được không?

$$H_0 : \beta_2 \geq 0.5; \quad H_1 : \beta_2 < 0.5$$

$$t_{qs} = \frac{0.393 - 0.5}{0.056} = -1.911; \quad W_{\alpha} = \{t : t < -t_{0.05, (30-3)} = -1.706\}$$

KL: Có thể ...



# CHƯƠNG III

## 3.3.2. Kiểm định giả thuyết về một ràng buộc giữa các hệ số hồi quy – kiểm định T

Ta có bảng sau:

$H_0$	$H_1$	Tiêu chuẩn KĐ	Bác bỏ $H_0$ nếu
$a\beta_j + b\beta_s = a^*$	$a\beta_j + b\beta_s \neq a^*$	$t = \frac{(a\widehat{\beta}_j + b\widehat{\beta}_s) - a^*}{se(a\widehat{\beta}_j + b\widehat{\beta}_s)}$	$ t_{qs}  > t_{\alpha/2}(n-k)$
$a\beta_j + b\beta_s \geq a^*$	$a\beta_j + b\beta_s < a^*$		$t_{qs} < -t_{\alpha}(n-k)$
$a\beta_j + b\beta_s \leq a^*$	$a\beta_j + b\beta_s > a^*$		$t_{qs} > t_{\alpha}(n-k)$

## 3.3.3. Giá trị xác suất P của các thống kê kiểm định

# CHƯƠNG III

## 3.3.4. Kiểm định giả thuyết về nhiều ràng buộc của hệ số hồi quy – kiểm định F

Xét mô hình

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + u \quad (\mathbf{U} - \text{unrestricted})$$

Nếu đồng thời ta có  $\beta_2 = 0$  và  $\beta_3 = 0$  thì sẽ được mô hình

$$Y = \beta_1 + \beta_4 X_4 + \dots + \beta_k X_k + u \quad (\mathbf{R} - \text{restricted})$$

Trong trường hợp này cặp giả thuyết là

$$H_0 : \beta_2 = 0, \beta_3 = 0 ; \quad H_1 : \beta_2^2 + \beta_3^2 \neq 0$$

Mô hình (R) sẽ được thiết lập hoàn toàn tương tự trong trường hợp có  $\beta_2 = \beta_3 = \dots = \beta_m = 0$

Các bước kiểm định như sau

# CHƯƠNG III

- Bước 1: Thiết lập cặp giả thuyết thống kê
- Bước 2: Ước lượng hàm hồi quy không có ràng buộc (**U**) và mô hình hồi quy có ràng buộc (**R**) ta thu được  $RSS(U)$  và  $RSS(R)$
- Bước 3: Tính giá trị quan sát

$$F_{qs} = \frac{[RSS(R) - RSS(U)] / m}{RSS(U) / (n - k(U))}$$

Với  $m$  là số ràng buộc,  $k(U) = k$  là số hệ số HQ trong (U)

- Bước 4: Nếu  $H_0$  đúng thì  $F \sim F(m, n-k(U))$  nên  
với mức ý nghĩa  $\alpha$  nếu  $F_{qs} > f_{\alpha}(m, n-k(U))$  thì bác bỏ  $H_0$

Có thể kiểm định bằng phương pháp P\_value

# CHƯƠNG III

**Ví dụ 3.3.** Sử dụng tập số liệu chitieu.wfl với các biến

CT: chi tiêu của hộ gia đình

TN: thu nhập từ lao động

TS: giá trị tài sản

SONGUOI: số người trong hộ gia đình

Mô hình

$$CT = \beta_1 + \beta_2 * TN + \beta_3 * TS + \beta_4 * SONGUOI + u \quad (\mathbf{U})$$

Giả sử cần kiểm định ý kiến cho rằng TS và SONGUOI đồng thời không ảnh hưởng đến CT. Ta có cặp giả thuyết

$$H_0 : \beta_3 = \beta_4 = 0; \quad H_1 : \beta_3^2 + \beta_4^2 \neq 0$$

# CHƯƠNG III

- Mô hình không có ràng buộc là

$$CT = 47.6 + 0.38*TN + 0.009*TS + 4.13*SONGUOI + e \quad (\mathbf{U})$$

$$RSS(\mathbf{U}) = 37775.29; \quad R^2(\mathbf{U}) = 0.830125$$

- Mô hình hồi quy có ràng buộc là

$$CT = 58.9 + 0.43*TN + e \quad (\mathbf{R})$$

$$RSS(\mathbf{R}) = 39013.88; \quad R^2(\mathbf{R}) = 0.824555$$

- Cặp giả thuyết  $H_0 : \beta_3 = \beta_4 = 0; \quad H_1 : \beta_3^2 + \beta_4^2 \neq 0$

$$F_{qs} = \frac{(39013.88 - 37775.29) / 2}{37775.29 / (30 - 4)} = 0.43; \quad f_{0.05}(2; 30 - 4) = 3.49$$

Do  $F_{qs} < f_{0.05}(2; 26)$  nên chưa có cơ sở để bác bỏ  $H_0$ , nghĩa là  $\beta_3$  và  $\beta_4$  (có) đồng thời bằng 0.

# CHƯƠNG III

- Kiểm định F sử dụng  $R^2$

Thống kê F trong bước 3 ở trên có thể tính theo công thức sau

$$F = \frac{\frac{R^2(U) - R^2(R)}{m}}{\frac{1 - R^2(U)}{n - k}}$$

Với VD 3.3 ta có  $F_{qs} = \frac{(0.830125 - 0.824555) / 2}{(1 - 0.830125) / (30 - 4)} = 0.43$

# CHƯƠNG III

## 3.3.5. Kiểm định về sự phù hợp của hàm hồi quy

Hàm hồi quy là không phù hợp nếu toàn bộ các biến độc lập trong mô hình đều không ảnh hưởng đến biến phụ thuộc.

Cặp giả thuyết

$$H_0 : \beta_2 = \beta_3 = \dots = \beta_k = 0$$

$$H_1 : \beta_2^2 + \beta_3^2 + \dots + \beta_k^2 \neq 0$$

Mô hình không có ràng buộc:

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + u, \quad \text{thu được } R^2$$

Mô hình có ràng buộc

$$Y = \beta_1 + u \quad \text{thu được } R^2 = 0$$

# CHƯƠNG III

Số ràng buộc là  $m = k - 1$

Thống kê kiểm định:

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

Nếu  $F_{qs} > f_{\alpha}(k-1, n-k)$  thì bác bỏ  $H_0$

Trong báo cáo Eviews  $F_{qs}$  và Prob. tương ứng được tính sẵn

## Ví dụ 3.3.

F – statistic = 21410.05

Prob. (F – statistic ) = 0.000000

Kết luận: bác bỏ  $H_0$ , mô hình là phù hợp.



# CHƯƠNG III

## 3.3.6. So sánh kiểm định T và kiểm định F

- Nếu kiểm định giả thuyết về 1 ràng buộc: sử dụng kiểm định T – có thể áp dụng cho ràng buộc dạng đẳng thức hoặc bất đẳng thức.
- Nếu kiểm định giả thuyết có 2 ràng buộc trở lên: sử dụng kiểm định F – chỉ sử dụng được cho các ràng buộc dạng đẳng thức.

