

Chương 3

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

Nội dung

1. Quy luật phân phối xác suất của một số thống kê mẫu
2. Bài toán xây dựng khoảng tin cậy (KTC) cho các hệ số hồi quy (HSHQ)
3. Bài toán kiểm định giả thuyết thống kê (KĐGT TK) về hệ số hồi quy
4. Một số kiểm định khác
5. Dự báo giá trị của biến phụ thuộc và sai số dự báo

1. Quy luật phân phối xác suất của một số thống kê mẫu

- Xét mô hình: $Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + u$
- Hàm hồi quy mẫu thu được từ mẫu ngẫu nhiên kích thước n : $\{(X_{2i}, \dots, X_{ki}, Y_i), i = 1, 2, \dots, n\}$

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

- Từ kết quả ước lượng, để đưa ra các suy diễn thống kê cho các hệ số hồi quy tổng thể, chúng ta cần biết quy luật phân phối của các $\hat{\beta}_j \quad (j = 1, 2, \dots, k)$
- Giả thiết 5: Sai số ngẫu nhiên tuân theo quy luật chuẩn

$$u_i \sim N(0, \sigma^2)$$

- Định lý 3.1: Khi các giả thiết 1 - 5 thỏa mãn ta có:

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

- Định lý 3.2: Khi các giả thiết 1- 5 thỏa mãn ta có:
với $j = 2, 3, \dots, k$ thì

$$T = \frac{\hat{\beta}_j - \beta_j}{\text{Se}\left(\hat{\beta}_j\right)} \sim T(n-k)$$

Tương tự ta có:

$$T = \frac{(a \cdot \hat{\beta}_j + b \cdot \hat{\beta}_s) - (a \cdot \beta_j + b \cdot \beta_s)}{\text{Se}\left(a \cdot \hat{\beta}_j + b \cdot \hat{\beta}_s\right)} \sim T(n - k)$$

2. Bài toán xây dựng KTC cho các HSHQ

2.1 Khoảng tin cậy cho một hệ số hồi quy: đánh giá tác động khi một biến độc lập thay đổi

- Chọn thống kê:
$$T = \frac{\hat{\beta}_j - \beta_j}{Se(\hat{\beta}_j)} \sim T(n-k)$$
- KTC với độ tin cậy $(1-\alpha)$ của β_j được xác định

$$P\left(-t_{\alpha_1}^{(n-k)} \leq \frac{\hat{\beta}_j - \beta_j}{Se(\hat{\beta}_j)} \leq t_{\alpha_2}^{(n-k)}\right) = 1 - \alpha$$

$$\text{với } \alpha_1, \alpha_2 \geq 0; \quad \alpha_1 + \alpha_2 = \alpha$$

$$\hat{\beta}_j - Se(\hat{\beta}_j)t_{\alpha_2}^{(n-k)} \leq \beta_j \leq \hat{\beta}_j + Se(\hat{\beta}_j)t_{\alpha_1}^{(n-k)}$$

- KTC đối xứng với độ tin cậy $(1 - \alpha)$ của β_j là:

$$\left(\hat{\beta}_j - Se(\hat{\beta}_j) t_{\alpha/2}^{(n-k)} \leq \beta_j \leq \hat{\beta}_j + Se(\hat{\beta}_j) t_{\alpha/2}^{(n-k)} \right)$$

- Khoảng tin cậy tối thiểu với độ tin cậy $(1 - \alpha)$ của β_j là:

$$\left(\hat{\beta}_j - Se(\hat{\beta}_j) t_{\alpha}^{(n-k)} \leq \beta_j \right)$$

- Khoảng tin cậy tối đa với độ tin cậy $(1 - \alpha)$ của β_j là :

$$\left(\beta_j \leq \hat{\beta}_j + Se(\hat{\beta}_j) t_{\alpha}^{(n-k)} \right)$$

