

Chương 1. MỞ ĐẦU
Chương 2. ƯỚC LƯỢNG VÀ KIỂM ĐỊNH MÔ HÌNH HỒI QUY ĐƠN
Chương 3. MÔ HÌNH HỒI QUY BỘI
Chương 4. ĐA CỘNG TUYẾN
Chương 5. PHƯƠNG SAI SAI SỐ THAY ĐỔI
Chương 6. TỰ TƯƠNG QUAN
Chương 7. HỒI QUY VỚI BIẾN GIẢ
Chương 8. LỰA CHỌN VÀ CHỈ ĐỊNH MÔ HÌNH

KINH TẾ LƯỢNG- ECONOMETRICS

Phạm Văn Khánh

Contents

- 1 Chương 1. MỞ ĐẦU
- 2 Chương 2. ƯỚC LƯỢNG VÀ KIỂM ĐỊNH MÔ HÌNH HỒI QUI ĐƠN
- 3 Chương 3. MÔ HÌNH HỒI QUY BỘI
- 4 Chương 4. ĐA CỘNG TUYẾN
- 5 Chương 5. PHƯƠNG SAI SAI SỐ THAY ĐỔI

Chương 1. MỞ ĐẦU

Chương 2. ƯỚC LƯỢNG VÀ KIỂM ĐỊNH MÔ HÌNH HỒI QUY ĐƠN

Chương 3. MÔ HÌNH HỒI QUY BỘI

Chương 4. ĐA CỘNG TUYẾN

Chương 5. PHƯƠNG SAI SAI SỐ THAY ĐỔI

Chương 6. TỰ TƯƠNG QUAN

Chương 7. HỒI QUY VỚI BIẾN GIẢ

Chương 8. LỰA CHỌN VÀ CHỈ ĐỊNH MÔ HÌNH

Chương 1. MỞ ĐẦU

Chương 1. MỞ ĐẦU
Chương 2. ƯỚC LƯỢNG VÀ KIỂM ĐỊNH MÔ HÌNH HỒI QUI ĐƠN
Chương 3. MÔ HÌNH HỒI QUY BỘI
Chương 4. ĐA CỘNG TUYẾN
Chương 5. PHƯƠNG SAI SAI SỐ THAY ĐỔI
Chương 6. TỰ TƯƠNG QUAN
Chương 7. HỒI QUY VỚI BIẾN GIẢ
Chương 8. LỰA CHỌN VÀ CHỈ ĐỊNH MÔ HÌNH

Chương 2. ƯỚC LƯỢNG VÀ KIỂM ĐỊNH MÔ HÌNH HỒI QUI ĐƠN

Chương 1. MỞ ĐẦU
Chương 2. ƯỚC LƯỢNG VÀ KIỂM ĐỊNH MÔ HÌNH HỒI QUY ĐƠN
Chương 3. MÔ HÌNH HỒI QUY BỘI
Chương 4. ĐA CỘNG TUYẾN
Chương 5. PHƯƠNG SAI SAI SỐ THAY ĐỔI
Chương 6. TỰ TƯƠNG QUAN
Chương 7. HỒI QUY VỚI BIẾN GIẢ
Chương 8. LỰA CHỌN VÀ CHỈ ĐỊNH MÔ HÌNH

Bài 1. Ước lượng mô hình

1. Ước lượng mô hình

Mô hình

- Mô hình hồi qui đơn (Simple regression)
 - một biến phụ thuộc (Y)
 - và một biến giải thích (X).
- Mô hình có dạng:

$$PRF : E(Y|X_i) = \beta_1 + \beta_2 X_i$$

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

2. Phương pháp bình phương nhỏ nhất

- Tìm $\hat{\beta}_1, \hat{\beta}_2$ sao cho

$$Q = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n e_i^2 \rightarrow \min$$

- Kết quả:

$$\hat{\beta}_2 = \frac{\overline{XY} - \bar{X}\bar{Y}}{\overline{X^2} - (\bar{X})^2}; \quad \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X}$$

- Đặt $x_i = X_i - \bar{X}$; $y_i = Y_i - \bar{Y}$; $\hat{y}_i = \hat{Y}_i - \bar{Y}$, ta được

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} \quad \hat{y}_i = \hat{\beta}_2 x_i$$

3. Các tính chất của SRF

- a. SRF đi qua điểm trung bình mẫu $(\bar{X}, \bar{Y})$
- b. Trung bình của các giá trị ước lượng bằng trung bình mẫu
 $\bar{\hat{Y}} = \bar{Y}$
- c. Tổng các phần dư bằng không $\sum_{i=1}^n e_i = 0$
- d. Các phần dư không tương quan với các giá trị của biến giải
thích: $\sum_{i=1}^n e_i X_i = 0$
- e. Các phần dư không tương quan với các giá trị ước lượng của
biến phụ thuộc Y: $\sum_{i=1}^n e_i \hat{Y}_i = 0$

4. Các giả thiết cơ bản của OLS

Một ước lượng sẽ dùng được khi nó là tốt nhất. Để ước lượng OLS là tốt nhất thì tổng thể phải thỏa mãn một số giả thiết sau:

- Giả thiết 1: Mô hình hồi quy có dạng tuyến tính đối với tham số.
- Giả thiết 2: Biến giải thích là phi ngẫu nhiên
- Giả thiết 3: Trung bình của các sai số ngẫu nhiên bằng 0

$$E(u_i) = 0 \quad \forall i$$

- Giả thiết 4: Phương sai sai số ngẫu nhiên bằng nhau

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

- Giả thiết 5: Các sai số ngẫu nhiên không tương quan

- Giả thiết 6: SSNN và biến giải thích không tương quan

$$\text{Cov}(u_i, X_i) = 0 \quad \forall i$$

- Giả thiết 7: Các giá trị của biến giải thích phải khác nhau càng nhiều càng tốt

$$\text{Var}(X) > 0$$

- Giả thiết 8: Kích thước mẫu phải lớn hơn số tham số cần ước lượng của mô hình: $n > k$.
- Giả thiết 9: Mô hình được chỉ định đúng.
- Giả thiết 10: Không có đa cộng tuyến giữa các biến giải thích của mô hình hồi quy bội.

Chương 1. MỞ ĐẦU
Chương 2. ƯỚC LƯỢNG VÀ KIỂM ĐỊNH MÔ HÌNH HỒI QUY ĐƠN
Chương 3. MÔ HÌNH HỒI QUY BỘI
Chương 4. ĐA CỘNG TUYẾN
Chương 5. PHƯƠNG SAI SAI SỐ THAY ĐỔI
Chương 6. TỰ TƯƠNG QUAN
Chương 7. HỒI QUY VỚI BIẾN GIẢ
Chương 8. LỰA CHỌN VÀ CHỈ ĐỊNH MÔ HÌNH

Định lý Gaus-Markov:

Nếu tổng thể thỏa mãn các giả thiết trên thì ước lượng OLS sẽ là ước lượng tuyến tính, không chệch, tốt nhất (trong số các ước lượng không chệch) của các tham số (Best Linear Unbiased Estimator - BLUE).

5. Các tính chất thống kê của các tham số ước lượng OLS

Các ước lượng là biến ngẫu nhiên tùy thuộc mẫu, nên có các tham số đặc trưng Kỳ vọng :

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

Phương sai :

$$Var(\hat{\beta}_1) = \frac{\sum_{i=1}^n X_i^2}{n \sum_{i=1}^n x_i^2} \sigma^2$$

$$Var(\hat{\beta}_2) = \frac{1}{\sum_{i=1}^n x_i^2} \sigma^2$$

6. Sự phù hợp của hàm hồi qui- Hệ số xác định R^2

Ký hiệu:

$$\sum_{i=1}^n y_i^2 = TSS$$

$$\sum_{i=1}^n \hat{y}_i^2 = ESS$$

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

$$TSS = ESS + RSS$$

TSS (Total Sum of Squares) : đo tổng biến động của biến phụ thuộc

ESS (Explained Sum of Squares): tổng biến động của biến phụ thuộc được giải thích bởi MH – biến giải thích.

RSS (Residual Sum of Squares) : tổng biến động của biến phụ thuộc được giải thích bởi các yếu tố nằm ngoài mô hình – Sai số ngẫu nhiên.

Đặt

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

gọi là hệ số xác định,

$$0 \leq R^2 \leq 1.$$

Ý nghĩa: Hệ số xác định R^2 là tỉ lệ (hoặc tỉ lệ %) sự biến động của biến phụ thuộc được giải thích bởi biến giải thích (theo mô hình, trong mẫu).

7. Phân phối của sai số ngẫu nhiên

Giả thiết 11: Các SSNN u_i có phân phối chuẩn.

Cơ sở của giả thiết này là:

+ Do ui thường là sự tổng hợp của một số lớn các nhân tố ngẫu nhiên độc lập và ảnh hưởng bề đều như nhau nên theo hệ quả của định lý giới hạn trung tâm thì có thể xem là ui phân phối chuẩn.

+ Phân phối chuẩn chỉ có hai tham số là μ và σ^2 nên dễ sử dụng.

+ Phân phối chuẩn có tính chất là nếu u_i phân phối chuẩn thì mọi hàm tuyến tính của nó cũng phân phối chuẩn.

+ Phân phối chuẩn có tính chất là tính độc lập và không tương quan là đồng nhất.