## 3.4. MỘT SỐ KIỂM ĐỊNH KHÁC

- Kiểm định Wald
- Kiểm định tỷ số hợp lý LR
- Kiểm định nhân tử Lagrange (LM)

Các KĐ này dùng được cho cả mô hình tuyến tính và phi tuyến và chỉ phù hợp cho trường hợp mẫu lớn

# CHƯƠNG III

## 3.5. DỰ BÁO GIÁ TRỊ CỦA BIẾN PHỤ THUỘC VÀ SAI SỐ DỰ BÁO

### 3.5.1. Dự báo giá trị của biến phụ thuộc

Xét mô hình 2 biến:  $Y = \beta_1 + \beta_2 X + u$

Khoảng tin cậy với độ tin cậy  $(1 - \alpha)$  cho giá trị trung bình của biến phụ thuộc khi  $X = X_i$  là

$$\hat{Y}_i - t_{\alpha/2}^{(n-2)} se(\hat{Y}_i) < E(Y | X_i) < \hat{Y}_i + t_{\alpha/2}^{(n-2)} se(\hat{Y}_i)$$

$$se(\hat{Y}_i) = \hat{\sigma} \left( \frac{1}{n} + \frac{x_i^2}{\sum_{j=1}^n x_j^2} \right)^{1/2}$$

# CHƯƠNG III

Với mô hình k biến, ta sử dụng dạng ma trận như sau

$$\mathbf{Y} = \mathbf{X} \cdot \boldsymbol{\beta} + \mathbf{u}$$

Ước lượng điểm tại  $\mathbf{X}_i = (1, X_{2i}, \dots, X_{ki})$  là

$$\widehat{Y}_i = \mathbf{X}_i \cdot \widehat{\boldsymbol{\beta}}$$

Ước lượng khoảng

$$\widehat{Y}_i - t_{\alpha/2}^{(n-k)} \text{se}(\widehat{Y}_i) < E(Y|\mathbf{X}_i) < \widehat{Y}_i + t_{\alpha/2}^{(n-k)} \text{se}(\widehat{Y}_i)$$

$$\text{se}(\widehat{Y}_i) = \widehat{\sigma} \left( \mathbf{X}_i' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_i \right)^{1/2}$$

# CHƯƠNG III

## 3.5.2. Đánh giá sai số dự báo

Căn bậc hai của trung bình phương sai sai số

$$RMSE = \sqrt{\frac{\sum (\hat{Y}_i - Y_i)^2}{n}}$$

Sai số trung bình tuyệt đối

$$MAE = \frac{\sum |\hat{Y}_i - Y_i|}{n}$$

Sai số trung bình tuyệt đối tính theo phần trăm

$$MAPE = \frac{\sum \left| \frac{\hat{Y}_i - Y_i}{Y_i} \right|}{n}$$

# CHƯƠNG IV

## PHÂN TÍCH HỒI QUY VỚI BIẾN ĐỊNH TÍNH

### 4.1. KHÁI NIỆM BIẾN GIẢ (dummy variable)

Biến giả là biến định tính chỉ có 2 phạm trù được gán là 0 và 1. Hai số 0 và 1 chỉ phản ánh 2 nhóm tính chất khác nhau, nó không dùng để so sánh.

Ví dụ

$N\grave{u} = 1$  nếu một người là nữ

$= 0$  nếu một người không phải là nữ

$B\grave{a}c = 1$  nếu hộ gia đình ở miền Bắc

$= 0$  nếu hộ gia đình không ở miền Bắc

# CHƯƠNG IV

## 4.2. MÔ HÌNH CÓ CHỨA BIẾN ĐỘC LẬP LÀ BIẾN GIẢ

Giả sử có biến định tính  $Z$  có 2 phạm trù và có tác động đến biến  $Y$ . Gọi  $D$  là biến giả thể hiện biến  $Z$ .

Xét bài toán hồi quy bội có chứa biến giả

$$Y = \beta_1 + \beta_2 D + \beta_3 X_3 + \dots + \beta_k X_k + u$$

- Các hệ số  $\beta_1, \beta_3, \dots, \beta_k$  có ý nghĩa như đã xét ở chương II
- Hệ số  $\beta_2$  thể hiện sự khác biệt giữa giá trị trung bình của  $Y$  trong nhóm các quan sát có  $D = 1$  với giá trị này trong nhóm các quan sát có  $D = 0$ , khi các biến  $X_j$  là như nhau.

# CHƯƠNG IV

- Ta có

$$E(Y | D = 0; X_3, \dots, X_k) = \beta_1 + \beta_3 X_3 + \dots + \beta_k X_k$$

$$E(Y | D = 1; X_3, \dots, X_k) = \beta_1 + \beta_2 + \beta_3 X_3 + \dots + \beta_k X_k$$

- Để kiểm tra xem  $D$  có tác động đến  $Y$  hay không ta kiểm định cặp giả thuyết

$$H_0 : \beta_2 = 0; \quad H_1 : \beta_2 \neq 0$$

**Ví dụ 4.1.** Với tập số liệu chitieu.wfl có  $n = 30$  quan sát

CT: chi tiêu của hộ gia đình (triệu đồng/năm)

TN: thu nhập của hộ gia đình (triệu đồng/năm)

TP = 1 nếu hộ gia đình ở thành phố

= 0 nếu ở nông thôn

# CHƯƠNG IV

- Hồi quy mô hình ta có

$$CT = 51.277 + 25.979*TP + 0.399*TN + e$$

$$(se) (12.717) \quad (14.713) \quad (0.039)$$

$$(Prob) (0.0004) \quad (0.0887) \quad (0.000)$$

$$n = 30; R^2 = 0.843 \text{ (Prob} = 0.000)$$

$$\text{Ở nông thôn: } CT = 51.277 + 0.399*TN + e$$

$$\begin{aligned} \text{Ở thành phố: } CT &= 51.277 + 25.979 + 0.399*TN + e \\ &= 77.256 + 0.399*TN + e \end{aligned}$$

Số 25.979 cho biết sự khác biệt của 2 nhóm trên.

- Khi thu nhập bằng 0 thì chi tiêu trung bình của hộ gia đình ở nông thôn ( $TP = 0$ ) là 51.277 triệu đồng/năm.



# CHƯƠNG IV

- Khi thu nhập bằng 0 thì chi tiêu trung bình của hộ gia đình ở thành phố ( $TP = 1$ ) là 77.256 triệu đồng/năm. Mức này cao hơn so với hộ gia đình cùng cùng mức thu nhập nhưng ở nông thôn là 25.979 triệu đồng/năm.
- Khi thu nhập tăng thêm 1 triệu đồng/năm thì chi tiêu trung bình của hộ gia đình ở thành phố hay nông thôn đều tăng lên 0.399 triệu đồng/năm.
- Dựa vào  $R^2$  ta thấy TN và TP giải thích được 84.3% sự thay đổi của CT. Kiểm định  $R^2 = 0$  có  $P\_value \approx 0.000$  nên mô hình là phù hợp.
- Trong mô hình thì 51.277 và 0.399 có ý nghĩa thống kê, số 25.979 không có ý nghĩa thống kê.

# CHƯƠNG IV

## 4.3. MÔ HÌNH VỚI BIẾN GIẢ VÀ BIẾN TƯƠNG TÁC

Trong ví dụ 4.1. ta thấy hai hàm hồi quy trong 2 trường hợp của biến giả chỉ khác nhau hệ số chặn. Trong nhiều trường hợp hai hàm hồi quy còn khác nhau cả hệ số góc. Khi đó cần đưa thêm biến tương tác của biến độc lập và biến giả vào mô hình.

**Ví dụ 4.2.** Với tập số liệu chitieu.wfl, đưa thêm biến  $TN*TP$  vào mô hình ta có

$$CT = 68.851 - 2.384*TP + 0.305*TN + 0.126(TN*TP) + e$$

(17.607) (24.699) (0.077) (0.089)

$$R^2 = 0.854$$

# CHƯƠNG IV

- Ở nông thôn:  $CT = 68.851 + 0.305 * TN + e$
- Ở thành phố:  $CT = 68.851 - 2.384 + 0.305 * TN - 0.126 * TN + e$   
 $= 66.467 + 0.431 * TN + e$

- Khi  $TN = 0$  thì chi tiêu trung bình 1 năm của hộ gia đình ở nông thôn là 68.851 triệu, ở thành phố là 66.467 triệu.
- Khi  $TN$  tăng thêm 1 triệu/năm thì  $CT$  trung bình của hộ gia đình ở nông thôn tăng thêm 0.305 triệu, còn hộ gia đình ở thành phố tăng thêm 0,431 triệu.
- Các biến độc lập trong mô hình giải thích được 85.4% sự thay đổi của  $CT$ .

# CHƯƠNG IV

## Kiểm định sự khác biệt giữa hàm hồi quy của hai nhóm

Dùng kiểm định Chow hoặc dùng biến giả.

### *Kiểm định Chow*

Giả sử các quan sát ứng với nhóm 1 được sắp xếp từ 1 đến  $n_1$ , và với nhóm 2 từ  $(n_1 + 1)$  đến  $n$ .