2.2 KTC cho biểu thức của hai HSHQ: đánh giá tác động khi hai biến độc lập cùng thay đổi

- Khi X_j tăng a đơn vị, X_s tăng b đơn vị thì giá trị trung bình của Y tăng $(a.\beta_j + b.\beta_s)$ đơn vị.
- Ước lượng điểm của $a.\beta_j + b.\beta_s$ là: $a.\hat{\beta}_j + b.\hat{\beta}_s$
- Sai số chuẩn:

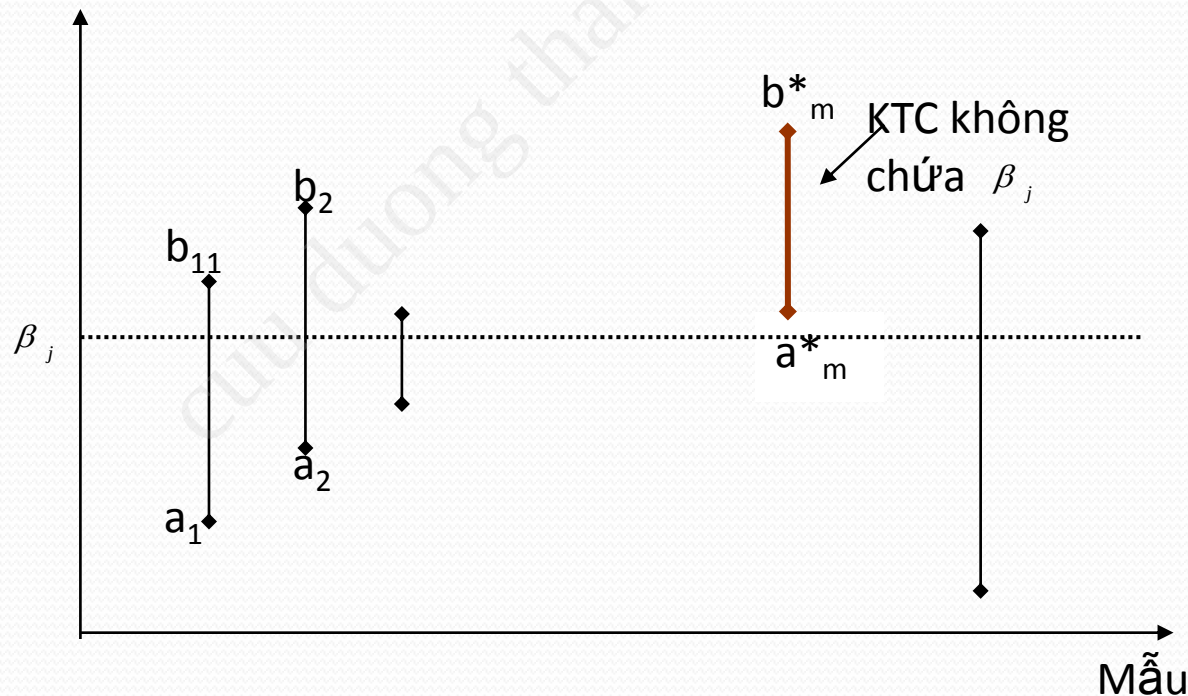
$$\begin{aligned} Se(a.\hat{\beta}_j + b.\hat{\beta}_s) &= \sqrt{a^2 \cdot Var(\hat{\beta}_j) + b^2 \cdot Var(\hat{\beta}_s) + 2.a.b.Cov(\hat{\beta}_j, \hat{\beta}_s)} \\ &= \sqrt{a^2 \cdot Se^2(\hat{\beta}_j) + b^2 \cdot Se^2(\hat{\beta}_s) + 2.a.b.Cov(\hat{\beta}_j, \hat{\beta}_s)} \end{aligned}$$

- KTC đối xứng với độ tin cậy $(1 - \alpha)$ của $a.\beta_j + b.\beta_s$ là:

$$\begin{aligned} (a.\hat{\beta}_j + b.\hat{\beta}_s) - Se(a.\hat{\beta}_j + b.\hat{\beta}_s).t_{\alpha/2}^{(n-k)} &\leq (a.\beta_j + b.\beta_s) \\ &\leq (a.\hat{\beta}_j + b.\hat{\beta}_s) + Se(a.\hat{\beta}_j + b.\hat{\beta}_s).t_{\alpha/2}^{(n-k)} \end{aligned}$$

2.3. Ý nghĩa của khoảng tin cậy

- Với độ tin cậy, chẳng hạn 95%, KTC của β_j được hiểu như sau: nếu lấy nhiều lần các mẫu một cách ngẫu nhiên từ cùng một tổng thể, thì có khoảng 95% số KTC được xây dựng từ các mẫu này là có chứa giá trị β_j .