Chương 1. MỞ ĐẦU	Chương 2. ƯỚC LƯỢNG VÀ KIỂM ĐỊNH MÔ HÌNH HỒI QUY ĐƠN
Chương 3. MÔ HÌNH HỒI QUY BỘI	Chương 4. ĐA CỘNG TUYẾN
Chương 5. PHƯƠNG SAI SAI SỐ THAY ĐỔI	Chương 6. TỰ TƯƠNG QUAN
Chương 7. HỒI QUY VỚI BIẾN GIẢ	Chương 8. LỰA CHỌN VÀ CHỈ ĐỊNH MÔ HÌNH

8. Các phân phối liên quan tới các ước lượng OLS.

9. Ước lượng khoảng

9.1 Ước lượng khoảng cho các hệ số hồi quy

Với độ tin cậy $1 - \alpha$ cho trước:

Khoảng tin cậy của các hệ số hồi quy β_1 và β_2

Khoảng tin cậy tổng quát:

$$\hat{\beta}_j - SE(\hat{\beta}_j)t_{\alpha/2}(n-2) < \beta_j < \hat{\beta}_j + SE(\hat{\beta}_j)t_{\alpha/2}(n-2)$$

Khoảng tin cậy đối xứng:

$$\hat{\beta}_j - SE(\hat{\beta}_j)t_{\alpha/2}(n-2) < \beta_j < \hat{\beta}_j + SE(\hat{\beta}_j)t_{\alpha/2}(n-2)$$

Khoảng tin cậy một phía

Khoảng tin cậy bên phải:

$$\hat{\beta}_j - SE(\hat{\beta}_j)t_{\alpha}(n-2) < \beta_j$$

Khoảng tin cậy bên trái:

$$\beta_j < \hat{\beta}_j + SE(\hat{\beta}_j)t_{\alpha}(n-2)$$

Khoảng tin cậy cho sai số ngẫu nhiên:

Khoảng tin cậy tổng quát:

$$\frac{\hat{\sigma}^2(n-2)}{\chi_{\alpha/2}^2(n-2)} < \sigma^2 < \frac{\hat{\sigma}^2(n-2)}{\chi_{\alpha/1}^2(n-2)}$$

Khoảng tin cậy hai phía:

$$\frac{\hat{\sigma}^2(n-2)}{\chi_{\alpha/2}^2(n-2)} < \sigma^2 < \frac{\hat{\sigma}^2(n-2)}{\chi_{1-\alpha/2}^2(n-2)}$$

Khoảng tin cậy 1 phía

Khoảng tin cậy bên phải:

$$\frac{\hat{\sigma}^2(n-2)}{\chi_{\alpha}^2(n-2)} < \sigma^2$$

Khoảng tin cậy bên trái:

$$\sigma^2 < \frac{\hat{\sigma}^2(n-2)}{\chi_{1-\alpha}^2(n-2)}$$

Chương 1. MỞ ĐẦU	
Chương 2. ƯỚC LƯỢNG VÀ KIỂM ĐỊNH MÔ HÌNH HỒI QUY ĐƠN	
Chương 3. MÔ HÌNH HỒI QUY BỘI	
Chương 4. ĐA CỘNG TUYẾN	
Chương 5. PHƯƠNG SAI SAI SỐ THAY ĐỔI	
Chương 6. TỰ TƯƠNG QUAN	
Chương 7. HỒI QUY VỚI BIẾN GIẢ	
Chương 8. LỰA CHỌN VÀ CHỈ ĐỊNH MÔ HÌNH	

Bài 2. Kiểm định giả thiết

1. Kiểm định giả thiết đối với hệ số hồi quy

Kiểm định hai phía

Với mức ý nghĩa α cho trước, kiểm định mối quan hệ thứ tự của hệ số với các số thực cho trước Cặp giả thiết

$$\begin{cases} H_0 : \beta_j = \beta_j^* \\ H_1 : \beta_j \neq \beta_j^* \end{cases} \quad \text{với } j = 1, 2.$$

Thông kê được chọn:

$$T_{qs} = \frac{\hat{\beta}_j - \beta_j^*}{SE(\hat{\beta}_j)}$$

Tiêu chuẩn kiểm định: Nếu $T_{qs} > t_{\alpha/2}(n-2)$ thì bác bỏ H_0 ,

Kiểm định 1 phía

Cặp giả thiết

$$\begin{cases} H_0 : \beta_j = \beta_j^* \\ H_1 : \beta_j > \beta_j^* \end{cases}$$

với $j = 1, 2$.

Thông kê được chọn:

$$T_{qs} = \frac{\hat{\beta}_j - \beta_j^*}{SE(\hat{\beta}_j)}$$

Tiêu chuẩn kiểm định: Nếu $T_{qs} > t_{\alpha}(n-2)$ thì bác bỏ H_0 và chấp nhận H_1 , ngược lại, chưa có cơ sở bác bỏ H_0 .

Kiểm định 1 phía

Cặp giả thiết

$$\begin{cases} H_0 : \beta_j = \beta_j^* \\ H_1 : \beta_j < \beta_j^* \end{cases}$$

với $j = 1, 2$.

Thông kê được chọn:

$$T_{qs} = \frac{\hat{\beta}_j - \beta_j^*}{SE(\hat{\beta}_j)}$$

Tiêu chuẩn kiểm định: Nếu $T_{qs} < -t_{\alpha}(n-2)$ thì bác bỏ H_0 và chấp nhận H_1 , ngược lại, chưa có cơ sở bác bỏ H_0 .

Trường hợp đặc biệt: Kiểm định xem hệ số hồi quy có ý nghĩa thống kê hay không?

Cặp giả thiết

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

với $j = 1, 2$.

Thống kê được chọn:

$$T_{qs} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$$

Kiểm định bằng P-value

Với kiểm định bên phải: $P - value = P(T > T_{qs})$

Với kiểm định bên trái: $P - value = P(T < T_{qs})$

Với kiểm định hai phía: $P - value = 2P(T > |T_{qs}|)$

Quy tắc kiểm định

Nếu cho trước α thì quy tắc kết luận như sau:

Nếu $P - value < \alpha$ thì bác bỏ H_0

Nếu $P - value > \alpha$ thì thừa nhận H_0

2. Kiểm định cho phương sai của sai số

Kiểm định 2 phía

Cặp giả thiết:

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases}$$

Thông kê kiểm định:

$$\chi_{qs}^2 = \frac{(n-2)\hat{\sigma}^2}{\sigma_0^2}$$

Tiêu chuẩn kiểm định:

Nếu $\chi_{qs}^2 < \chi_{1-\alpha/2}^2$ hoặc $\chi_{qs}^2 > \chi_{\alpha/2}^2$ thì bác bỏ H_0

Kiểm định 1 phía

Cặp giả thiết:

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases}$$

Thống kê kiểm định:

$$\chi_{qs}^2 = \frac{(n-2)\hat{\sigma}^2}{\sigma_0^2}$$

Tiêu chuẩn kiểm định:

Nếu $\chi_{qs}^2 > \chi_{\alpha}^2$ thì bác bỏ H_0

Kiểm định 1 phía

Cặp giả thiết:

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{cases}$$

Thống kê kiểm định:

$$\chi_{qs}^2 = \frac{(n-2)\hat{\sigma}^2}{\sigma_0^2}$$

Tiêu chuẩn kiểm định:

Nếu $\chi_{qs}^2 < \chi_{1-\alpha}^2$ thì bác bỏ H_0

Kiểm định về sự phù hợp của mô hình.

$$\text{Cặp giả thiết } \begin{cases} H_0 : R^2 = 0 \\ H_1 : R^2 \neq 0 \end{cases}$$
$$\iff \begin{cases} H_0 : \beta_2 = 0 \\ H_1 : \beta_2 \neq 0 \end{cases} \quad \text{Kiểm định F:}$$

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

- Nếu $F_{qs} > F_{\alpha}(1; n - 2)$ thì bác bỏ H_0 : biến giải thích giải thích được cho sự biến động của biến phụ thuộc, hàm hồi qui được gọi là phù hợp.
- Ngược lại, Y không phụ thuộc vào biến giải thích, hàm hồi qui không phù hợp.
- Vì hai cặp giả thiết tương đương, kiểm định F tương đương kiểm định T: $F_{qs} = (T_{qs})^2$.
- Kiểm định F nói trên cũng có thể tiến hành bằng phương pháp P-value.

Dự báo

Là ước lượng khoảng cho giá trị trung bình và cá biệt của biến phụ thuộc khi biến giải thích nhận giá trị xác định $X = X_0$

Dự báo giá trị trung bình

Khoảng tin cậy đối xứng:

$$\hat{Y}_0 - SE(\hat{Y}_0)t_{\alpha/2}(n-2) < E(Y/X_0) < \hat{Y}_0 + SE(\hat{Y}_0)t_{\alpha/2}(n-2)$$

Khoảng tin cậy bên phải:

$$\hat{Y}_0 - SE(\hat{Y}_0)t_{\alpha}(n-2) < E(Y/X_0)$$

Khoảng tin cậy bên trái:

$$E(Y/X_0) < \hat{Y}_0 + SE(\hat{Y}_0)t_{\alpha}(n-2)$$

Trong đó $\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$ và $SE(\hat{Y}_0) = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}}$

Dự báo giá trị cá biệt

Khoảng tin cậy đối xứng:

$$\hat{Y}_0 - SE(\hat{Y}_0 - Y_0)t_{\alpha/2}(n-2) < E(Y/X_0) < \hat{Y}_0 + SE(\hat{Y}_0 - Y_0)t_{\alpha/2}(n-2)$$

Khoảng tin cậy bên phải:

$$\hat{Y}_0 - SE(\hat{Y}_0 - Y_0)t_{\alpha}(n-2) < E(Y/X_0)$$

Khoảng tin cậy bên trái:

$$E(Y/X_0) < \hat{Y}_0 + SE(\hat{Y}_0 - Y_0)t_{\alpha}(n-2)$$

Trong đó $\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$ và $SE(\hat{Y}_0 - Y_0) = \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}}$

Ví dụ 1

Giả sử có số liệu thống kê về lãi suất ngân hàng (X - %/năm), tổng vốn đầu tư (Y - tỉ đồng) trên địa bàn của tỉnh A qua 10 năm liên tiếp như sau:

X	7.0	6.5	6.5	6.0	6.0	6.0	5.5	5.5	5.0	4.5
Y	28	32	30	34	32	35	40	42	48	50

Yêu cầu:

- 1) Lập mô hình hồi quy tuyến tính mô tả quan hệ giữa tổng vốn đầu tư và lãi suất ngân hàng. nêu ý nghĩa kinh tế của hệ số hồi quy ước lượng được.
- 2) Kiểm định giả thiết hệ số hồi quy của X trong hàm hồi quy tổng thể bằng 0 với mức ý nghĩa 2% và nêu ý nghĩa kết quả.
- 3) Với mức ý nghĩa 5%, có thể xem hệ số góc của mô hình hồi quy bằng - 11 được không?
- 4) Xác định khoảng tin cậy của hệ số chặn và hệ số góc với độ tin cậy 90%.

- 5) Đánh giá mức độ phù hợp của mô hình hồi quy ước lượng được với mức ý nghĩa 1%.
- 6) Xác định khoảng tin cậy cho phương sai của nhiễu với độ tin cậy 95%.
- 7) Có tài liệu cho rằng: phương sai của nhiễu là 2, có nên tin số liệu này không với mức ý nghĩa 10%.
- 8) Dự báo giá trị cầu tổng vốn đầu tư (Y - tỉ đồng) trên địa bàn của tỉnh A khi lãi suất là 6.8 % với độ tin cậy 95%.