Với nhóm 1 ta có  $n_1$  phương trình

$$Y_i = \beta_1 + \beta_2 X_i + u_1 \quad (i = 1, \dots, n_1) \quad (1)$$

Với nhóm 2 có  $(n - n_1)$  phương trình

$$Y_i = \alpha_1 + \alpha_2 X_i + u_2 \quad (i = n_1 + 1, \dots, n) \quad (2)$$

Cần kiểm định cặp giả thuyết

$$H_0 : \beta_1 = \alpha_1; \beta_2 = \alpha_2 ; \quad H_1 : \text{không phải } H_0$$

# CHƯƠNG IV

Mô hình *không có ràng buộc* là  $\{(1);(2)\}$ , ta tìm được

$$RSS(U) = RSS_1 + RSS_2$$

Mô hình *có ràng buộc* là

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (i = 1, \dots, n) \quad (R)$$

Hồi quy (R) ta tìm được  $RSS(R) = RSS$

Dùng kiểm định F ta có thống kê kiểm định là

$$F = \frac{[RSS - (RSS_1 + RSS_2)] / k}{[RSS_1 + RSS_2] / (n - 2k)}$$

Nếu  $F_{qs} > f_\alpha(k, n-2k)$  thì bác bỏ  $H_0$

Với mô hình nhiều biến thì làm tương tự như trên

# CHƯƠNG IV

## *Kiểm định sử dụng biến giả*

**Ví dụ 4.3.** Sử dụng ví dụ 4.2 ta có

$$CT = \beta_1 + \beta_2 TP + \beta_3 TN + \beta_4 TN * TP + u \quad (U)$$

Nếu không có sự khác biệt về chi tiêu của hộ gia đình ở thành phố và ở nông thôn thì  $\beta_2 = \beta_4 = 0$

$$H_0 : \beta_2 = \beta_4 = 0; \quad H_1 : \beta_2^2 + \beta_4^2 > 0$$

Hồi quy mô hình (U) ta có  $R^2(U) = 0.854$

Mô hình có ràng buộc là

$$CT = \beta_1 + \beta_3 TN + u \quad (R)$$

Có  $R^2(R) = 0.825$

$F_{qs} = 2.582 < f_{0.05}(2, 26) = 3.49$  nên chưa có cơ sở bác bỏ  $H_0$

# CHƯƠNG IV

## 4.4. TRƯỜNG HỢP BIẾN ĐỊNH TÍNH CÓ NHIỀU PHẠM TRÙ

Khi xem xét biến định tính có nhiều hơn 2 phạm trù, ví dụ như nơi cư trú (miền Bắc, Trung, Nam), loại hình doanh nghiệp phân theo sở hữu (DN nhà nước, DN tư nhân, DN có vốn đầu tư nước ngoài),... thì không thể chỉ dùng 1 biến giả để mô hình hóa các biến trên.

Nếu biến định tính có  $m$  phạm trù thì phải sử dụng  $(m-1)$  biến giả.

Phạm trù cơ sở là phạm trù có tất cả các biến giả (ứng với biến định tính đang xét) nhận giá trị bằng 0.

# CHƯƠNG IV

**Ví dụ 4.4.** Với tập số liệu chitieu.wfl xét biến CT, TN và

BAC = 1 nếu hộ gia đình ở miền Bắc

= 0 nếu ngược lại

NAM = 1 nếu hộ gia đình ở miền Nam

= 0 nếu ngược lại

$$CT = 46.93 + 0.43*TN + 12.4*BAC + 18.84*NAM + e$$

Hộ gia đình ở miền trung (phạm trừ cơ sở)

$$CT = 46.93 + 0.43*TN + e$$

Hộ gia đình ở miền Bắc

$$CT = 59.33 + 0.43*TN + e$$

Hộ gia đình ở miền Nam

$$CT = 65.77 + 0.43*TN + e$$



# CHƯƠNG IV

## *Bẫy biến giả*

Giả sử trong ví dụ 4.4 ta đưa thêm vào biến giả

$TRUNG = 1$  nếu hộ gia đình ở miền Trung  
 $= 0$  nếu ngược lại

thì luôn có  $BAC + TRUNG + NAM = 1$ . Khi hồi quy mô hình có hệ số chặn thì ta gặp phải vấn đề đa công tuyến hoàn hảo.

- Chỉ sử dụng  $(m - 1)$  biến giả
- Nếu muốn sử dụng cả  $m$  biến giả thì cần sử dụng mô hình không có hệ số chặn.

# CHƯƠNG IV

*Khi biến định tính là biến thứ bậc* như trình độ học vấn (tiểu học trở xuống  $HV = 0$ , trung học  $HV = 1$ , cao đẳng – đại học  $HV = 2$ , sau đại học  $HV = 3$ ) hay mức độ hài lòng về một loại dịch vụ,...

Nếu đưa trực tiếp biến này vào mô hình dưới dạng biến số thì kết quả thu được thiếu chính xác. Khi đó ta dùng biến giả để giải quyết. Ví dụ biến  $HV$  ta có

$HV_1 = 1$  nếu  $HV = 1$  và bằng 0 cho trường hợp ngược lại

$HV_2 = 1$  nếu  $HV = 2$  và bằng 0 cho trường hợp ngược lại

$HV_3 = 1$  nếu  $HV = 3$  và bằng 0 cho trường hợp ngược lại

Khi  $HV_1 = HV_2 = HV_3 = 0$  thì trình độ học vấn là tiểu học trở xuống (phạm trù cơ sở).

# CHƯƠNG V

## KIỂM ĐỊNH VÀ LỰA CHỌN MÔ HÌNH

- Khi phân tích hồi quy bằng phương pháp OLS thì giả thiết 1 mặc nhiên công nhận.
- Nếu mô hình vi phạm 1 trong các giả thiết còn lại thì ta nói mô hình có khuyết tật. Một số loại khuyết tật:
  - + Dạng hàm sai, thiếu biến quan trọng, thừa biến
  - + Phương sai sai số thay đổi
  - + Có đa cộng tuyến hoàn hảo hoặc đa cộng tuyến cao

....

# CHƯƠNG V

## 5.1. KỲ VỌNG CỦA SAI SỐ NGẪU NHIÊN KHÁC 0.

Xét mô hình

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + u$$

Giả thiết 2:  $E(u | X_2, \dots, X_k) = 0$

Nếu giả thiết này thỏa mãn thì

$$(1) E(u) = 0$$

$$(2) \text{cov}(X_j, u) = 0 \text{ với mọi } j = 2, \dots, k$$

Nếu (1) hoặc (2) không thỏa mãn thì giả thiết 2 không thỏa mãn.

# CHƯƠNG V

## 5.1.1. Nguyên nhân

*a. Nguyên nhân 1: Mô hình thiếu biến quan trọng*

Giả sử  $Z$  là một biến không có trong mô hình. Mô hình được gọi là thiếu biến quan trọng  $Z$  nếu

- (i)  $Z$  có tác động đến biến phụ thuộc  $Y$
- (ii)  $Z$  có tương quan với ít nhất 1 trong các biến độc lập của mô hình.

*b. Nguyên nhân 2: **Dạng hàm sai***

*c. Nguyên nhân 3: Tính tác động đồng thời của số liệu*

*d. Nguyên nhân 4: Sai số đo lường của các biến độc lập.*

# CHƯƠNG V

## 5.1.2. Hậu quả của kỳ vọng SSNN khác 0

### *a. Ước lượng OLS sẽ là ước lượng chệch*

Lượng chệch của ước lượng  $\hat{\beta}_j$  định nghĩa bởi

$$E(\hat{\beta}_j) - \beta_j$$

### *b. Các suy diễn thống kê không còn đáng tin cậy*

# CHƯƠNG V

## 5.1.3. Phát hiện sự khác 0 của kỳ vọng SSNN

### *a. Mô hình bỏ sót biến quan trọng*

- + Nếu muốn kiểm định mô hình có thiếu biến  $Z$  hay không thì đưa thêm biến  $Z$  vào mô hình rồi sử dụng kiểm định  $T$  để kiểm tra xem hệ số của nó có bằng 0 hay không
- + Muốn kiểm định mô hình thiếu các biến  $Z_1, Z_2, \dots, Z_m$  thì đưa thêm các biến này vào mô hình rồi sử dụng kiểm định  $F$  về sự đồng thời bằng 0 của các hệ số của các biến đó

# CHƯƠNG V

## Ví dụ 5.1. Xét mô hình

- $CT = \beta_1 + \beta_2 TN + \beta_3 TS + u$ ; có  $R^2(1) = 0,829$
- Khi đưa thêm biến SONGUOI và biến TP vào thì ta được mô hình mới có  $R^2(2) = 0,9$ . Có nên thêm 2 biến này không?
- Thêm biến ta có mô hình

$$CT = \alpha_1 + \alpha_2 TN + \alpha_3 TS + \alpha_4 SONGUOI + \alpha_5 TP + v$$

- Kiểm định:  $H_0: \alpha_4 = \alpha_5 = 0$ ;  $H_1: \alpha_4^2 + \alpha_5^2 \neq 0$

$$F_{qs} = \frac{(R^2(2) - R^2(1)) / m}{(1 - R^2(2)) / (n - k)} = \frac{(0,9 - 0,829) / 2}{(1 - 0,9) / 25} = 8,875$$

Do  $F_{qs} > f_{0,05}(2,25) = 3,49$  nên bác bỏ  $H_0$ , nghĩa là nên thêm 2 biến này vào mô hình.