2.4 Các yếu tố ảnh hưởng đến độ dài khoảng tin cậy

- KTC hẹp cho biết chính xác hơn về giá trị của hệ số cần ước lượng; ngược lại, một KTC quá rộng thì thông tin về giá trị của hệ số cần ước lượng là kém chính xác.
- Độ dài KTC đối xứng cho các hệ số β_j bằng $2 se(\hat{\beta}_j) t_{\alpha/2}(n-k)$
- Giá trị này phụ thuộc vào các yếu tố:
 - Số bậc tự do (n-k): số bậc tự do càng bé thì $t_{\alpha/2}(n-k)$ càng lớn và như vậy KTC càng rộng;
 - Việc thêm biến làm giảm bậc tự do nhưng tăng R^2
 - Mối tương quan tuyến tính giữa X_j và các biến độc lập còn lại trong mô hình.

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

3. Bài toán KĐGT Thống kê về HSHQ

3.1 Kiểm định giả thuyết về một hệ số hồi quy

- Để KĐGT $H_0 : \beta_j = \beta_j^*$ ta giả sử H_0 đúng và có Tiêu chuẩn kiểm định:

$$T = \frac{\hat{\beta}_j - \beta_j^*}{Se\left(\hat{\beta}_j\right)} \sim T(n - k)$$

- Tùy theo giả thuyết H_1 ta có các miền bác bỏ khác nhau.

Loại giả thuyết	Giả thuyết H_0	Giả thuyết đối H_1	Miền bác bỏ giả thuyết H_0 với mức ý nghĩa α
Hai phía	$\beta_j = \beta_j^*$	$\beta_j \neq \beta_j^*$	$W_\alpha = \{t, t > t_{\alpha/2}^{(n-k)}\}$
Phía phải	$\beta_j \leq \beta_j^*$	$\beta_j > \beta_j^*$	$W_\alpha = \{t, t > t_\alpha^{(n-k)}\}$
Phía trái	$\beta_j \geq \beta_j^*$	$\beta_j < \beta_j^*$	$W_\alpha = \{t, t < -t_\alpha^{(n-k)}\}$

- Trường hợp đặc biệt: Kiểm định giả thuyết $H_0 : \beta_j = 0$

- Khi đó:

$$t = \frac{\hat{\beta}_j - 0}{Se(\hat{\beta}_j)} = T - statistic (X_j)$$

- Nếu kết quả kiểm định cho thấy $\beta_j \neq 0$ có nghĩa là hệ số $\hat{\beta}_j$ có ý nghĩa thống kê.

3.2 Kiểm định giả thuyết về một ràng buộc giữa các hệ số hồi quy - kiểm định T

- Để kiểm định giả thuyết $H_0: a.\beta_j + b.\beta_s = a^*$ (a^* : hằng số bất kỳ) ta giả sử H_0 đúng và có tiêu chuẩn kiểm định:

$$T = \frac{(a.\hat{\beta}_j + b.\hat{\beta}_s) - a^*}{Se\left(a.\hat{\beta}_j + b.\hat{\beta}_s\right)} \sim T(n - k)$$