Kết quả

Sample(adjusted): 1 10				
Included observations: 10 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	94.55224	5.277127	17.91737	0.0000
X	-9.820896	0.895522	-10.96667	0.0000
R-squared	0.937630	Mean dependent var		37.10000
Adjusted R-squared	0.929834	S.D. dependent var		7.578478
S.E. of regression	2.007449	Akaike info criterion		4.408463
Sum squared resid	32.23881	Schwarz criterion		4.468980
Log likelihood	-20.04231	F-statistic		120.2678
Durbin-Watson stat	1.002633	Prob(F-statistic)		0.000004

Ví dụ 2

Giả sử ta có bảng số liệu dưới đây mô tả mức vốn huy động được (biến Y - đơn vị : triệu /năm) và lãi suất tiền gửi (biến X – đơn vị : % năm) vào năm 2000 tại các đơn vị tín dụng trong một thành phố. Cho rằng Y có dạng tuyến tính đối với tham số lẫn biến số X :

X	11.9	9.4	7.5	4.0	11.3	6.3	2.2	10.3	7.6
Y	7.2	4	3.1	1.6	4.8	5.1	2.0	6.6	4.4

- (1) Hãy thiết lập bảng tính và ước lượng hàm hồi quy
- (2) Ý nghĩa kinh tế các hệ số hàm quy
- (3) Phương sai và sai số chuẩn của các hệ số hồi quy
- (4) Tính khoảng tin cậy 90% của các hệ số hồi quy và cho biết ý nghĩa của nó
- (5) Từ mẫu số liệu trên, có ý kiến cho rằng lãi suất không ảnh hưởng gì đến mức vốn huy động được từ khách hàng, bạn hãy cho biết nhận xét trên đúng hay không đúng với mức ý nghĩa 5%
- (6) Hãy dự báo mức vốn huy động được trung bình và mức vốn huy động cá biệt nếu lãi suất 4% với độ tin cậy 90%

Ví dụ 3

Xét ví dụ về tiêu dùng và thu nhập

Included observations:				
12				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-231.7951	94.52751	-2.452144	0.0341
X	0.719433	0.021750	33.07810	0.0000
R-squared	0.990943	Mean dependent var		2880.600
Adjusted R-squared	0.990038	S.D. dependent var		314.4417
S.E. of regression	31.38488	Akaike info criterion		9.881541
Sum squared resid	9850.106	Schwarz criterion		9.962359
Log likelihood	-57.28925	F-statistic		1094.160
Durbin-Watson stat	1.284182	Prob(F-statistic)		0.000000

1. Kiểm định sự phù hợp của hàm hồi quy.
2. Nêu ý nghĩa kinh tế của β_2
3. Kiểm định về nhận xét của Keynes về quy luật của tiêu dùng.

Bài tập 1

Một công ty sản xuất một loại sản phẩm. Lượng sản phẩm (Q) mà công ty sản xuất phụ thuộc vào giá sản phẩm này trên thị trường (P). Dựa trên số liệu trong 20 tháng từ tháng 1 năm 2006 đến tháng 8 năm 2007, người ta ước lượng được mô hình dưới đây. Cho 5%; Q tính bằng 1000 sản phẩm, P tính bằng nghìn đồng.

Included observations: 20				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1170.608	270.8481		0.0004
P	-135.7035	51.41326		0.0167
R-squared		Mean dependent var		460.2000
Adjusted R-squared	0.238989	S.D. dependent var		155.3125
S.E. of regression	135.4883	Akaike info criterion		12.75029
Sum squared resid	330427.5	Schwarz criterion		12.84986
Durbin-Watson stat	2.794879	Prob(F-statistic)		0.016655

- Viết hàm hồi quy tổng thể, hàm hồi quy mẫu.
- Hệ số chặn và hệ số góc được ước lượng bằng bao nhiêu?
- Các hệ số thu được từ hàm hồi quy mẫu có phù hợp lý thuyết kinh tế không?
- Các hệ số của mô hình có ý nghĩa thống kê không? Con số [prob] cho biết điều gì?
- Có thể nói rằng khi giá thay đổi thì lượng bán ra thay đổi không?
- Hệ số xác định đo độ phù hợp của hàm hồi quy bằng bao nhiêu, giá trị đó có ý nghĩa gì?
- Hàm có thể coi là phù hợp không?
- Tìm ước lượng điểm cho phương sai yếu tố ngẫu nhiên

- i. Tổng bình phương phần dư bằng bao nhiêu?
- j. TSS và ESS bằng bao nhiêu?
- k. Tìm khoảng tin cậy cho hệ số chặn và hệ số góc của mô hình
- l. Khi giá tăng thêm 1 nghìn đồng thì lượng cung thay đổi trung bình trong khoảng nào?
- m. Khi giá tăng 1 nghìn thì lượng cung tăng tối đa bao nhiêu?
- n. Có thể nói khi giá giảm 1 nghìn thì lượng cung giảm 0,5 đơn vị được không?
- o. Tìm một ước lượng điểm cho lượng cung khi giá là 10,5 nghìn đồng.
- p. Tìm lượng cung trung bình và cá biệt khi giá là 10,55 nghìn đồng.

Bài tập 2

Một cơ quan nghiên cứu mối quan hệ giữa số đơn vị sản phẩm và các yếu tố đầu vào của quá trình sản xuất ở một số cơ sở sản xuất đã đưa ra những mô hình hồi quy. Lúc đầu người nghiên cứu chú trọng vào quản lý nguồn nhân lực nên đưa ra mô hình sau: Với Q là sản lượng, L là lao động (người). Cho $\alpha = 5\%$

Included observations: 16				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-48643.96	58633.59		
L	62.24053	24.66142		
R-squared		Mean dependent var		98446.75
Adjusted R-squared		S.D. dependent var		29910.63
S.E. of regression		Akaike info criterion		23.26029
Sum squared resid	9.22E+09	Schwarz criterion		23.35687
Log likelihood	-184.0823	F-statistic		6.369576
Durbin-Watson stat	0.182702	Prob(F-statistic)		0.024321

- a. Viết hàm hồi quy tổng thể, hàm số đo và các tham số có ý nghĩa như thế nào?
- b. Viết hàm hồi quy mẫu. Các hệ số của hàm hồi quy mẫu có phù hợp lý thuyết kinh tế không?
- c. Theo lý thuyết thì khi không có lao động sẽ không có sản lượng, nhưng trong hàm hồi quy mẫu thì khi không có lao động ước lượng điểm mức sản lượng lại không bằng không. Trên thực tế giá trị đó có thể coi là bằng 0 hay không?
- d. Hệ số góc của mô hình có ý nghĩa thống kê không?
- e. Hệ số xác định bằng bao nhiêu %, giá trị đó có ý nghĩa như thế nào?
- f. Có thể nói hàm hồi quy phù hợp không?
- g. Tìm ước lượng điểm và khoảng phương sai yếu tố ngẫu nhiên.

- h. RSS, TSS, ESS bằng bao nhiêu?
- i. Tìm khoảng tin cậy cho hệ số chặn của mô hình.
- j. Khi doanh nghiệp thêm 1 lao động thì sản lượng tăng trong khoảng nào?
- k. Khi giảm bớt 1 lao động thì sản lượng giảm tối đa bao nhiêu đơn vị?
- l. Có thể cho rằng khi bớt 1 lao động thì sản lượng giảm 30 đơn vị được không?
- m. Nếu tăng 1 lao động thì sản lượng tăng nhiều hơn 20 đơn vị có đúng không?
- n. Tìm ước lượng điểm mức sản lượng với doanh nghiệp có 30 lao động.
- o. Tìm mức sản lượng trung bình và cá biệt khi doanh nghiệp có 30 lao động?

Bài tập 3

Giả sử ta có các số liệu về tỷ lệ tăng trưởng GRV và lạm phát INFLV của Việt Nam tính theo phần trăm trong các năm từ 1980 tới 1996. Xét mô hình $GRV_t = b_1 + b_2 INFLV_t + v_t$, sử dụng thủ tục OLS trong Eviews để ước lượng phương trình này ta có các kết quả thu được như sau:

Included observations: 17				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.147850	0.878923		
INFLV	-0.008207	0.005233		
R-squared		Mean dependent var		6.290000
Adjusted R-squared		S.D. dependent var		2.963431
S.E. of regression		Akaike info criterion		5.033368
Sum squared resid	120.7123	Schwarz criterion		5.131393
Log likelihood	-40.78363	F-statistic		2.460203
Durbin-Watson stat	0.767860	Prob(F-statistic)		0.137614

- a. Các kết quả thu được có ý nghĩa gì về mặt kinh tế không?
 b. Ước lượng mô hình sau

$$\ln GRV_t = g_1 + g_2 \ln GRV_t + w_t$$

ta có kết quả sau:

Included observations: 16 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.598414	0.254701		
LOG(INFLV)	-0.191480	0.061184		
R-squared		Mean dependent var		1.843671
Adjusted R-squared		S.D. dependent var		0.412754
S.E. of regression		Akaike info criterion		0.723145
Sum squared resid	1.503592	Schwarz criterion		0.819719
Log likelihood	-3.785164	F-statistic		9.794281
Durbin-Watson stat	1.884796	Prob(F-statistic)		0.007385

So sánh với kết quả ước lượng theo mô hình trước, theo bạn mô hình nào cho kết quả tốt hơn. Tại sao?

c. Tỷ lệ tăng trưởng khi lạm phát bằng 3% là bao nhiêu theo mô hình ở câu b?

- Chương 1. MỞ ĐẦU
- Chương 2. ƯỚC LƯỢNG VÀ KIỂM ĐỊNH MÔ HÌNH HỒI QUY ĐƠN
- Chương 3. MÔ HÌNH HỒI QUY BỘI**
- Chương 4. ĐA CỘNG TUYẾN
- Chương 5. PHƯƠNG SAI SAI SỐ THAY ĐỔI
- Chương 6. TỰ TƯƠNG QUAN
- Chương 7. HỒI QUY VỚI BIẾN GIẢ
- Chương 8. LỰA CHỌN VÀ CHỈ ĐỊNH MÔ HÌNH

Chương 3. MÔ HÌNH HỒI QUY BỘI

1. Mô hình hồi quy tổng quát k biến - Dạng ma trận của mô hình

Mô hình

Mô hình hồi qui trong đó biến phụ thuộc Y phụ thuộc vào $k - 1$ biến giải thích $X_2, \dots, X_k$ có dạng

$$PRF : E(Y_i) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} \quad (1)$$

$$PRM : Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i \quad (2)$$

trong đó: β_1 gọi là hệ số chặn

β_j ($j = 2, k$) gọi là các hệ số góc riêng phần.