# CHƯƠNG V

## *b. Mô hình có dạng hàm sai*

*Kiểm định chung về dạng hàm sai: **kiểm định Ramsey***

Dùng để phát hiện mô hình thiếu biến là hàm của các biến trong mô hình

Tư tưởng: nếu mô hình ban đầu có dạng hàm phù hợp thì khi đưa thêm các biến  $X_j^2, X_j^3, \dots, X_j X_s, \dots$  vào mô hình thì hệ số của các biến này sẽ không có ý nghĩa thống kê.

Việc đưa thêm biến như trên vào mô hình tiêu tốn nhiều bậc tự do nên Ramsey sử dụng các biến  $\hat{Y}^m$  để thay thế.

# CHƯƠNG V

Các bước thực hiện:

- Bước 1: Ước lượng mô hình hồi quy ban đầu thu được  $\hat{Y}$
- Bước 2: Ước lượng mô hình mới

$$Y = \alpha_1 + \alpha_2 X_2 + \dots + \alpha_k X_k + \alpha_{k+1} \hat{Y} + \alpha_{k+2} \hat{Y}^2 + v$$

- Bước 3: Kiểm định cặp giả thuyết

$$H_0 : \alpha_{k+1} = \alpha_{k+2} = 0; \quad H_1 : \alpha_{k+1}^2 + \alpha_{k+2}^2 > 0$$

(dùng kiểm định F)

# CHƯƠNG V

## **Ví dụ 5.2.** Xét mô hình

# CHƯƠNG V

*Một số kiểm định khác*

***Kiểm định Davidson – Mac Kinnon (kiểm định J)***

Dùng để lựa chọn giữa các hàm hồi quy không bao nhau

$$H_0 : Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + u; \quad u \sim N(0; \sigma_u^2) \quad (1)$$

$$H_1 : Y = \alpha_1 + \alpha_2 Z_2 + \dots + \alpha_k Z_k + v; \quad v \sim N(0; \sigma_v^2) \quad (2)$$

Ý tưởng: Nếu dạng hàm (1) đúng thì khi đưa thêm  $\hat{Y}'$  của hàm (2) vào thì hệ số của biến mới này không có ý nghĩa thống kê. Ngược lại, nếu dạng hàm (2) đúng thì khi đưa thêm biến  $\hat{Y}$  của hàm (1) vào thì hệ số của nó không có ý nghĩa thống kê.

# CHƯƠNG V

## Thực hiện

- Bước 1: Ước lượng (1) và (2) bằng OLS thu được các ước lượng của Y, kí hiệu là  $\hat{Y}, \hat{Y}'$ .

- Bước 2: Ước lượng các mô hình sau  
$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + \beta_{k+1} Y' + u$$

$$Y = \alpha_1 + \alpha_2 X_2 + \dots + \alpha_k X_k + \alpha_{k+1} \hat{Y} + v$$

- Bước 3: Thực hiện các kiểm định T để KĐ các cặp giả thuyết

$$H_{01} : \beta_{k+1} = 0; \quad H_{11} : \beta_{k+1} \neq 0;$$

$$H_{02} : \alpha_{k+1} = 0; \quad H_{12} : \alpha_{k+1} \neq 0;$$

# CHƯƠNG V

$\text{Giả thiết } \alpha_{k+1} = 0$ $\text{Giả thiết } \beta_{k+1} = 0$	Không bác bỏ	Bác bỏ
Không bác bỏ	(1) và (2) đều có thể chấp nhận	(1) không thể, (2) có thể chấp nhận
Bác bỏ	(1) có thể chấp nhận, (2) không thể	(1) và (2) đều không thể chấp nhận

## *Kiểm định sử dụng hàm gộp*

Đưa tất cả các biến  $X_2, \dots, X_k, Z_2, \dots, Z_k$  vào 1 mô hình rồi kiểm định sự đồng thời bằng 0 của các hệ số của  $X$ , sự đồng thời bằng 0 của các hệ số của  $Z$ . Việc ra quyết định lựa chọn tương tự như kiểm định J.

# CHƯƠNG V

## 5.1.4. Một số biện pháp khắc phục

- Nếu nguyên nhân là thiếu biến  $Z$  (có số liệu) thì đưa thêm biến  $Z$  vào mô hình
- Nếu nguyên nhân là dạng hàm sai, được phát hiện từ kiểm định Ramsey, thì xem xét các mô hình thay thế như mô hình dạng logarit, đưa thêm biến  $X^2$  vào,...
- Nếu nguyên nhân là thiếu biến (không quan sát được) thì  
+ Dùng biến đại diện: biến  $Z$  là biến đại diện cho  $Z^*$  nếu nó là quan sát được và có tương quan với biến  $Z^*$

VD có thể dùng chỉ số IQ để đại diện cho năng lực bẩm sinh

+ Sử dụng phương pháp biến công cụ.

# CHƯƠNG V

## 5.2. PHƯƠNG SAI SAI SỐ THAY ĐỔI

Giả thiết 3:  $\text{var}(u|X_{2i}, \dots, X_{ki}) = \sigma^2$

Nếu giả thiết trên bị vi phạm thì mô hình có PSSS thay đổi

### 5.2.1. Nguyên nhân của phương sai sai số thay đổi

- + Do bản chất của số liệu
- + Do mô hình thiếu biến quan trọng hoặc dạng hàm sai

### 5.2.2. Hậu quả của phương sai sai số thay đổi

- + Các ước lượng OLS vẫn là ước lượng không chệch
- + Phương sai của các hệ số ước lượng là chệch
- + Khoảng tin cậy và kiểm định giả thuyết các hệ số không còn giá trị sử dụng
- + Các ước lượng hệ số không còn là ước lượng tốt nhất



# CHƯƠNG V

## 5.2.3. Phát hiện phương sai sai số thay đổi

Giả sử mô hình không vi phạm giả thiết 2 thì  $E(u|X) = 0$  nên

$$\text{Var}(u | X_{2i}, \dots, X_{ki}) = e(u^2 | X_{2i}, \dots, X_{ki})$$

Nếu mô hình thỏa mãn giả thiết 3 thì

$$e(u^2 | X_{2i}, \dots, X_{ki}) = \sigma^2 \quad (1)$$

Khi (1) bị vi phạm thì mô hình có PSSS thay đổi.

### *a. Sử dụng đồ thị phần dư*

Trong công thức (1) giá trị của  $u^2$  chưa biết nên phải thay thế bằng ước lượng của nó là  $e_i^2$ . Nhìn vào đồ thị của  $e_i$  hoặc  $e_i^2$  trên đồ thị (vẽ theo 1 biến  $X_j$  nào đó) có thể thấy được sự hiện diện của PSSS thay đổi.

# CHƯƠNG V

## *b. Kiểm định Breusch – Pagan (BP)*

Ý tưởng: Nếu  $u^2$  có tương quan với ít nhất một trong các biến  $X_j$  thì (1) không thỏa mãn. Do đó ta xét mô hình hồi quy phụ sau

$$u_i^2 = a_1 + a_2 X_{2i} + \dots + a_k X_{ki} + v_i$$

Do  $u_i^2$  không quan sát được nên thay thế nó bởi  $e_i^2$  ta có mô hình sau:

$$e_i^2 = b_1 + b_2 X_{2i} + \dots + b_k X_{ki} + w_i \quad (*)$$

Cần kiểm định cặp giả thuyết

$$H_0 : b_2 = \dots = b_k = 0; \quad H_1 : b_2^2 + \dots + b_k^2 > 0$$

Ta có thể sử dụng kiểm định F hoặc dùng thống kê

$$LM = nR_e^2 \sim \chi^2(k)$$

# CHƯƠNG V

- Bước 1: Ước lượng mô hình hồi quy gốc thu được các phần dư  $e_i$ .
- Bước 2: Ước lượng mô hình (\*) thu được  $R_e^2$
- Bước 3: Tính giá trị quan sát của các thống kê kiểm định

$$F_{qs} = \frac{R_e^2 / (k - 1)}{(1 - R_e^2) / (n - k)}; \quad LM_{qs} = nR_e^2$$

Nếu  $F_{qs} > f_{\alpha}(k-1, n-k)$  hoặc  $LM_{qs} > \chi^2_{\alpha}(k)$  thì bác bỏ  $H_0$  và kết luận mô hình có PSSS thay đổi.