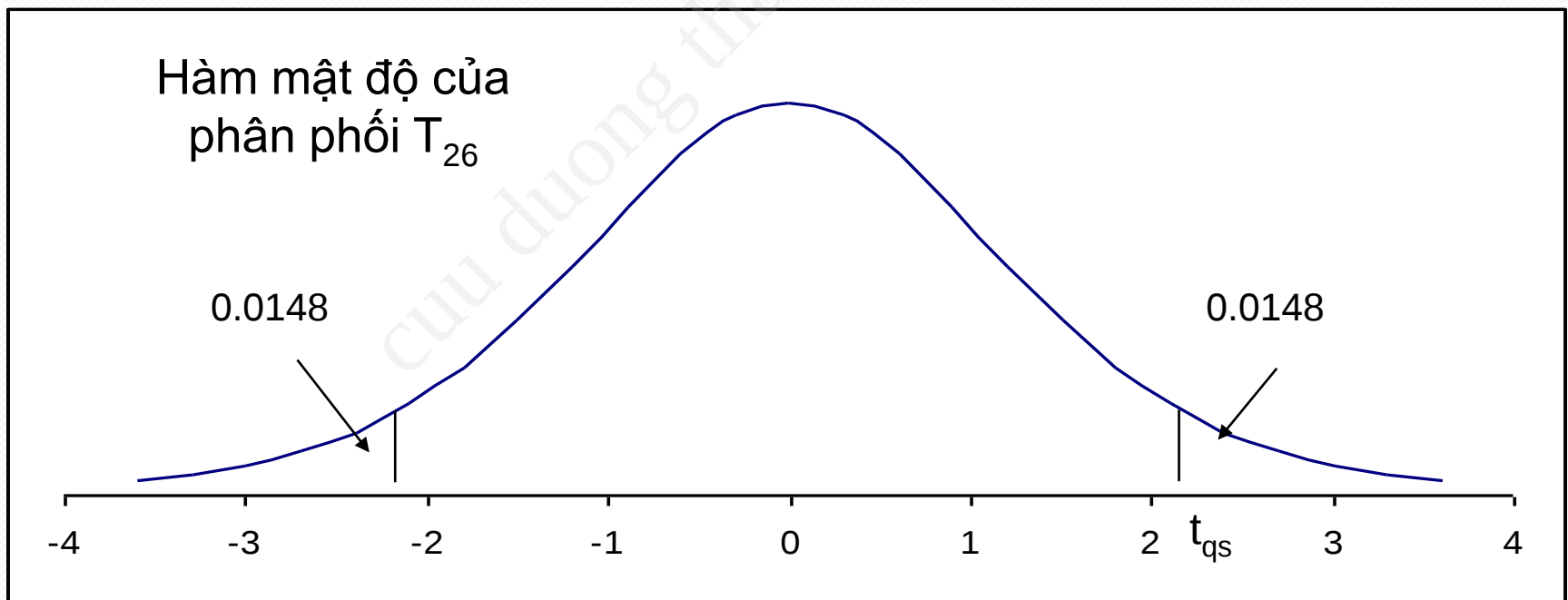
- Tùy theo giả thuyết H_1 ta có các miền bác bỏ khác nhau (Tương tự trường hợp kiểm định một hệ số hồi quy)

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

- Các kết luận trong quá trình kiểm định giả thuyết nhiều khi phụ thuộc vào việc lựa chọn mức ý nghĩa α .
- Giá trị quan sát của thống kê kiểm định khác nhau thể hiện mức độ mạnh yếu của chứng cứ ủng hộ giả thuyết H_1 là khác nhau.
- Giá trị xác suất P được định nghĩa là **mức ý nghĩa nhỏ nhất mà giả thuyết H_0 bị bác bỏ** tương ứng với giá trị quan sát của thống kê kiểm định này

$$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{cases}$$

- $t_{qs} = 2.3$, với $n - k = 26$ ta có giá trị P tương ứng là $P = P(|T(26)| > 2.3) = 2 \times 0.0148 = 0.0296$
- Con số $P = 0.0296$ cho thấy với mọi mức ý nghĩa lớn hơn 0.0296 thì giả thuyết H_0 nói trên đều bị bác bỏ



- **Quy tắc kiểm định sử dụng giá trị xác suất:**
“Nếu giá trị xác suất P nhỏ hơn mức ý nghĩa α thì ta sẽ bác bỏ giả thuyết H_0 ”.
- Với các kiểm định khác chúng ta sẽ có các khái niệm tương tự về giá trị xác suất P , do đó sẽ không được nhắc lại trong các phần tiếp theo.