Với mẫu $W = (X_{2i}, X_{3i}, \dots, X_{ki}, Y_i); \quad i = 1, \dots, n,$

$$SRF : \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} \quad (3)$$

$$SRM : Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} + e_i \quad (4)$$

2.2 Dạng ma trận

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_{n-1} \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{21} & \dots & X_{k1} \\ 1 & X_{22} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots \\ 1 & X_{2n-1} & \dots & X_{kn-1} \\ 1 & X_{2n} & \dots & X_{kn} \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_{n-1} \\ u_n \end{bmatrix}$$

$$Y_{(n \times 1)} = X_{(n \times k)}\beta_{(k \times 1)} + U_{(n \times 1)}$$

$$E(Y) = X\beta$$

$$Y = X\hat{\beta} + e$$

Phương pháp bình phương nhỏ nhất

Tìm $\hat{\beta}$ sao cho $\sum_{i=1}^n e_i^2 = e'e \rightarrow \min$

$$(Y - X\hat{\beta})'(Y - X\hat{\beta}) \rightarrow \min \iff X'X\hat{\beta} = X'Y$$

Nếu tồn tại $(X'X)^{-1}$ thì $\hat{\beta} = (X'X)^{-1}X'Y$

Khi đó $\hat{\beta} = (X'X)^{-1}X'Y$ là ước lượng tuyến tính không chệch tốt nhất của β

Các tham số của ước lượng

Kì vọng: $E(\hat{\beta}) = \beta$

Phương sai – hiệp phương sai

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

Với σ^2 được ước lượng bởi $\sigma^2 = \frac{\sum_{i=1}^n e_i^2}{n-k}$

Sự phù hợp của hàm hồi qui

Hệ số xác định bội.

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

Hệ số xác định bội hiệu chỉnh.

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

$\bar{R}^2$ có các tính chất sau:

- $\bar{R}^2$ có thể nhận giá trị âm.
- Khi số biến giải thích của mô hình tăng lên thì $\bar{R}^2$ tăng chậm hơn R^2 .
- $\bar{R}^2 \leq R^2 \leq 1$

Tính chất này được dùng làm căn cứ xem xét việc đưa thêm biến giải thích vào mô hình. Khi đưa thêm biến vào mô hình mà $\bar{R}^2$ còn tăng hoặc khi giá trị t của kiểm định về sự bằng không của hệ số hồi quy tương ứng với biến đưa thêm còn lớn hơn 1 thì việc đưa thêm biến còn hợp lý.

Suy diễn thống kê

3.1 Ước lượng khoảng

- Ước lượng khoảng 2 phía cho hệ số hồi quy
- Ước lượng khoảng 1 phía cho hệ số hồi quy
- Ước lượng khoảng 2 phía cho phương sai σ^2
- Ước lượng khoảng 1 phía cho phương sai σ^2

Kiểm định giả thiết

Kiểm định T cho các hệ số hồi quy

Kiểm định χ^2 cho σ^2

Kiểm định sự phù hợp của hàm hồi quy

$$\text{Cặp giả thiết } \begin{cases} H_0 : R^2 = 0 \\ H_1 : R^2 \neq 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} H_0 : \beta_2 = \beta_3 = \dots = \beta_k = 0 \\ H_1 : \exists j : \beta_j \neq 0 \end{cases} \quad \text{Kiểm định F:}$$

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

- Nếu $F_{qs} > F_{\alpha}(k - 1; n - 2)$ thì bác bỏ H_0 : biến giải thích giải thích được cho sự biến động của biến phụ thuộc, hàm hồi qui được gọi là phù hợp.
- Ngược lại, Y không phụ thuộc vào biến giải thích, hàm hồi qui không phù hợp.

Kiểm định thu hẹp hồi qui (Kiểm định Wald):

Giả sử rằng chúng ta có hồi quy (Mô hình UR)

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i$$

Bây giờ chúng ta kiểm định giả thiết tổng quát sau:

$$H_0 : \beta_{k-m+1} = \beta_{k-m+2} = \dots = \beta_k = 0$$

$$H_1 : \exists j : \beta_j \neq 0, \quad j = k - m + 1, \dots, k$$

H_0 : Tất cả m biến giải thích không giải thích cho Y

H_1 : Ít nhất một biến giải thích trong các biến $X_{k-m+1}, \dots, X_k$ có giải thích cho Y

Nếu H_0 đúng thì hàm hồi đã cho có dạng:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_{k-m} X_{k-m} + u_i$$

(được gọi là hàm hồi quy thu hẹp hay hàm hồi quy có điều kiện ràng buộc: kí hiệu là mô hình R).

Ký hiệu

e_R : vectơ phần dư từ hàm hồi có điều kiện ràng buộc

e_{UR} : Vectơ phần dư từ hàm hồi quy ban đầu (không điều kiện ràng buộc)

m : Số biến bị loại ra khỏi mô hình ban đầu (số điều kiện ràng buộc)

n : Số quan sát

Khi đó tiêu chuẩn kiểm định giả thiết H_0 ;

$$\frac{(e'_R e_R - e'_{UR} e_{UR}) / m}{e'_{UR} e_{UR} / (n - k)} \sim F(m, (n - k))$$

Nếu $F > F_\alpha(m, (n - k))$ thì giả thiết H_0 bị bác bỏ.

Đôi khi ta dùng công thức sau đây:

$$\frac{(e'_R e_R - e'_{UR} e_{UR})/m}{e'_{UR} e_{UR}/(n-k)} = \frac{(TSS - ESS_R) - (TSS - ESS_{UR}/m)}{(TSS - ESS_R)/(n-k)}$$

$$= \frac{(ESS_{UR} - ESS_R)/m}{(TSS - ESS_{UR})/(n-k)}.$$

chia cả hai vế và mẫu số cho TSS, ta được công thức rút gọn:

$$F = \frac{(R^2_{UR} - R^2_R)/m}{(1 - R^2_{UR})/(n-k)} \sim F(m, (n-k)).$$

Dự báo

Chúng ta có thể sử dụng mô hình vào dự báo: dự báo giá trị trung bình và dự báo giá trị cá biệt. Cho

$$X_0 = \begin{bmatrix} 1 \\ X_{20} \\ X_{30} \\ \cdot \\ \cdot \\ X_{k0} \end{bmatrix}$$

Dự báo giá trị trung bình: $E(Y|X_0)$

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

Với $X = X_0$ ta có

$$\begin{aligned}(\hat{Y} | X_0) &= X_0' \hat{\beta} \Rightarrow \text{var}(\hat{Y}_0 | X_0') = X_0' \text{Var}(\hat{\beta}) X_0 \\ \text{Var}(\hat{Y}_0 | X_0) &= \sigma^2 X_0' (X'X)^{-1} X_0,\end{aligned}$$

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

Nhưng σ^2 chưa biết nên phải dùng ước lượng không chệch lệch $\hat{\sigma}^2$ của nó:

$$\hat{Var}(\hat{Y}|X_0) = \hat{\sigma}^2 X_0'(X'X)^{-1}X_0$$

$$SE(\hat{Y}_0|X_0) = \sqrt{\hat{\sigma}^2 X_0'(X'X)^{-1}X_0}$$

$$\hat{Y}_0 - t_{\alpha/2(n-k)} SE(\hat{Y}_0|X_0) \leq \hat{Y}_0 + t_{\alpha/2(n-k)} SE(\hat{Y}_0|X_0)$$

Dự báo giá trị cá biệt

$$Y_i = X_i' \hat{\beta} + e_i \Rightarrow \text{Var}(Y_0|X_0) = \text{Var}(X_0' \hat{\beta}) + \sigma^2$$
$$\hat{\text{Var}}(Y_0|X_0) = \hat{\sigma}^2 [1 + X_0'(X'X)^{-1}X_0]$$
$$SE(Y_0|X_0) = \sqrt{\hat{\text{Var}}(Y_0|X_0)}$$

$$\hat{Y}_0 - t_{\alpha/2}(n-k) SE(Y_0|X_0) \leq (Y_0|X_0) \leq \hat{Y}_0 + t_{\alpha/2}(n-k) SE(Y_0|X_0)$$

Với việc trình bày hồi quy bằng ngôn ngữ ma trận đã cung cấp cho chúng ta một công nghệ mà nhờ đó có thể sử dụng kỹ thuật tính toán, tự động hoá toàn bộ quá trình tính toán, phân tích và dự đoán.

Ví dụ

Bảng sau đây cho Tỷ lệ lạm phát $Y(\%)$, Tỷ lệ thất nghiệp $X_2(\%)$ và Tỷ lệ lạm phát kỳ vọng $X_3(\%)$ của Mỹ giai đoạn 1970- 1982:

- a. Hồi quy Y với X_2 và cho nhận xét.
- b. Hồi quy Y với X_2 và X_3 và so sánh với kết quả thu được ở phần a.
- c. Hãy phân tích kết quả thu được ở mô hình 3 biến.