# CHƯƠNG V

## **Ví dụ 5.3.** Xét mô hình

# CHƯƠNG V

## *c. Kiểm định White (thường dùng)*

Xét mô hình

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

**Giả thiết 3'**:  $u_i^2$  không tương quan với các biến độc lập, bình phương của các biến độc lập và tích chéo của các biến độc lập.

Nếu giả thiết 3' bị vi phạm thì mô hình có khuyết tật PSSS thay đổi.

Do  $u_i^2$  không quan sát được nên khi thực hiện kiểm định White thì dùng  $e_i^2$  để thay thế.

Các bước kiểm định như sau:

# CHƯƠNG V

- Bước 1: Ước lượng mô hình ban đầu thu được  $e_i^2$
- Bước 2: Ước lượng mô hình hồi quy phụ

$$e_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + v_i$$

thu được  $R_e^2$

- Bước 3: Xét cặp giả thuyết

$$H_0 : \alpha_2 = \dots = \alpha_6 = 0 \quad (\Leftrightarrow R_e^2 = 0)$$

$$H_1 : \alpha_2^2 + \dots + \alpha_6^2 > 0 \quad (R_e^2 > 0)$$

Sử dụng kiểm định F để kiểm định cặp giả thuyết này.

- Có thể sử dụng thống kê sau để kiểm định:

$$\chi^2 = nR_e^2 \sim \chi^2(k_e - 1)$$

$k_e$  là số hệ số của hồi quy phụ. Bác bỏ  $H_0$  nếu  $\chi_{qs}^2 > \chi_\alpha^2(k_e - 1)$

# CHƯƠNG V

## **Ví dụ 5.4.** Xét mô hình

# CHƯƠNG V

## *c. Một số kiểm định khác*

- ❖ Kiểm định Park (1986): nếu đồ thị phần dư gợi ý PSSS là hàm của biến  $X_j$  theo dạng

$$\sigma_i^2 = \sigma^2 X_{ji}^{\alpha_j}$$

thì ta cần kiểm định  $\alpha_j \neq 0$ .

Lấy loga tự nhiên 2 vế rồi thay  $\sigma_i^2$  bởi  $e_i^2$  rồi thực hiện KĐ:

- B1: ƯL hàm hồi quy gốc để thu được  $e_i^2$
  - B2: ƯL mô hình HQ phụ:  $\ln e_i^2 = \alpha_1 + \alpha_j \ln X_{ji} + v_i$
  - B3: Kiểm định:  $H_0: \alpha_j = 0$ ;  $H_1: \alpha_j \neq 0$   
bằng kiểm định t hoặc kiểm định F.
- ❖ Kiểm định Gleizer (1969)



# CHƯƠNG V

## 5.2.4. Khắc phục phương sai sai số thay đổi

PSSS thay đổi có thể do thiếu biến hoặc dạng hàm sai, do đó trước khi khắc phục hậu quả do PSSS thay đổi gây ra ta cần khắc phục các khuyết tật này trước.

2 cách giải quyết vấn đề PSSS thay đổi:

*a. Phương pháp bình phương bé nhất tổng quát (GLS – generalized least squares)*

Giả sử PSSS là thay đổi theo dạng:  $\sigma_i^2 = \sigma^2 X_{2i}^2$

thì thực hiện chia 2 vế của mô hình gốc cho  $X_{2j}$  sẽ thu được mô hình mới có PSSS bằng  $\sigma^2$

Phương pháp này còn gọi là PP bình phương bé nhất có trọng số (WLS = weighted least squares).

# CHƯƠNG V

## *b. Ước lượng sai số chuẩn*

White (1980) đề xuất phương pháp *sai số chuẩn vững* (robust standard error) với tư tưởng: vẫn sử dụng các hệ số UL từ phương pháp OLS nhưng phương sai của các hệ số được tính toán lại mà không sử dụng giả thiết PSSS không đổi.

Khi các giả thiết khác  
thỏa mãn thì

$$\text{var}(\hat{\beta}_2) = \frac{\sum_{i=1}^n x_{2i}^2 \sigma_i^2}{\left( \sum_{i=1}^n x_{2i}^2 \right)^2} \quad (*)$$

Công thức White đề xuất

$$\text{var}(\hat{\beta}_2) = \frac{\sum_{i=1}^n x_{2i}^2 e_i^2}{\left( \sum_{i=1}^n x_{2i}^2 \right)^2} \quad (**)$$

sẽ hội tụ về (\*) khi n đủ lớn

Sai số chuẩn vững là căn bậc 2 của (\*\*).

# CHƯƠNG V

## 5.3. SAI SỐ NGẪU NHIÊN KHÔNG TUÂN THEO QUY LUẬT CHUẨN

Giả thiết 5: Sai số ngẫu nhiên u tuân theo quy luật Chuẩn.

Kết hợp với các giả thiết khác thì có  $u_i \sim N(0; \sigma^2)$ .

Giả thiết này tương đương với tại mỗi quan sát thì

$$Y |_{X_{2i}, \dots, X_{ki}} \sim N(\beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki}, \sigma^2)$$

Trong nhiều trường hợp điều này không thỏa mãn.

### 5.3.1. Hậu quả của SSNN không tuân theo quy luật Chuẩn

- + Các ƯL OLS vẫn tốt nhất
- + Các thống kê t và F không tuân theo quy luật Student và Fisher
- + Khi n nhỏ thì các suy diễn thống kê không đáng tin cậy
- + Khi n lớn thì suy diễn thống kê vẫn có giá trị.

# CHƯƠNG V

## 5.3.2. Phát hiện SSNN không tuân theo quy luật Chuẩn

### a. Xem xét đồ thị phần dư

Xem xét đồ thị tần suất (histogram plot) của phần dư.

### b. Kiểm định Jarque – Bera (JB) (1987)

$H_0$ : u tuân theo quy luật chuẩn

$H_1$ : u **không** tuân theo quy luật chuẩn

B1: ƯL mô hình gốc thu được phần dư  $e_i$

B2: Tính giá trị qua sát của thống kê kiểm định

$$JB = n \left( \frac{S^2}{6} + \frac{(K-3)^2}{24} \right) \quad \text{với } S \text{ là hệ số bất đối xứng (Skewness)}$$

$K$  là hệ số nhọn (Kurtosis)

Nếu  $H_0$  đúng thì  $JB \sim \chi^2(2)$

B3: Nếu  $JB_{qs} > \chi_{\alpha}^{2(2)}$  thì bác bỏ  $H_0$ , nghĩa là u không pp chuẩn<sub>116</sub>

# CHƯƠNG V

## **Ví dụ 5.5.** Xét mô hình

# CHƯƠNG V

## 5.4. VẤN ĐỀ ĐA CỘNG TUYẾN

Giả thiết 4: Giữa các biến độc lập không có đa cộng tuyến hoàn hảo.

Thực tế thường không gặp đa cộng tuyến hoàn hảo nhưng hay gặp vấn đề gần như đa cộng tuyến hoàn hảo – đa cộng tuyến cao.

### 5.4.1. Khái niệm đa cộng tuyến trong mô hình hồi quy

- Nếu giữa các biến số  $X_2, \dots, X_k$  có quan hệ đa cộng tuyến hoàn hảo thì sẽ tồn tại ít nhất một biến  $X_j$  trong các biến này sao cho kho hồi quy nó theo các biến còn lại:

$$X_j = \alpha_1 + \alpha_2 X_2 + \dots + \alpha_{j-1} X_{j-1} + \alpha_{j+1} X_{j+1} + \alpha_k X_k + v$$

thì sẽ có hệ số xác định  $R^2 = 1$

- Khi  $R^2 \approx 1$  thì các biến  $X_2, \dots, X_k$  có quan hệ đa cộng tuyến cao.