Ví dụ 3.1: Y phụ thuộc K, L

Dependent Variable: **Y**

Method: Least Squares

Sample: 1 100

Included observations: 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-485.9608	95.85601	-5.069695	0.0000
K	1.292811	0.044404	29.11470	0.0000
L	2.214092	0.050943	43.46253	0.0000
R-squared	0.964118	Sum sq. resid	7221985.	
F-statistic	1303.136	Prob(F-statistic)	0.000000	

- Với Y là sản lượng, K là vốn, L là lao động
- Hiệp phương sai ước lượng hai hệ số góc = 0,00012.

Ví dụ 3.1 (tiếp)

- (a) Giải thích ý nghĩa kết quả ước lượng
- (b) Biến K có thực sự giải thích cho Y ?
- (c) Hệ số nào có ý nghĩa thống kê?
- (d) L tăng 1 đơn vị, K không đổi thì Y thay đổi thế nào?
- (e) K tăng 1 đơn vị, L không đổi thì Y có tăng 2 đơn vị?
- (f) K tăng 1 đơn vị, L giảm một đơn vị thì Y có giảm?
- (g) K và L cùng tăng 1 đơn vị thì Y thay đổi thế nào?
- (h) Kiểm định giả thuyết tổng hai hệ số góc = 3,6?

3.4 Kiểm định giả thuyết về nhiều ràng buộc của các hệ số hồi quy- kiểm định F

- Xét mô hình hồi quy về tiền lương sau:

$$wage = \beta_1 + \beta_2 Edu + \beta_3 Medu + \beta_4 Ssibs + u$$

trong đó:

- *Wage*: tiền lương, *Edu*: trình độ học vấn
- *Meduc*: trình độ học vấn người mẹ, *Ssibs*: số anh chị em trong gia đình của người lao động;
- Có phải “học vấn người mẹ” và “số anh chị em trong gia đình” đồng thời không ảnh hưởng tới tiền lương người lao động.
- Tức là kiểm định giả thuyết:
$$\begin{cases} H_0 : \beta_3 = \beta_4 = 0 \\ H_1 : \beta_3^2 + \beta_4^2 \neq 0 \end{cases}$$

- Xét mô hình: $Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + U_i$ 

$$\begin{cases} H_0 : \beta_{k-m+1} = \beta_{k-m+2} = \dots = \beta_k = 0 \\ H_1 : \beta_{k-m+1}^2 + \beta_{k-m+2}^2 + \dots + \beta_k^2 \neq 0 \end{cases}$$

- Giả sử giả thuyết H_0 đúng, mô hình k biến sẽ bị thu hẹp:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_{k-m} X_{(k-m)i} + V_i$$
 

- Tiêu chuẩn kiểm định:

$$F = \frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(n-k)}$$

$$= \frac{(R_{UR}^2 - R_R^2)/m}{(1 - R_{UR}^2)/(n-k)}$$

$$W_\alpha = \left\{ F, F > F_\alpha^{(m, (n-k))} \right\}$$



$$\begin{array}{ccccccc} \text{Wage} & = & 2404 & + & 86.12 \text{ Edu} & - & 14.88 \text{ Meduc} & - & 30.25 \text{ Ssibs} & + & e \\ (\text{se}) & & (454) & & (36.78) & & (19.61) & & (39.18) & & \end{array}$$

$$\text{RSS} = 3649563, n=32$$

Đề kiểm định $H_0 : \beta_3 = \beta_4 = 0; H_1 : \beta_3^2 + \beta_4^2 \neq 0$

Ước lượng mô hình hồi quy có ràng buộc:

$$wage = \alpha_1 + \alpha_2 Edu + u \quad \text{thu được RSS} = 3770332$$

Do đó giá trị quan sát của thống kê kiểm định bằng:

$$F_{qs} = \frac{(3770332 - 3649563) / 2}{3649563 / 28} = 0.46$$

$F_{qs} < f_{0.05}(2, 28) = 3.49$ suy ra không đủ chứng cứ để bác bỏ H_0 .