Included observations: 13				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.127172	4.285283	1.429817	0.1806
X2	0.244934	0.630456	0.388502	0.7051
R-squared	0.013536	Mean dependent var		7.756923
Adjusted R-squared	-0.076143	S.D. dependent var		3.041892
S.E. of regression	3.155577	Akaike info criterion		5.276858
Sum squared resid	109.5343	Schwarz criterion		5.363773
Log likelihood	-32.29958	F-statistic		0.150934
Durbin-Watson stat	0.969568	Prob(F-statistic)		0.705058

Included observations: 13

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.193357	1.594789	A	0.0011
X2	-1.392472	0.305018	B	0.0010
X3	1.470032	0.175786	C	0.0000
R-squared	D	Mean dependent var		7.756923
Adjusted R-squared	0.851907	S.D. dependent var		3.041892
S.E. of regression	1.170605	Akaike info criterion		3.352092
Sum squared resid	E	Schwarz criterion		3.482465
Log likelihood	-18.78860	F-statistic	F	
Durbin-Watson stat	2.225465	Prob(F-statistic)		0.000029

Bài tập 1

Bảng dưới đây cho: Y - Thu nhập/ đầu người(Y) tính bằng 100 USD; X_2 - Tỷ lệ phần trăm của lao động nông nghiệp; X_3 - Số năm trung bình đào tạo đối với những người trên 25 tuổi của 1 địa phương. Cho $\alpha = 5\%$.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Y	18	24	24	21	21	36	27	24	27	30	30	33	27	30	33
X_2	9	10	8	7	10	4	5	5	6	8	7	4	9	5	8
X_3	8	13	11	10	12	16	10	10	12	14	12	16	14	10	12

Giả thiết rằng $E(Y/X_1, X_2) = \beta_1 + \beta_1 X_2 + \beta_2 X_3$. Kết quả hồi quy mẫu được cho dưới bảng sau:

Dependent Variable: Y				
Included observations: 15				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	18.60894	5.586760	3.330900	0.0060
X2	-1.128492	0.398171	-2.834186	0.0151
X3	1.357542	0.358533	3.786374	0.0026
R-squared	0.693203	Mean dependent var	27.00000	
Adjusted R-squared	0.642070	S.D. dependent var	5.070926	
S.E. of regression	3.033795	Akaike info criterion	5.234362	
Sum squared resid	110.4469	Schwarz criterion	5.375972	
Log likelihood	-36.25771	F-statistic	13.55690	
Durbin-Watson stat	0.946397	Prob(F-statistic)	0.000834	

- Viết hàm hồi quy tổng thể, mô hình hồi quy tổng thể và hàm hồi quy mẫu.
- Hãy kiểm định sự bằng không của từng hệ số hồi quy và ý nghĩa rút ra từ các kiểm định ấy.
- Hãy tìm khoảng tin cậy đối xứng cho các hệ số hồi quy. Hãy cho biết ý nghĩa của kết quả tìm được
- Phải chăng cả hai yếu tố “ Tỷ lệ lao động nông nghiệp” và “ số năm được đào tạo” đều không cùng ảnh hưởng đến Thu nhập theo đầu người.
- Tìm khoảng tin cậy 95% cho phương sai của sai số.
- Kiểm định sự phù hợp của mô hình.
- Nếu một người học đại học 4 năm và cao học 2 năm trong một địa phương có 5% lao động nông nghiệp thì thu nhập của người đó nằm trong khoảng nào?

	C	X2	X3
C	31.2118834722	-1.41829348855	-1.72251958428
X2	-1.41829348855	0.158540359747	0.0257092475266
X3	-1.72251958428	0.0257092475266	0.128546237633

Bài tập 2

Xét mô hình hồi quy biến sản lượng (Q) theo Lao động(L: người) và biến K là vốn(triệu đồng) $Q = \beta_1 + \beta_2 K + \beta_3 L + u$. Cho 5%. Kết quả ước lượng mô hình trên phần mềm Eviews như sau:

Included observations: 16				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
K	2615.988	424.8504	6.157434	0.0000
L	6.117142	15.82129	0.386640	0.7053
C	-22336.50	31041.66	-0.719565	0.4845
R-squared		Mean dependent var	98446.75	
Adjusted R-squared	0.797512	S.D. dependent var	29910.63	
S.E. of regression	13459.40	Akaike info criterion	22.02010	
Sum squared resid	2.36E+09	Schwarz criterion	22.16496	
Log likelihood	-173.1608	F-statistic		
Durbin-Watson stat	0.337815	Prob(F-statistic)		

- a. Viết hàm hồi quy tổng thể, hồi quy mẫu.
- b. Các ước lượng nhận được có phù hợp về lý thuyết không?
- c. Tìm ước lượng điểm mức sản lượng doanh nghiệp có 2000 lao động, nguồn vốn 300 triệu đồng.
- d. Các giá trị ước lượng có ý nghĩa thống kê không?
- e. Tính hệ số xác định bội bằng các cách.
- f. Phải chăng các biến độc lập không giải thích được cho sự biến động của sản lượng?
- g. Có thể nói vốn, lao động cùng tác động thuận chiều đến sản lượng không?

- h. Khi lao động không đổi, nếu thêm vốn 1 triệu thì sản lượng tăng trong khoảng nào? ít hơn 20 đơn vị được không?
- i. Có thể nói khi lao động không đổi, tăng vốn thêm 1 triệu thì sản lượng tăng ít hơn 10 đơn vị được không?
- j. Nguồn vốn không đổi, thêm 1 lao động thì sản lượng tăng có bằng 20 đơn vị không?
- k. Dùng kiểm định thu hẹp hồi quy để đánh giá việc đưa thêm biến K vào mô hình hay không nếu biết với mô hình Q phụ thuộc L có hệ số chặn hệ số xác định bằng 0.312700 và RSS bằng 9.22×10^9

Bài tập 3

Với bài tập 2, một người đưa ra dạng khác của mô hình và hồi quy được kết quả sau, với LQ, LK, LL là logarit cơ số tự nhiên của các biến tương ứng.

Dependent Variable: LQ				
Included observations: 16				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.529063	3.183826	2.050697	0.0610
LL	0.187768	0.460393	0.407842	0.6900
LK	0.944096	0.191086	4.940674	0.0003
R-squared		Mean dependent var		11.45483
Adjusted R-squared	0.738277	S.D. dependent var		0.299201
S.E. of regression	0.153068	Akaike info criterion		-0.748512
Sum squared resid	0.304586	Schwarz criterion		-0.603652
Log likelihood	8.988096	F-statistic		
Durbin-Watson stat	0.306011	Prob(F-statistic)		

Cho biết ma trận hiệp phương sai của các hệ số ước lượng ứng với các biến C, LK và LL là

	C	LL	LK
C	10.13675	-1.443758	0.291585
LL	-1.443758	0.211962	-0.054847
LK	0.291585	-0.054847	0.036514

- Viết hàm số kinh tế ban đầu với các biến Q, K, L.
- Viết hàm hồi quy mẫu. Cho biết ý nghĩa của các ước lượng nhận được.
- Các ước lượng nhận được có phù hợp với lý thuyết không?
- Các biến giải thích giải thích được bao nhiêu phần trăm sự biến động của biến phụ thuộc.
- Kiểm định sự phù hợp của hàm hồi quy.

- f. Khi lao động tăng 1% thì sản lượng tăng trong khoảng bao nhiêu % ?
- g. Khi vốn giảm 1% thì sản lượng giảm tối đa bao nhiêu %?
- h. Nguồn vốn tăng lên bằng 1,2 lần so với trước thì sản lượng có tăng tương ứng bằng 1,2 lần không?
- i. Khi yếu tố khác không đổi, nếu nguồn vốn tăng lên t lần mà sản lượng tăng nhỏ hơn t lần thì ta nói sản lượng tăng chậm hơn với tăng nguồn vốn, nếu sản lượng tăng lớn hơn t lần ta gọi là tăng nhanh hơn so với tăng nguồn vốn và bằng đúng t thì gọi là tăng bằng với tăng nguồn vốn. Theo kết quả hồi quy trên thì sản lượng tăng là nhanh, chậm hay bằng so với tăng nguồn vốn?

- Chương 1. MỞ ĐẦU
- Chương 2. ƯỚC LƯỢNG VÀ KIỂM ĐỊNH MÔ HÌNH HỒI QUY ĐƠN
- Chương 3. MÔ HÌNH HỒI QUY BỘI
- Chương 4. ĐA CỘNG TUYẾN**
- Chương 5. PHƯƠNG SAI SAI SỐ THAY ĐỔI
- Chương 6. TỰ TƯƠNG QUAN
- Chương 7. HỒI QUY VỚI BIẾN GIẢ
- Chương 8. LỰA CHỌN VÀ CHỈ ĐỊNH MÔ HÌNH

Chương 4. ĐA CỘNG TUYẾN

1. Bản chất của đa cộng tuyến (Multicollinearity)

1.1. Hiện tượng

Xét MH:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i$$

Giả thiết 10: Các biến giải thích không có quan hệ cộng tuyến.
Nếu giả thiết bị vi phạm thì gọi là hiện tượng đa cộng tuyến.

Đa cộng tuyến hoàn hảo (Perfect Multicollinearity):

$\exists \lambda_j \neq 0 (j \neq 1)$ sao cho:

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

$\Rightarrow$ Ma trận X là suy biến, không có lời giải duy nhất.

Đa cộng tuyến không hoàn hảo (Imperfect Multicollinearity) :

$\exists \lambda_j \neq 0 (j \neq 1)$ sao cho:

$$\lambda_2 X_{2i} + \dots + \lambda_k X_{ki} + v_i = 0 \forall i$$

với v_i là SSNN có phương sai dương $\Rightarrow$ vẫn có lời giải.

1.2. Nguyên nhân

Đa cộng tuyến hoàn hảo gần như không bao giờ xảy ra.

Đa cộng tuyến không hoàn hảo thường xuyên xảy ra, do các nguyên nhân:

- Bản chất các biến giải thích có quan hệ tương quan với nhau.
- Do số liệu mẫu không ngẫu nhiên.
- Do kích thước mẫu không đủ.
- Do quá trình làm tròn số liệu.

2. Hậu quả

2.1. Đa cộng tuyến hoàn hảo: không giải được

Vì lúc đó $\hat{\beta}_j = \frac{0}{0}$ và $var(\hat{\beta}_j) = \infty \quad \forall j$.

2.2. Đa cộng tuyến không hoàn hảo:

- Các ước lượng có phương sai lớn, là ước lượng không hiệu quả.
- Khoảng tin cậy rộng không còn ý nghĩa.
- Các kiểm định t có thể sai.
- Các kiểm định t và kiểm định F có thể cho kết luận mâu thuẫn nhau.
- Các ước lượng có thể sai về dấu.
- Mô hình trở nên nhạy cảm với mỗi sự thay đổi của số biến giải thích và của tập số liệu.

3. Phát hiện đa cộng tuyến.

3.1. Sự mâu thuẫn giữa kiểm định t và F

Có mâu thuẫn: Kiểm định F có ý nghĩa, hầu hết các kiểm định t về các hệ số góc không có ý nghĩa.