# CHƯƠNG V

## 5.4.2. Nguyên nhân và hậu quả của đa cộng tuyến cao

### *a. Nguyên nhân*

- + Do bản chất mối quan hệ giữa các biến số
- + Mô hình có dạng đa thức, các biến  $X$ ,  $X^2$ ,  $X^3$  thường có quan hệ tuyến tính khá chặt.
- + Mẫu không mang tính đại diện.

### *b. Hậu quả*

Đa cộng tuyến cao không vi phạm giả thiết nào của định lý Gauss – Markov nên nó không ảnh hưởng đến tính ‘tốt nhất’ của các ước lượng OLS.

# CHƯƠNG V

Hiện tượng đa cộng tuyến cao ảnh hưởng đến phương sai của các hệ số ước lượng như sau:

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

Khi  $R_j^2$  lớn sẽ có thể làm cho phương sai lớn, gây ra:

- Khoảng tin cậy cho  $\beta_j$  trở nên rộng, nghĩa là các ước lượng trở nên kém chính xác.
- Hệ số ước lượng dễ mất ý nghĩa thống kê (do  $t_{qs}$  trong kiểm định  $t$  trở nên quá bé).
- Dấu của hệ số ước lượng của biến  $X_j$  có thể ngược với kì vọng.
- Một sự thay đổi nhỏ trong mẫu cũng có thể gây ra một sự thay đổi khá lớn trong kết quả ước lượng.

Các hậu quả trên cũng có thể là hậu quả của mô hình thiếu biến hoặc dạng hàm sai.



# CHƯƠNG V

## 5.4.3. Phát hiện đa cộng tuyến cao

- a. Xem xét hệ số xác định của các mô hình hồi quy phụ  $R_j^2$ . Nếu  $R_j^2$  là quá lớn thì có thể là hậu quả của hiện tượng đa cộng tuyến cao.
- b. Xem xét hệ số phóng đại phương sai (VIF)  
Nếu  $VIF > 10$  thì đó là dấu hiệu của đa cộng tuyến cao.

$$VIF_j = \frac{1}{1 - R_j^2}$$

- c. Tính hệ số tương quan cặp của các biến  $X_j$ . Nếu hệ số tương quan cặp giữa 2 biến nào đó lớn hơn 0,8 thì có thể xem như mô hình có đa cộng tuyến cao.

# CHƯƠNG V

## 5.4.4. Một số biện pháp khắc phục

a. Nếu đa cộng tuyến cao nhưng không gây hậu quả nghiêm trọng thì không cần biện pháp khắc phục.

Cụ thể là:

- + Tồn tại đa cộng tuyến cao trong mô hình nhưng các  $se(\hat{\beta}_j)$  là không quá lớn so với  $\hat{\beta}_j$  thì không cần khắc phục.
- + Đa cộng tuyến cao nhưng  $se(\hat{\beta}_j)$  là không lớn, trong đó  $X_j$  là biến mà ta đang đo lường tác động đến biến  $Y$ .

# CHƯƠNG V

**b.** Khi đa cộng tuyến gây hậu quả nghiêm trọng, với số liệu chéo thì khắc phục như sau:

+ Gia tăng kích thước mẫu

+ Thay đổi dạng hàm ước lượng

+ Bỏ bớt biến là nguyên nhân gây ra đa cộng tuyến cao

+ Sử dụng kỹ thuật phân tích nhân tố

# CHƯƠNG V

## 5.5. MÔ HÌNH CHỨA BIẾN KHÔNG THÍCH HỢP

(irrelevant variable)

### 5.5.1. Hậu quả của việc chứa biến không thích hợp

Khi đưa thêm biến không thích hợp vào mô hình thì các ước lượng của mô hình thừa biến vẫn là ước lượng không chệch nhưng phương sai của các hệ số ước lượng sẽ lớn làm cho

- + Khoảng tin cậy của các hệ số hồi quy sẽ rộng hơn
- + Các giá trị  $t_{qs}$  bé hơn làm mất ý nghĩa thống kê của các hệ số ước lượng.

# CHƯƠNG V

## 5.5.2. Phát hiện biến không thích hợp

- + Để phát hiện 1 biến không thích hợp thì dùng kiểm định t
- + Để phát hiện 2 hay nhiều biến không thích hợp thì dùng kiểm định F.

# CHƯƠNG VI

## MÔ HÌNH CHUỖI THỜI GIAN

- Hai cách tiếp cận cơ bản với số liệu chuỗi thời gian:
  - + Mô hình hồi quy với số liệu chuỗi thời gian
  - + Phương pháp Box – Jenkins (phần B)
- Chuỗi thời gian có tính chất phụ thuộc thống kê giữa các quan sát nên sẽ không thỏa mãn giả thiết về mẫu ngẫu nhiên.
- Một số chuỗi thời gian: GDP hàng năm, mức lạm phát hàng tháng, chỉ số VN index hàng ngày, ...

# CHƯƠNG VI

## 6.1. SỐ LIỆU CHUỖI THỜI GIAN – MỘT SỐ KHÁI NIỆM

*Khái niệm chuỗi thời gian: chuỗi các quan sát được thu thập trên cùng một đối tượng tại các mốc thời gian khác cách đều nhau gọi là chuỗi thời gian.*

- + Thường dùng chỉ số  $t$  để chỉ thứ tự các quan sát.
- + Số liệu chuỗi thời gian phải được sắp xếp theo trình tự thời gian nhất định: các quan sát xảy ra sau luôn được sắp xếp ngay sau quan sát xảy ra trước nó

# CHƯƠNG VI

*Hiện tượng tự tương quan (autocorrelation): chuỗi  $X_t$  được gọi là có tự tương quan bậc  $p$  nếu  $\text{corr}(X_t, X_{t+p}) \neq 0$  với  $p \neq 0$ .*

Tự tương quan với số liệu chuỗi thời gian đôi khi được gọi là tương quan chuỗi (serial correlation)



# CHƯƠNG VI

## *Một số đặc trưng của số liệu chuỗi thời gian*

- + Số liệu chuỗi thời gian thường có tính tự tương quan, nghĩa là  $\text{corr}(Y_t, Y_{t-s})$  thường khác 0.
- + Số liệu chuỗi thời gian (về kinh tế, xã hội) thường chịu tác động của yếu tố mùa vụ. Yếu tố mùa vụ thường xuất hiện với các số liệu có tần suất xuất hiện bé hơn 1 năm (số liệu theo quý, theo tháng,...)
- + Đa phần chuỗi thời gian còn có yếu tố xu thế, chỉ xu thế tăng hay giảm trong một thời kì khá dài của chuỗi số.

# CHƯƠNG VI

## 6.2. MÔ HÌNH HỒI QUY VỚI CHUỖI THỜI GIAN

Xét mô hình hồi quy tuyến tính với số liệu chuỗi thời gian như sau

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + u_t$$

### 6.2.1. Các giả thiết của mô hình

Giả thiết TS1: Sai số ngẫu nhiên không tự tương quan

$$\text{corr}(u_t, u_s |_{X_2, \dots, X_k}) = 0; \quad \forall t \neq s$$

Giả thiết này còn gọi là mô hình không có tự tương quan.

Giả thiết TS1 thay thế cho giả thiết 1 của MHHQ với số liệu chéo vì giả thiết về mẫu ngẫu nhiên là không phù hợp với số liệu chuỗi thời gian.