Như vậy có thể cho rằng các biến *Medu* và *Ssibs* là cùng không tác động đến mức lương của người lao động.

3.5 Kiểm định về sự phù hợp của hàm hồi quy

- Xét mô hình: $Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + U_i$

- Kiểm định giả thuyết:

$$\begin{cases} H_0 : \beta_2 = \beta_3 = \dots = \beta_k = 0 \\ H_1 : \beta_2^2 + \beta_3^2 + \dots + \beta_k^2 \neq 0 \end{cases}$$

Trường hợp đặc biệt của kiểm định giả thuyết về nhiều ràng buộc của các hệ số hồi quy

- Giả thuyết H_0 ngụ ý toàn bộ các biến độc lập trong mô hình đều không ảnh hưởng đến biến phụ thuộc.
- Nếu H_0 là đúng thì mô hình là không có ý nghĩa, hay còn được gọi là mô hình không phù hợp.
- Do đó kiểm định trên còn được gọi là kiểm định về sự phù hợp của hàm hồi quy.



Thủ tục kiểm định

- Bước 1: ước lượng mô hình hồi qui gốc – mô hình không có điều kiện ràng buộc:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + U_i \quad \text{Thu được } R^2$$

- Bước 2: Ước lượng mô hình với điều kiện ràng buộc:

$Y_i = \beta_1 + V_i$ Mô hình này không có biến độc lập nên có hệ số xác định bằng 0.

- Tiêu chuẩn kiểm định:

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

- Miền bác bỏ:

$$W_\alpha = \left\{ F, F > F_{\alpha}^{((k-1), (n-k))} \right\}$$

Một số dạng ràng buộc

- MH gốc (U): $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u$

- **Dạng 1: Các hệ số bằng 0**

$$H_0: \beta_3 = \beta_4 = 0 \quad (m = 2)$$

$$\text{MH có ràng buộc (R): } Y = \beta_1 + \beta_2 X_2 + u$$

- **Dạng 2: Tổng hiệu các hệ số bằng 0**

$$H_0: \beta_3 + \beta_4 = 0 \quad (m = 1)$$

$$\text{MH (R): } Y = \beta_1 + \beta_2 X_2 + \beta_3 (X_3 - X_4) + u$$

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 (X^*) + u$$

Một số dạng ràng buộc

- MH gốc (U): $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u$

- **Dạng 3: Tổng các hệ số bằng số khác 0**

$$H_0: \beta_3 + \beta_4 = 1 \quad (m = 1)$$

$$\text{MH (R):} \quad Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + (1 - \beta_3) X_4 + u$$

$$Y - X_4 = \beta_1 + \beta_2 X_2 + \beta_3 (X_3 - X_4) + u$$

$$Y^* = \beta_1 + \beta_2 X_2 + \beta_3 (X^*) + u$$

Lưu ý: Biến phụ thuộc thay đổi

- **Dạng 4: Các ràng buộc khác**

$$H_0: \beta_2 = 2 \text{ và } \beta_3 + \beta_4 = 1 \quad (m = 2)$$

Ví dụ 3.1 (tiếp)

- (h) Kiểm định sự phù hợp của mô hình
- (i) Khi bớt biến L thì tổng bình phương phần dư tăng lên đến $1,48E+08$. Có nên bỏ biến đó không?
- (j) Khi bớt biến K thì hệ số xác định giảm xuống còn $0,65$. Vậy có nên bỏ biến đó không? So sánh kết quả với kiểm định T
- (k) Khi thêm hai biến K^2 và K^3 vào mô hình thì hệ số xác định tăng lên đến $0,9664$. Vậy có nên thêm hai biến đó vào không?