3.2. Hồi qui phụ

Nghi ngờ biến giải thích X_j phụ thuộc tuyến tính vào các biến giải thích khác, hồi qui mô hình hồi qui phụ:

$$X_j = \alpha_1 + \alpha_2 X_2 + \dots + \alpha_{j-1} X_{j-1} + \alpha_{j+1} X_{j+1} + \dots + v_i \quad (*)$$

H_0 : Mô hình ban đầu không có Đa cộng tuyến

H_1 : Mô hình ban đầu có Đa cộng tuyến

Dependent Variable: Y

Method: Least Squares

Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	33.87971	19.11513	1.772403	0.0902
X2	-26.00263	34.95897	-0.743804	0.4649
X3	6.709261	8.740550	0.767602	0.4509
R-squared	0.741695	Mean dependent var	169.3680	
Adjusted R-squared	0.718213	S.D. dependent var	79.05857	
S.E. of regression	41.96716	Akaike info criterion	10.42382	
Sum squared resid	38747.34	Schwarz criterion	10.57008	
Log likelihood	-127.2977	F-statistic	31.58532	
Durbin-Watson stat	2.785912	Prob(F-statistic)	0.000000	

Dependent Variable: X2

Sample: 1 25

Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.104988	0.111892	-0.938303	0.3578
X3	0.250022	0.000157	1594.459	0.0000
R-squared	0.999991	Mean dependent var	159.4480	
Adjusted R-squared	0.999991	S.D. dependent var	81.46980	
S.E. of regression	0.250315	Akaike info criterion	0.144426	
Sum squared resid	1.441126	Schwarz criterion	0.241936	
Log likelihood	0.194678	F-statistic	2542299.	
Durbin-Watson stat	2.245068	Prob(F-statistic)	0.000000	

Dependent Variable: X3

Sample: 1 25

Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.425686	0.447288	0.951703	0.3511
X2	3.999613	0.002508	1594.459	0.0000
R-squared	0.999991	Mean dependent var	638.1560	
Adjusted R-squared	0.999991	S.D. dependent var	325.8492	
S.E. of regression	1.001168	Akaike info criterion	2.916830	
Sum squared resid	23.05376	Schwarz criterion	3.014340	
Log likelihood	-34.46038	F-statistic	2542299.	
Durbin-Watson stat	2.245125	Prob(F-statistic)	0.000000	

Dependent Variable: Y
Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	36.73575	18.58133	1.977025	0.0601
X2	0.831821	0.104206	7.982448	0.0000
R-squared	0.734777	Mean dependent var		169.3680
Adjusted R-squared	0.723246	S.D. dependent var		79.05857
S.E. of regression	41.59070	Akaike info criterion		10.37025
Sum squared resid	39785.09	Schwarz criterion		10.46776
Log likelihood	-127.6281	F-statistic		63.71948
Durbin-Watson stat	2.919889	Prob(F-statistic)		0.000000

4. Khắc phục đa cộng tuyến.

4.1. Dùng thông tin tiên nghiệm.

Ví dụ: Xét mô hình

$$TD_i = \beta_1 + \beta_2 TN_i + \beta_3 SK_i + u_i$$

Nếu thấy TD_i có cộng tuyến với SK_i
và có thể biết rằng $\beta_3 = 0, 1\beta_2$ Thì mô hình trở thành

$$TD_i = \beta_1 + \beta_2(TN_i + 0, 1SK_i) + u_i$$

Và đã khắc phục được đa cộng tuyến.

4.2. Bỏ bớt biến nếu có thể.

Việc lựa chọn biến bị loại khỏi mô hình có thể căn cứ vào kết quả của hồi quy phụ.

4.3. Tăng kích thước mẫu hoặc lấy mẫu mới nếu có thể.

4.4. Đổi dạng của mô hình.

4.5. Dùng sai phân cấp 1.

Chương 1. MỞ ĐẦU	
Chương 2. ƯỚC LƯỢNG VÀ KIỂM ĐỊNH MÔ HÌNH HỒI QUY ĐƠN	
Chương 3. MÔ HÌNH HỒI QUY BỘI	
Chương 4. ĐA CỘNG TUYẾN	
Chương 5. PHƯƠNG SAI SAI SỐ THAY ĐỔI	
Chương 6. TỰ TƯƠNG QUAN	
Chương 7. HỒI QUY VỚI BIẾN GIẢ	
Chương 8. LỰA CHỌN VÀ CHỈ ĐỊNH MÔ HÌNH	

Chương 5. PHƯƠNG SAI SAI SỐ THAY ĐỔI

1. Bản chất của hiện tượng phương sai của sai số thay đổi

1.1. Hiện tượng

MH ban đầu:

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

Gt 4: Phương sai các sai số ngẫu nhiên là đồng nhất
 $Var(u_i) = \sigma^2$ không đổi.

1.2. Nguyên nhân

- Bản chất hiện tượng Kinh tế xã hội.
- Số liệu không đúng bản chất hiện tượng.
- Quá trình xử lý số liệu.

2. Hậu quả

- Các ước lượng là không chệch, nhưng không hiệu quả và do đó không phải là tốt nhất.
- Các kiểm định t , F có thể sai, khoảng tin cậy rộng.

3. Phát hiện

3.1. Đồ thị phần dư

Dùng đồ thị của e_i , $|e_i|$ hoặc e_i^2 để đánh giá.

3.2. Kiểm định Park

Giả thiết:

$$\sigma_i^2 = \sigma^2 X_i^{\alpha_2}$$

Trong đó σ^2 là một hằng số.

MH hồi qui phụ

$$\ln e_i^2 = \alpha_1 + \alpha_2 \ln X_i + v_i(*)$$

Bước 1. Hồi quy mô hình ban đầu tìm được e_i

Bước 2. Hồi quy mô hình (*)

Bước 3. Kiểm định cặp giả thuyết

$$H_0 : \alpha_2 = 0; H_1 : \alpha_2 \neq 0$$

3.3. Kiểm định Glejer

Tùy vào giả thiết mà thực hiện hồi qui phụ để kiểm định.

Giả thiết:

$$\sigma_i^2 = \sigma^2 X_i$$

MH hồi qui phụ

$$|e_i| = \alpha_1 + \alpha_2 X_i + v_i(*)$$

Dùng kiểm định T hoặc F để kiểm định

3.4. Kiểm định White

Dùng cho mô hình nhiều biến giải thích. Hồi qui bình phương phân dư theo tổ hợp bậc cao dần của các biến giải thích.

VD : MH ban đầu

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

MH hồi qui phụ :

$$e^2 = \alpha_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_2^2 + \alpha_5 X_3^2 + \alpha_6 X_2 X_3 (+ \dots +) + v_i (*)$$

Kiểm định χ^2 $\chi_{qs}^2 = nR_*^2$, nếu $\chi_{qs}^2 > \chi_\alpha^2(k_* - 1)$ thì bác bỏ H_0

Ví dụ

Cho biết D là nợ nước ngoài và Y là tổng sản phẩm trong nước của 73 nước đang phát triển. Hãy kiểm định hiện tượng phương sai sai số thay đổi. Giả sử kết quả hồi quy D theo Y như sau:

Dependent Variable: D

Included observations: 73

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3521.957	1066.371	3.302752	0.0015
Y	0.275643	0.017613	15.64971	0.0000
R-squared	0.775255	Mean dependent var	10982.11	
Adjusted R-squared	0.772090	S.D. dependent var	17071.75	
S.E. of regression	8150.046	Akaike info criterion	20.87645	
Sum squared resid	4.72E+09	Schwarz criterion	20.93920	
Log likelihood	-759.9904	F-statistic	244.9136	
Durbin-Watson stat	2.081218	Prob(F-statistic)	0.000000	

Kiểm định Park cho kết quả sau:

Dependent Variable: LOG(E^2)

Included observations: 73

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.549893	1.619225	5.897819	0.0000
LOG(Y)	0.675378	0.177575	3.803343	0.0003
R-squared	0.169255	Mean dependent var	15.62156	
Adjusted R-squared	0.157554	S.D. dependent var	2.521699	
S.E. of regression	2.314538	Akaike info criterion	4.543312	
Sum squared resid	380.3532	Schwarz criterion	4.606065	
Log likelihood	-163.8309	F-statistic	14.46542	
Durbin-Watson stat	2.267623	Prob(F-statistic)	0.000299	

Kiểm định Gleijer cho kết quả sau:

Dependent Variable: ABS(E)

Included observations: 73

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2890.521	715.5377	4.039649	0.0001
Y88	0.067345	0.011819	5.698195	0.0000
R-squared	0.313807	Mean dependent var	4713.169	
Adjusted R-squared	0.304142	S.D. dependent var	6555.777	
S.E. of regression	5468.703	Akaike info criterion	20.07849	
Sum squared resid	2.12E+09	Schwarz criterion	20.14124	
Log likelihood	-730.8647	F-statistic	32.46943	
Durbin-Watson stat	2.155443	Prob(F-statistic)	0.000000	

Kiểm định White cho kết quả sau:

White Heteroskedasticity Test:

F-statistic	32.12701	Probability	0.000000
Obs*R-squared	34.93782	Probability	0.000000

Test Equation:

Dependent Variable: RESID^2

Included observations: 73

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-32588022	20003195	-1.629141	0.1078
Y	5760.776	817.7877	7.044341	0.0000
Y^2	-0.016020	0.003069	-5.219110	0.0000

R-squared	0.478600	Mean dependent var	64603429
Adjusted R-squared	0.463703	S.D. dependent var	1.84E+08
S.E. of regression	1.35E+08	Akaike info criterion	40.31714
Sum squared resid	1.27E+18	Schwarz criterion	40.41127
Log likelihood	-1468.576	F-statistic	32.12701

3.5. Kiểm định tương quan hạng của Spearman

3.6. Kiểm định Goldfeld-Quandt

3.7. Kiểm định Breusch-Pagan

3. Khắc phục

Dựa trên giả thiết về sự thay đổi của PSSS thay đổi mà ta có các phương pháp khắc phục khác nhau.

3.1. Nếu biết σ_i^2 . Dùng WLS- Phương pháp bình phương nhỏ nhất có trọng số.

Chia hai vế mô hình cho σ_i

$$\frac{Y_i}{\sigma_i} = \beta_1 \frac{1}{\sigma_i} + \beta_2 \frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i}$$

Ta được

$$Y'_i = \beta_1 X_{0i} + \beta_2 X'_i + u'_i$$

3.2. Nếu chưa biết σ_i^2 – Dùng GLS – Phương pháp bình phương nhỏ nhất tổng quát.

Tùy thuộc vào tính chất của σ_i^2 mà biến đổi mô hình gốc sao cho phương sai của sai số ngẫu nhiên trở nên đồng đều.

Giả thiết 1: $\sigma_i^2 = \sigma^2 X_i$

Lúc đó chia hai vế cho $\sqrt{X_i}$:

$$\frac{Y_i}{\sqrt{X_i}} = \beta_1 \frac{1}{\sqrt{X_i}} + \beta_2 \sqrt{X_i} + \frac{u_i}{\sqrt{X_i}}$$

$\Rightarrow$ PSSS sẽ bằng σ^2

Giả thiết 2: $\sigma_i^2 = \sigma^2 X_i^2$

Lúc đó chia hai vế cho X_i .