# CHƯƠNG VI

Giả thiết TS2: Kỳ vọng có điều kiện của sai số ngẫu nhiên bằng 0

$$E(u_t | X_2, \dots, X_k) = 0; \quad \forall t$$

Từ giả thiết TS2 ta chứng minh được

(i)  $E(u_t) = 0$ ; mọi  $t$

(ii)  $\text{Cov}(u_t, X_s) = 0$  với mọi  $t \neq s$

Giả thiết TS2 đòi hỏi  $X$  và  $u$  không tương quan tại các thời điểm chéo nhau (mạnh hơn giả thiết 2)

**Biến ngoại sinh chặt**: Biến độc lập  $X_j$  được gọi là biến ngoại sinh chặt nếu  $\text{cov}(u_t, X_s) = 0$  với mọi  $t \neq s$

*Ví dụ 6.1*:  $CT_t = \beta_1 + \beta_2 TN_t + u$

$$LP_t = \beta_1 + \beta_2 M_t + u$$

# CHƯƠNG VI

Giả thiết TS3: Phương sai sai số là bằng nhau tại mọi thời điểm

$$\text{var}(u_t | X_2, \dots, X_k) = \sigma^2; \quad \forall t$$

Giả thiết TS4: Giữa các biến độc lập không có đa cộng tuyến hoàn hảo

Giả thiết TS5: Sai số ngẫu nhiên tuân theo quy luật chuẩn  
→  $u_t \sim N(0; \sigma^2)$

# CHƯƠNG VI

## 6.2.2. Tính chất của các ước lượng và bài toán suy diễn thống kê

Định lý 6.1: (ĐL Gauss – Markov cho mô hình hồi quy chuỗi thời gian) *Khi các giả thiết TS1 – TS4 thỏa mãn thì các ước lượng OLS là các ước lượng tuyến tính, không chệch và tốt nhất trong lớp các ước lượng tuyến tính không chệch.*

Định lý 6.2: *Khi các giả thiết TS1 – TS4 thỏa mãn thì phương sai của các ước lượng OLS được tính như sau:*

$$\text{var}(\hat{\beta}_j) = \frac{\sigma^2}{(1 - R_j^2) \sum_t x_{jt}^2}; \quad j = 2, 3, \dots, k$$

(t = 1, 2, ..., n, ...)

# CHƯƠNG VI

$$se(\hat{\beta}_j) = \frac{\hat{\sigma}}{\sqrt{(1-R_j^2) \sum_t x_{jt}^2}}; \quad j = 2, 3, \dots, k; \quad \hat{\sigma}^2 = \frac{\sum_t e_t^2}{n-k}$$

Trong đó  $\hat{\sigma}$  UL không chệch của  $\sigma$ .

**Định lý 6.3.** Khi các giả thiết TS1 – TS5 thỏa mãn thì các hệ số ước lượng có phân phối chuẩn

$$\hat{\beta}_j \sim N(\beta_j, \text{var}(\hat{\beta}_j))$$

Khi các giả thiết TS1 – TS5 thỏa mãn thì việc thực hiện các bài toán suy diễn thống kê và dự báo tương tự như chương 3.

# CHƯƠNG VI

## 6.3. MỘT SỐ MÔ HÌNH HỒI QUY CHUỖI THỜI GIAN CƠ BẢN

### 6.3.1. Mô hình hồi quy tĩnh

Xét mô hình hồi quy

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + u_t$$

Mô hình này được gọi là tĩnh theo nghĩa rằng chúng ta chỉ xét đến mối quan hệ giữa các biến số tĩnh tại cùng một thời điểm, nó cho phép xem xét mối quan hệ tức thời hoặc mối quan hệ cân bằng dài hạn giữa các biến số.

# CHƯƠNG VI

## 6.3.2. Mô hình hồi quy động

### a. Nhiễu trắng

*Chuỗi thời gian có kì vọng bằng 0, phương sai không đổi và hiệp phương sai bằng 0 được gọi là nhiễu trắng (white noise), kí hiệu là  $\varepsilon_t$*

(i)  $E(\varepsilon_t) = 0; \forall t$

(ii)  $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2; \forall t$

(iii)  $\text{cov}(\varepsilon_t, \varepsilon_s) = 0; \forall t \neq s$

Các chuỗi số trong kinh tế thường không phải là nhiễu trắng.



# CHƯƠNG VI

## *b. Mô hình có trễ phân phối hữu hạn*

Mô hình hồi quy trong đó có chứa cả giá trị hiện tại và các giá trị trễ của biến giải thích được gọi là mô hình có trễ phân phối (distributed lag model)

Mô hình có trễ phân phối bậc  $p$ :

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \dots + \beta_p X_{t-p} + u_t$$

Trong đó  $u_t$  là nhiễu trắng.

Mô hình trên cho rằng tác động của biến  $X$  lên biến  $Y$  sẽ triệt tiêu sau  $p$  thời kì.

# CHƯƠNG VI

**Ví dụ 6.2:** Cho rằng sự thay đổi của thu nhập có tác động trễ sau không quá 3 tháng lên chi tiêu, ta có mô hình:

$$CT_t = \alpha + \beta_0 TN_t + \beta_1 TN_{t-1} + \beta_2 TN_{t-2} + \beta_3 TN_{t-3} + u_t$$

Chẳng hạn

$$CT_t = 2,5 + 0,6TN_t + 0,15TN_{t-1} + 0,05TN_{t-2} + 0,01TN_{t-3} + e_t$$

Khi thu nhập một tháng nào đó bỗng nhiên tăng 1 triệu đồng, còn các tháng sau lại trở lại mức cũ thì

- + chi tiêu trong tháng đó sẽ tăng 0,6 triệu đồng,
- + chi tiêu tháng sau tăng 0,15 triệu đồng,
- + chi tiêu sau hai tháng tăng 0,05 triệu đồng,
- + sau 3 tháng tăng 0,01 triệu đồng
- + sau 4 tháng trở đi thì tác động này triệt tiêu.

# CHƯƠNG VI

*Ví dụ 6.2:*

$$CT_t = 2,5 + 0,6TN_t + 0,15TN_{t-1} + 0,05TN_{t-2} + 0,01TN_{t-3} + e_t$$

Khi thu nhập một tháng tăng 1 triệu đồng một cách lâu dài (ví dụ tăng lương) thì

- + chi tiêu trong tháng đó sẽ tăng 0,6 triệu đồng,
- + chi tiêu tháng sau tăng  $0,6 + 0,15 = 0,75$  triệu đồng,
- + chi tiêu sau 2 tháng tăng  $0,6 + 0,15 + 0,05 = 0,8$  triệu đồng,
- + sau 3 tháng tăng  $0,8 + 0,01 = 0,81$  triệu đồng
- + sau 4 tháng trở đi thì chi tiêu tăn 0,81 triệu đồng.

Như vậy tác động dài hạn của sự gia tăng thu nhập vĩnh viễn 1 triệu đồng lên chi tiêu là 0,81 triệu đồng

# CHƯƠNG VI

## *c. Mô hình tự hồi quy*

Mô hình tự hồi quy bậc p (AR(p) – autoregressive model)

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \dots + \beta_p Y_{t-p} + u_t$$

Trong đó  $u_t$  là nhiễu trắng.

Mô hình tự hồi quy có thể chứa biến khác

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \dots + \beta_p Y_{t-p} + \beta_{p+1} X_t + u_t$$

# CHƯƠNG VI

## 6.3.3. Mô hình với xu thế thời gian và yếu tố mùa vụ

### *a. Mô hình hồi quy với xu thế thời gian*

Nhiều chuỗi số kinh tế có xu thế gia tăng hoặc giảm sút theo thời gian. Các dạng phổ biến là

❖ *Xu thế tuyến tính:*  $Z_t = a + bt + e_t$

❖ *Xu thế phi tuyến*

▪ *Xu thế bậc hai:*  $Z_t = a + bt + ct^2 + e_t$

▪ *Xu thế dạng mũ:*  $\ln(Z_t) = a + bt + e_t$

Trong đó  $a, b, c$  là các hệ số thông thường,  $e_t$  là sai số ngẫu nhiên thể hiện tác động của các biến khác lên biến  $Z_t$ .

$e_t$  trong các mô hình trên thể hiện sự khác biệt của biến  $Z_t$  xung quanh xu thế tất định của nó (là  $bt, bt + ct^2$ )

# CHƯƠNG VI

❖ *Sự cần thiết của biến xu thế trong mô hình chuỗi thời gian*

Giả sử  $Y_t$  có xu thế tuyến tính và có mô hình dạng

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 t + u_t \quad (1)$$

So sánh với mô hình không có biến xu thế

$$Y_t = \alpha_1 + \alpha_2 X_t + v_t \quad (2)$$

Khi đó sai số ngẫu nhiên  $v_t$  sẽ chứa yếu tố xu thế, do đó kì vọng toán của nó sẽ thay đổi theo thời gian và giả thiết TS2 bị vi phạm: nếu  $\beta_3 > 0$  thì kì vọng của  $v_t$  tăng theo thời gian, nếu  $\beta_3 < 0$  thì kì vọng của  $v_t$  giảm theo thời gian.