3.6 So sánh kiểm định T và kiểm định F

Trường hợp kiểm định một ràng buộc

- Kiểm định cặp giả thuyết: $H_0 : \beta_j = \beta^*; H_1 : \beta_j \neq \beta^*$
- Có thể áp dụng hai loại kiểm định T và F;
- Kết luận là hoàn toàn giống nhau; giá trị xác suất của hai thống kê quan sát cũng như nhau, vì:

$$(t_{qs})^2 = \left(\frac{\hat{\beta}_j - \beta^*}{se(\hat{\beta}_j)} \right)^2 = F_{qs} \quad \text{và} \quad t_{\alpha/2, (n-k)}^2 = F_{\alpha}(1, n-k)$$

- Do đó giá trị xác suất P của hai thống kê quan sát là bằng nhau khi kiểm định trên cùng một mẫu.
- Như vậy, trong trường hợp này ta sẽ sử dụng kiểm định T vì tính toán dễ dàng, thuận lợi hơn.

Trường hợp kiểm định đồng thời nhiều hơn một ràng buộc

- Nếu kiểm định sự **bằng không** của hai hệ số thì có thể dùng kiểm định T để kiểm định lần lượt cho từng hệ số được không?
- **Không nên**, vì việc sử dụng kiểm định T như vậy là không xác đáng và không đáng tin cậy trong một số tình huống.
- Một trong những nguyên nhân gây ra sự khác biệt trong kết luận khi sử dụng kiểm định T và kiểm định F là sự **đa cộng tuyến** giữa các biến độc lập có trong mô hình.

Hộp 3.4: Kiểm định hệ số hồi quy

- Nếu kiểm định giả thuyết về một 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 hai ràng buộc trở lên: sử dụng kiểm định F - chỉ sử dụng được cho các ràng buộc dạng đẳng thức.

3.5. Kiểm định Khi – Bình phương

- Kiểm định m ràng buộc, sử dụng hàm hợp lý $\ln(L)$, so sánh χ_{qs}^2 với $\chi_{\alpha}^2(m)$
- Ví dụ 3.1, kiểm định giả thuyết $\beta_K + \beta_L = 3,6$

Wald Test:

Test Statistic Value		df	Prob
t-statistic	-1.341821	97	0.1828
F-statistic	1.800483	(1, 97)	0.1828
Chi-square	1.800483	1	0.1797

Null Hypothesis: **C(2) + C(3) = 3.6**

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
-3.6 + C(2) + C(3)	-0.093098	0.069382

Ví dụ 3.2: $\ln(Y)$ phụ thuộc $\ln(K)$, $\ln(L)$

- Với Y là sản lượng, K là vốn, L là lao động

Dependent Variable: **LOG(Y)**

Sample: 1 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.416571	0.114175	3.648529	0.0004
LOG(K)	0.621661	0.014506	42.85566	0.0000
LOG(L)	0.476395	0.005883	80.97390	0.0000
R-squared	0.988628	Sum sq. resid	0.204993	
F-statistic	4216.348	Prob(F-statistic)	0.000000	

- Hiệp phương sai ước lượng hai hệ số góc xấp xỉ 0
- (a) Kiểm định tính hiệu quả theo quy mô bằng kiểm định T

Ví dụ 3.2 (tiếp)

- (b) Kiểm định giả thuyết “quá trình sản xuất có hiệu quả không đổi theo quy mô” qua kết quả dưới đây

Dependent Variable: LOG(Y/L)			Included observations: 100	
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.131938	0.007945	142.4783	0.0000
LOG(K/L)	0.537318	0.006444	83.37731	0.0000
R-squared	0.986099	Mean dep var	1.615989	
Sum sq resid	0.288245	F-statistic	6951.776	

4. Một số kiểm định khác

- Khi giả thiết 5 được thỏa mãn;
- Khi giả thiết 5 không được thỏa mãn;
- (sinh viên tự nghiên cứu giáo trình)

5. Dự báo giá trị của biến phụ thuộc và sai số dự báo

5.1 Dự báo giá trị của biến phụ thuộc

- Xét mô hình:

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

- SRF:

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

- Giả sử cho $X=X_0$, cần dự báo $E(Y/X_0)$
- Ước lượng điểm của $E(Y/X_0)$ là: $\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$
- Khi giả thiết 1-4 thỏa mãn ta có:

$$T = \frac{\hat{Y}_0 - E(Y/X_0)}{Se\left(\hat{Y}_0\right)} \sim T(n-2)$$

$$Var (\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right] = \frac{\sigma^2}{n} + (X_0 - \bar{X})^2 \cdot \frac{\sigma^2}{\sum x_i^2}$$

$$Var (\hat{Y}_0) = \frac{\sigma^2}{n} + (X_0 - \bar{X})^2 \cdot Var (\hat{\beta}_2)$$

$$Var (\hat{Y}_0) = \frac{\hat{\sigma}^2}{n} + (X_0 - \bar{X})^2 \cdot [Se (\hat{\beta}_2)]^2 ; Se (\hat{Y}_0) = \sqrt{Var (\hat{Y}_0)}$$

- Khoảng tin cậy với độ tin cậy $(1-\alpha)$ cho giá trị trung bình của biến phụ thuộc khi $X = X_0$ là:

$$\left(\hat{Y}_0 - t_{\alpha/2}^{(n-2)} Se \left(\hat{Y}_0 \right) \leq E(Y / X_0) \leq \hat{Y}_0 + t_{\alpha/2}^{(n-2)} Se \left(\hat{Y}_0 \right) \right)$$

5.2. Đánh giá sai số dự báo

- Sai số dự báo được tính dựa trên sự sai lệch giữa giá trị thực tế và giá trị ước lượng của biến phụ thuộc.

- Căn bậc hai của trung bình bình phương sai số;

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{n}}$$

- Sai số trung bình tuyệt đối

$$MAE = \frac{\sum_{i=1}^n |\hat{Y}_i - Y_i|}{n}$$

- Sai số trung bình tuyệt đối tính theo phần trăm

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{\hat{Y}_i - Y_i}{Y_i} \right|}{n}$$

- Chỉ có MAPE là không phụ thuộc vào đơn vị đo.

- Thông thường với các số liệu kinh tế, yêu cầu sai số dự báo nhỏ hơn 5%.
- Tuy nhiên với một số biến số như chỉ số VNINDEX hay chỉ số giá CPI theo tháng thì sai số 5% được cho là quá lớn.

Ví dụ 3.3: Dự báo cho Y theo K, L

- Đánh giá dự báo cho Y qua các mô hình của K và L
- MH với biến phụ thuộc $\ln(Y)$ đã quy về giá trị của Y

Mô hình	RMSE	MAE	MAPE
2.1(a) : lin-lin	268.74	218.47	6.86
2.1(b) : log-log	151.79	111.37	3.30
2.1(c) : lin-log	446.67	312.39	11.99
2.1(d) : log-lin	524.88	360.99	11.17
2.1(e) : tương tác	190.32	147.39	5.74

Ví dụ 3.3 (tiếp)

- MAPE của các mô hình theo các nhóm quan sát

Mô hình	100 qs	50 qs đầu	50 qs sau
2.1(a) : lin-lin	6.86	7.25	6.47
2.1(b) : log-log	3.30	3.20	3.40
2.1(c) : lin-log	11.99	9.94	14.04
2.1(d) : log-lin	11.17	12.41	9.94
2.1(e) : tương tác	5.74	5.98	5.51

Trình bày kết quả nhiều mô hình

- Kí hiệu *, **, *** : tương ứng hệ số có ý nghĩa thống kê ở mức 10%, 5%, 1%
- Kết quả mô hình với một số hộ gia đình

Biến phụ thuộc CONS (Chỉ tiêu)	MH (1)	MH (2)	MH (3)
C	10.6	2.12***	-2.8
INCOME (Thu nhập)	0.56***		0.58***
SIZE (Số người)		3.12**	3.31***
R-sq	0.45	0.257	0.74
MAPE	13.17%	12.72%	8.48%

Tóm tắt Chương 3

- Giả thiết: Sai số ngẫu nhiên phân phối chuẩn
- Khoảng tin cậy cho từng hệ số, nhiều hệ số
- Kiểm định T về các hệ số, hệ số có ý nghĩa thống kê
- Kiểm định F về hệ số và sự phù hợp
- Kiểm định thêm, bớt biến, ràng buộc
- Dự báo và đánh giá sai số dự báo