Giả thiết 3: $\sigma_i^2 = \sigma^2 (EY_i)^2$

Lúc đó chia hai vế cho $\hat{Y}_i$

Chú ý

Có thể thay đổi dạng hàm để khắc phục phương sai của sai số thay đổi.

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

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

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

Ví dụ

Tiếp tục sử dụng tập số liệu trên, giả sử $\sigma_i^2 = \sigma^2 Y_i$ lúc đó ta hồi quy mô hình sau:

$$\frac{D_i}{\text{sqr}(Y)} = \beta_1 \frac{1}{\text{sqr}(Y_i)} + \beta_2 \text{sqr}(Y_i) + \frac{u_i}{\text{sqr}(Y_i)}$$

Kết quả hồi quy như sau:

Dependent Variable: D/SQR(Y)

Included observations: 73

Variable	Coefficient	Std. Error	t-Statistic	Prob.
1/SQR(Y)	1024.466	270.9094	3.781581	0.0003
SQR(Y)	0.367922	0.028848	12.75364	0.0000
R-squared	0.308884	Mean dependent var		64.11631
Adjusted R-squared	0.299150	S.D. dependent var		45.42712
S.E. of regression	38.03013	Akaike info criterion		10.14165
Sum squared resid	102686.7	Schwarz criterion		10.20440
Log likelihood	-368.1702	Durbin-Watson stat		2.013380

Kết quả kiểm định White như sau:

White Heteroskedasticity Test:

F-statistic	2.248267	Probability	0.072848
Obs*R-squared	8.526662	Probability	0.074083

Test Equation:

Dependent Variable: RESID^2

Included observations: 73

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-735.5216	5167.096	-0.142347	0.8872
1/SQR(Y88)	-3833.260	285895.2	-0.013408	0.9893
(1/SQR(Y88))^2	910480.8	4365419.	0.208567	0.8354
SQR(Y88)	19.78189	28.10322	0.703901	0.4839
(SQR(Y88))^2	-0.018487	0.040348	-0.458186	0.6483
R-squared	0.116804	Mean dependent var	1406.667	
Adjusted R-squared	0.064851	S.D. dependent var	3300.866	

Chương 1. MỞ ĐẦU	
Chương 2. ƯỚC LƯỢNG VÀ KIỂM ĐỊNH MÔ HÌNH HỒI QUY ĐƠN	
Chương 3. MÔ HÌNH HỒI QUY BỘI	
Chương 4. ĐA CỘNG TUYẾN	
Chương 5. PHƯƠNG SAI SAI SỐ THAY ĐỔI	
Chương 6. TỰ TƯƠNG QUAN	
Chương 7. HỒI QUY VỚI BIẾN GIẢ	
Chương 8. LỰA CHỌN VÀ CHỈ ĐỊNH MÔ HÌNH	

Chương 6. TỰ TƯƠNG QUAN

1. Hiện tượng tự tương quan (Autocorrelation or Serial correlation)

1.1. Hiện tượng

MH ban đầu:

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

Giả thiết 5: Các sai số ngẫu nhiên không tương quan

$$\text{Cov}(u_i, u_j) = 0 (\forall i \neq j)$$

hoặc

$$\text{Cov}(u_t, u_{t-p}) = 0 (p \neq 0)$$

Nếu giả thiết bị vi phạm: hiện tượng tự tương quan bậc p.

Xét trường hợp $p = 1$

u_t và u_{t-1} có cùng trung bình và phương sai.

$u_t = \rho u_{t-1} + \epsilon_t$ ($-1 \leq \rho \leq 1$, ϵ_t thỏa mãn các giả thiết của OLS)

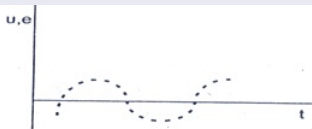
- $\rho = -1$ tự tương quan âm hoàn hảo
- $-1 < \rho < 0$ tự tương quan âm
- $\rho = 0$ không có tự tương quan
- $0 < \rho < 1$ tự tương quan dương
- $\rho = 1$ tự tương quan dương hoàn hảo

$\Rightarrow$ Lược đồ AR(1).

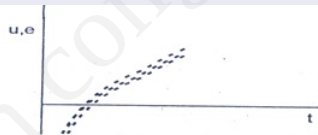
Tổng quát: tự tương quan bậc p :

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \epsilon_t$$

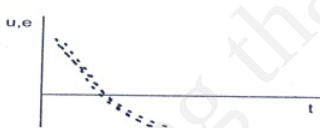
với $\rho_p \neq 0 \Rightarrow$ Lược đồ AR(p): Autoregressive Procedure order p .



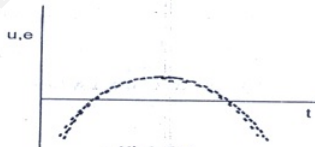
Hình 5.1



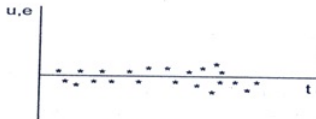
Hình 5.2



Hình 5.3



Hình 5.4



Hình 5.5

- Nhưng hình 5.5 cho thấy không có một mẫu có hệ thống nào và nó ủng hộ cho giả thiết không có tự tương qua trong mô hình hồi quy tuyến tính cổ điển.

1.2. Nguyên nhân

- Bản chất, tính quán tính trong hiện tượng kinh tế xã hội
- Hiện tượng mạng nhện trong kinh tế
- Quá trình xử lý, nội suy số liệu
- Mô hình thiếu biến hoặc dạng hàm sai

2. Hậu quả

1. Ước lượng bình phương nhỏ nhất thông thường vẫn là ước lượng tuyến tính không chệch, nhưng chúng không phải là ước lượng hiệu quả nữa.

2. Phương sai ước lượng được của các ước lượng bình phương nhỏ nhất thông thường là chệch. Hậu quả sẽ ngộ nhận rằng một hệ số nào đó khác không có ý nghĩa về mặt thống kê, nhưng thực tế lại không phải như vậy.

3. Do đó các kiểm định t và F nói chung không đáng tin cậy.

4. Công thức thông thường để tính phương sai của sai số là ước lượng chệch của σ^2 thực, và trong một số trường hợp, nó dường như ước lượng thấp.

5. Kết quả là R^2 được tính toán như đã biết, có thể là độ đo không đáng tin cậy cho R^2 thực và dường như ước lượng cao R^2 .

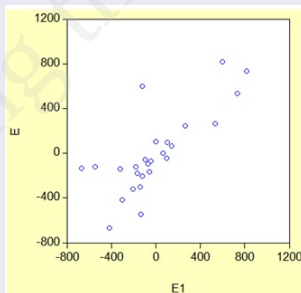
6. Các phương sai và sai số tiêu chuẩn của dự đoán đã tính được cũng có thể không hiệu quả.

3. Phát hiện

3.1. Quan sát đồ thị của e_t theo e_{t-1}

Bước 1. Hồi quy mô hình gốc để tìm e_t và e_{t-1}

Bước 2. Vẽ đồ thị của e_t theo e_{t-1} và nhận xét.



3.2. Kiểm định Durbin – Watson

Kiểm định Durbin – Watson dựa trên thống kê

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \cong 2(1 - \hat{\rho})$$

với

$$\hat{\rho} = \frac{\sum_{i=1}^n e_i e_{i-1}}{\sum_{i=1}^n e_i^2}$$

là ước lượng cho ρ

Chú ý

Kiểm định DW sẽ chỉ áp dụng được nếu thoả mãn các điều kiện sau:

- + Mô hình phải có hệ số chặn.
- + Biến giải thích phải là phi ngẫu nhiên.
- + Nếu có hiện tượng tự tương quan thì đó chỉ là lược đồ $AR(1)$.
- + Mô hình không chứa biến trễ của biến phụ thuộc làm biến giải thích.
- + Không có quan sát nào bị mất trong tệp số liệu.

Do không tìm được chính xác phân phối xác suất của d nên dựa vào tính chất của nó để kết luận.

$$\text{Do } -1 \leq \hat{\rho} \leq 1 \implies 0 \leq d \leq 4.$$

Với $n, k' = k - 1$ và $\alpha = 0,05$ cho trước, tra bảng ta tìm được d_L và d_U .

Quy tắc ra quyết định

<u>Tự tương quan dương</u> $\rho > 0$	<u>Không có kết luận</u>	<u>Không có tự tương quan</u> $\rho = 0$	<u>Không có kết luận</u>	<u>Tự tương quan âm</u> $\rho < 0$		
0	d_L	d_U	2	$4 - d_U$	$4 - d_L$	4

Hạn chế của kiểm định DW:

- + Vẫn còn hai miền không có kết luận $\Rightarrow$ Dùng kiểm định DW cải biên.
- + Không áp dụng được với các mô hình tự hồi quy.

3.3 Hồi qui phụ

Kiểm định

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \epsilon_t$$

MH hồi qui phụ :

$$e_t = \alpha_0 + \alpha_1 e_{t-1} + \dots + \alpha_p e_{t-p} + v_t \quad (*)$$

$$\begin{cases} H_0 : R_*^2 = 0 \\ H_1 : R_*^2 \neq 0 \end{cases} \iff \begin{cases} H_0 : \alpha_1 = \dots = \alpha_p = 0 \\ H_1 : \exists \alpha_j \neq 0 : (j \neq 0) \end{cases}$$

H_0 : Không có tự tương quan đến bậc p. H_1 Có tự tương quan ở bậc tương ứng.