# CHƯƠNG VI

Trường hợp cả  $X$  và  $Y$  đều có xu thế thời gian thì sẽ mắc phải vấn đề thiếu biến quan trọng vì  $X$  và  $u$  có tương quan với nhau do cùng chứa yếu tố xu thế.

Nếu ước lượng mô hình (2) trong trường hợp này sẽ thấy  $\hat{\alpha}_2$  có ý nghĩa thống kê,  $R^2$  thu được cao kể cả khi biến  $X$  và biến  $Y$  hoàn toàn không có quan hệ với nhau.

Hiện tượng trên gọi là ‘hồi quy giả mạo’ (spurious regression)

# CHƯƠNG VI

## *b. Mô hình hồi quy với yếu tố mùa vụ*

Nếu biến phụ thuộc  $Y$  chịu ảnh hưởng của yếu tố mùa vụ thì việc không đưa yếu tố này vào mô hình sẽ gây ra hiện tượng thiếu biến quan trọng.

Việc đưa thêm yếu tố mùa vụ vào mô hình hồi quy được thực hiện thông qua biến giả

Ví dụ: Mô hình hồi quy GDP theo vốn và lao động có yếu tố mùa vụ là

$$GDP = \beta_1 + \beta_2 K + \beta_3 L + \alpha_1 Q_1 + \alpha_2 Q_2 + \alpha_3 Q_3 + u$$

Với  $Q_i = 1$  với các số liệu quý  $i$  ( $i = 1, 2, 3$ )

$Q_i = 0$  với các quý khác.

Các hệ số  $\alpha_j$  có ý nghĩa tương tự như ở chương 4.



# CHƯƠNG VI

## 6.3.4. Mô hình với thay đổi cấu trúc

Khi một yếu tố cơ bản của nền kinh tế thay đổi thì sẽ gây ra sự thay đổi trong mối quan hệ giữa các biến số kinh tế.

- Quan hệ giữa tăng trưởng kinh tế và đầu tư tư nhân sẽ khác nhau giữa giai đoạn trước và sau khi đất nước hội nhập với kinh tế thế giới.
- Quan hệ giữa xuất khẩu và tăng trưởng thay đổi khi chính sách tỷ giá thay đổi từ cố định sang thả nổi.
- Tác động của lao động lên sản lượng thay đổi khi có đột phá trong tiến bộ công nghệ.
- Quan hệ giữa lãi suất tiền gửi và lượng tiền gửi hàng tháng thay đổi khi có sự xuất hiện của thị trường chứng khoán.

Khi có sự thay đổi cấu trúc như trên thì sử dụng biến giả<sup>145</sup>

# CHƯƠNG VI

## 6.4. TÍNH CHẤT MẪU LỚN CỦA ƯỚC LƯỢNG OLS

### 6.4.1. Một số khái niệm

#### *a. Chuỗi dừng*

Một chuỗi  $X_t$  với  $E(X_t^2)$  hữu hạn được gọi là chuỗi dừng (stationary series) nếu thỏa mãn 3 điều kiện sau đây:

- (i) Kỳ vọng không đổi:  $E(X_t) = 0$
- (ii) Phương sai không đổi:  $\text{Var}(X_t) = \sigma^2$
- (iii) Hiệp phương sai không phụ thuộc vào thời điểm tính toán:  $\text{cov}(X_t, X_{t-s}) = \rho_s$  với mọi  $t$ .

Chuỗi không thỏa mãn một trong 3 điều kiện trên gọi là chuỗi không dừng. Đa phần các chuỗi số kinh tế đều là chuỗi không dừng.

# CHƯƠNG VI

## *b. Chuỗi phụ thuộc yếu*

Chuỗi dừng  $X_t$  được gọi là chuỗi phụ thuộc yếu (weakly dependent) nếu hệ số tương quan giữa  $X_t$  và  $X_{t+h}$  tiến đến 0 với một tốc độ khá nhanh khi  $h$  tiến ra vô cùng.

Ví dụ chuỗi tự hồi quy bậc 1

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \varepsilon_t$$

là chuỗi dừng khi  $|\beta_2| < 1$  (chứng minh?)

# CHƯƠNG VI

## 6.4.2. Các giả thiết thay thế

Xét mô hình

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + u_t$$

Trong đó các biến  $X_{jt}$  có thể là biến trễ của biến độc lập hoặc biến phụ thuộc.

Giả thiết TS0': Các chuỗi  $\{Y_t, X_{2t}, \dots, X_{kt}\}$  là chuỗi dừng và phụ thuộc yếu.

Giả thiết TS1': Sai số ngẫu nhiên không tự tương quan

$$\text{corr}(u_t, u_s |_{X_2, \dots, X_k}) = 0; \quad \forall t \neq s$$

Giả thiết này trùng với giả thiết TS1

# CHƯƠNG VI

Giả thiết TS2': Tại mỗi  $t$  ta có  $E(u_t | X_{2t}, \dots, X_{kt}) = 0$

Giả thiết này chỉ đòi hỏi tính ngoại sinh thông thường của các biến giải thích và hoàn toàn tương tự như giả thiết 2 trong mô hình với số liệu chéo. Giả thiết TS2' yếu hơn giả thiết TS2 (đòi hỏi biến ngoại sinh chặt).

Giả thiết TS3': Phương sai sai số là bằng nhau tại mọi thời điểm

$$\text{var}(u_t | X_{2t}, \dots, X_{kt}) = \sigma^2; \quad \forall t$$

Giả thiết TS4': Giữa các biến độc lập không có quan hệ đa cộng tuyến hoàn hảo.

# CHƯƠNG VI

## 6.4.3. Tính chất mẫu lớn của ước lượng

**Định lý 6.4:** Khi các giả thiết TS0'-TS2' thỏa mãn thì các hệ số ước lượng thu được từ phương pháp OLS là các ước lượng vững, nghĩa là

$$P(|\widehat{\beta}_j^{(n)} - \beta_j| < \varepsilon) \xrightarrow{n \rightarrow \infty} 1; \quad \forall \varepsilon > 0$$

Trong đó  $\widehat{\beta}_j^{(n)}$  là ước lượng của  $\beta_j$  với mẫu kích thước  $n$ .

**Định lý 6.5:** Khi các giả thiết TS0'-TS4' thỏa mãn thì các ước lượng OLS là tiệm cận hiệu quả trong lớp các ước lượng tuyến tính và vững, nghĩa là

$$\lim_{n \rightarrow \infty} \left[ \text{var} \sqrt{n} \left( \widehat{\beta}_j - \beta_j \right) \right] \leq \lim_{n \rightarrow \infty} \left[ \text{var} \sqrt{n} \left( \widetilde{\beta}_j - \beta_j \right) \right]; \quad j = 1, 2, \dots, k$$

$\widetilde{\beta}_j$  là ước lượng tuyến tính và vững bất kì của  $\beta_j$ .

# CHƯƠNG VI

**Định lý 6.6:** Khi các giả thiết TS0'-TS4' thỏa mãn thì các hệ số ước lượng thu được từ phương pháp OLS có phân phối xấp xỉ chuẩn, khi đó các thống kê  $t$ ,  $F$  có phân phối xấp xỉ quy luật Student và quy luật Fisher tương ứng.

**Kết:** Khi  $n$  đủ lớn và các giả thiết TS0'-TS4' thỏa mãn thì các kết quả phân tích sử dụng ước lượng OLS là đáng tin cậy.

Các chuỗi số được xem xét đến trong các định lý nói trên là chuỗi dừng và phụ thuộc yếu. Thường chuỗi dừng cũng phụ thuộc yếu nên ở đây chỉ nói đơn giản là chuỗi dừng.

Sử dụng bộ giả thiết TS0'-TS4', trước khi hồi quy số liệu chuỗi thời gian ta cần phải xem xét tính dừng của chuỗi số trong mô hình.

# CHƯƠNG VI

Để kiểm tra tính dừng của chuỗi số ta xem xét đồ thị của chuỗi số theo thời gian.

Số liệu trong phân tích kinh tế xã hội thường là chuỗi không dừng. Để biến đổi chuỗi không dừng thành chuỗi dừng thì lấy sai phân hoặc xem xét biến mức tăng trưởng của các chuỗi số này.

**Ví dụ:** Xét mô hình

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

Giả sử  $Y_t$  là chuỗi không dừng thì sai phân của chuỗi này là

$$Y'_t = Y_t - Y_{t-1}$$