Kiểm định T hoặc F

3.4. Kiểm định Breusch- Goldfrey.

MH hồi qui phụ

$$e_t = [\beta_1 + \beta_2 X_t] + \alpha_1 e_{t-1} + \dots + \alpha_p e_{t-p} + v_t \quad (*)$$

$$\begin{cases} H_0 : \alpha_1 = \dots = \alpha_p = 0 \\ H_1 : \exists \alpha_j \neq 0 : (j \neq 0) \end{cases}$$

Kiểm định

$$\chi^2 : \chi_{qs}^2 = n_* R_*^2 = (n - p) R_*^2$$

Nếu

$$\chi_{qs}^2 > \chi_{\alpha}^2(p)$$

thì bác bỏ H_0

Kiểm định F: Hồi qui

$$e_t = [\beta_1 + \beta_2 X_t] + v_t \quad (**)$$

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

Nếu

$$F_{qs} > F_{\alpha}(k_* - 1; n_* - k_*)$$

thì bác bỏ H_0

Ví dụ: Tập số liệu ch7bt4 gồm hai biến Tiêu dùng và thu nhập của Nigeria. Hãy kiểm định hiện tượng tự tương quan của mô hình.
Hồi quy Tiêu dùng theo Thu nhập được kết quả sau:

Dependent Variable: CONS

Included observations: 27

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	155.2239	203.4712	0.762879	0.4527
GDP	0.597069	0.060594	9.853648	0.0000
R-squared	0.795240	Mean dependent var		2037.449
Adjusted R-squared	0.787050	S.D. dependent var		789.2231
S.E. of regression	364.1989	Akaike info criterion		14.70446
Sum squared resid	3316021.	Schwarz criterion		14.80045
Log likelihood	-196.5103	F-statistic		97.09438
Durbin-Watson stat	0.462830	Prob(F-statistic)		0.000000

Dùng phương pháp hồi quy phụ cho kết quả sau:

Dependent Variable: E

Included observations: 24 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.256609	52.50711	0.004887	0.9961
E(-1)	0.843549	0.221024	3.816555	0.0011
E(-2)	0.020077	0.293835	0.068326	0.9462
E(-3)	-0.173636	0.233916	-0.742300	0.4665
R-squared	0.604411	Mean dependent var	-8.530418	
Adjusted R-squared	0.545072	S.D. dependent var	378.2292	
S.E. of regression	255.1093	Akaike info criterion	14.07227	
Sum squared resid	1301615.	Schwarz criterion	14.26862	
Log likelihood	-164.8673	F-statistic	10.18582	
Durbin-Watson stat	2.022007	Prob(F-statistic)	0.000278	

Kiểm định BG cho kết quả sau:

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	34.31433	Probability	0.000005
Obs*R-squared	15.88781	Probability	0.000067

Test Equation:

Dependent Variable: RESID

Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-60.84700	133.6292	-0.455342	0.6530
GDP	0.021511	0.039844	0.539890	0.5942
RESID(-1)	0.777523	0.132732	5.857844	0.0000

R-squared	0.588437	Mean dependent var	-8.42E-14
Adjusted R-squared	0.554140	S.D. dependent var	357.1264
S.E. of regression	238.4630	Akaike info criterion	13.89074
Sum squared resid	1364750.	Schwarz criterion	14.03473
Log likelihood	-184.5250	F-statistic	17.15717
Durbin-Watson stat	1.846319	Prob(F-statistic)	0.000024

4. Khắc phục

Mục đích là chuyển MH ban đầu có khuyết tật tự tương quan thành MH mới có cùng hệ số cũ nhưng không có tự tương quan. Mô hình ban đầu:

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

Có tự tương quan :

$$u_t = \rho u_{t-1} + \epsilon_t$$

với $\rho \neq 0$, ϵ_t thỏa mãn các giả thiết OLS.

4.1 Khi ρ đã biết – Dùng phương pháp sai phân tổng quát.

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

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \iff \rho Y_{t-1} = \rho\beta_1 + \beta_2 \rho X_{t-1} + \rho u_{t-1}$$

$$Y_t - \rho Y_{t-1} = \beta_1(1 - \rho) + \beta_2(X_t - \rho X_{t-1}) + u_t - \rho u_{t-1}$$

(phương trình sai phân tổng quát)

$$Y_t^* = \beta_1^* + \beta_2^* X_t + v_t$$

Ước lượng bằng OLS $\hat{\beta}_1 = \frac{\hat{\beta}_1^*}{1-\rho}$ và $\hat{\beta}_2 = \frac{\hat{\beta}_2^*}{1-\rho}$

4.2. Khi ρ chưa biết

Ước lượng ρ bằng các phương pháp khác nhau

i. Từ thống kê Durbin-Watson

$$d = 2(1 - \rho) \Rightarrow \rho = 1 - \frac{d}{2}$$

ii. Từ hồi qui phụ

$$e_t = \alpha_0 + \alpha_1 e_{t-1} + v_t$$

lấy $\rho = \alpha_1$

iii. Phương pháp Cochran-Orcutt

Hồi qui mô hình ban đầu:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \Rightarrow \hat{\beta}_1^{(1)}, \hat{\beta}_2^{(1)} e_t^{(1)}, ,$$

Hồi qui mô hình

$$e_t^{(1)} = \alpha + \rho e_{t-1}^{(1)} + v_t \Rightarrow \hat{\rho}^{(1)}$$

Lấy $\hat{\rho}^{(1)}$ thay vào phương trình sai phân tổng quát, thu được
 $\hat{\beta}_1^{(2)}, \hat{\beta}_2^{(2)} e_t^{(2)}$

Hồi qui mô hình

$$e_t^{(2)} = \alpha + \rho e_{t-1}^{(2)} + v_t \Rightarrow \hat{\rho}^{(2)}$$

Lấy $\hat{\rho}^{(2)}$ thay vào phương trình sai phân tổng quát, thu được
 $\hat{\beta}_1^{(3)}, \hat{\beta}_2^{(3)}, e_t^{(3)}$

... Quá trình lặp cho đến khi $\hat{\rho}$ ở hai bước kế tiếp chênh lệch nhau
không đáng kể thì $\hat{\beta}_1$ và $\hat{\beta}_2$ ở bước cuối cùng là ước lượng cho β_1
và β_2

iv. Phương pháp Durbin – Watson hai bước.

Trước hết viết lại mô hình sai phân tổng quát dưới dạng:

$$Y_t = \beta_1(1 - \rho) + \beta_2 X_t - \beta_2 \rho X_{t-1} + \rho Y_{t-1} + (u_t - \rho u_{t-1})$$

Bước 1. Ước lượng mô hình trên thu được $\hat{\rho}$

Bước 2. Thiết lập mô hình sai phân tổng quát và ước lượng nó để tìm $\hat{\beta}_1$ và $\hat{\beta}_2$

v. Đưa thêm biến vào mô hình.

Một trong các nguyên nhân của tự tương quan là mô hình bị chỉ định sai do bỏ sót biến giải thích vì vậy việc đưa thêm biến vào mô hình cũng có thể khắc phục tự tương quan.

Ví dụ:

Tiếp tục ví dụ đã cho, ta tìm cách khắc phục hiện tượng tự tương quan của mô hình.

Nếu sử dụng thống kê DW, từ báo cáo ta có

$$d = 0,46283 \iff \hat{\rho} = 1 - \frac{d}{2} = 0,768585$$

Thiết lập các biến sai phân tổng quát:

$$Cons1_t = cons_t - 0,768585 cons_{t-1}$$

$$gdp1_t = gdp_t - 0,768585 gdp_{t-1}$$

Dependent Variable: CONS1

Included observations: 26 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	143.5202	123.3592	1.163433	0.2561
GDP1	0.467009	0.140706	3.319044	0.0029
R-squared	0.314600	Mean dependent var		523.5229
Adjusted R-squared	0.286042	S.D. dependent var		277.1436
S.E. of regression	234.1754	Akaike info criterion		13.82382
Sum squared resid	1316115.	Schwarz criterion		13.92060
Log likelihood	-177.7097	F-statistic		11.01605
Durbin-Watson stat	1.718728	Prob(F-statistic)		0.002875

Để sử dụng phương pháp lặp Cochrane- Orcutt, ta khai báo thêm vào danh sách biến giải thích bậc của tự tương quan, chẳng hạn ở đây là AR(1) và hồi quy bằng OLS thu được kết quả sau:

Dependent Variable: CONS

Included observations: 26 after adjusting endpoints

Convergence achieved after 39 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1290.375	864.3354	1.492910	0.1491
GDP	0.320087	0.180066	1.777612	0.0887
AR(1)	0.895753	0.097838	9.155465	0.0000
R-squared	0.917399	Mean dependent var		2068.732
Adjusted R-squared	0.910216	S.D. dependent var		787.5967
S.E. of regression	235.9953	Akaike info criterion		13.87367
Sum squared resid	1280957.	Schwarz criterion		14.01883
Log likelihood	-177.3577	F-statistic		127.7230
Durbin-Watson stat	1.884378	Prob(F-statistic)		0.000000
Inverted AR Roots	.90			

Kết quả thu được là ước lượng ở bước cuối cùng, đồng thời kết quả cũng cho ước lượng của lược đồ AR(1).

Ta cũng có thể khắc phục tự tương quan của mô hình bằng cách đưa thêm biến trễ vào mô hình vì theo lý thuyết kinh tế thì tiêu dùng thường còn phụ thuộc vào tiêu dùng ở kỳ trước, vì vậy có thể thử nghiệm mô hình sau:

$$Cons_t = \beta_1 + \beta_2 gdp_t + \beta_3 Cons_{t-1} + u_t$$

Dependent Variable: CONS

Included observations: 26 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.962031	122.4302	0.032362	0.9745
GDP	0.207495	0.064345	3.224707	0.0038
CONS(-1)	0.695319	0.094013	7.396025	0.0000
R-squared	0.937124	Mean dependent var		2068.732
Adjusted R-squared	0.931656	S.D. dependent var		787.5967
S.E. of regression	205.8988	Akaike info criterion		13.60081
Sum squared resid	975068.9	Schwarz criterion		13.74598
Log likelihood	-173.8106	F-statistic		171.3986
Durbin Watson stat	1.910159	Prob(F-statistic)		0.000000

Chương 7. HỒI QUY VỚI BIẾN GIẢ

- Trong các mô hình hồi quy tuyến tính mà chúng ta đã xem xét từ các chương trước cho đến nay thì các biến giải thích đều là các biến lượng. Các biến đó có thể nhận giá trị bằng số.
- Chẳng hạn, tiền lương của cán bộ, doanh số bán ra của một cửa hàng, chi tiêu cho quảng cáo, cung tiền... đều là những biến số lượng. Nhưng trong thực tế có nhiều trường hợp các biến giải thích (hoặc thậm chí cả biến phụ thuộc) là biến chất lượng.
- Trong chương này ta sẽ nghiên cứu hồi quy khi biến giải thích là biến chất lượng, đặc biệt trong lý thuyết phân tích chính sách.

I. HỒI QUY VỚI CÁC BIẾN GIẢ THÍCH LÀ BIẾN CHẤT

Biến chất lượng như ta đã nói ở trên thường chỉ ra có hoặc không có một thuộc tính nào đó, chẳng hạn nam hay nữ; khu vực tư nhân thay nhà nước. Vấn đề đặt ra là làm thế nào để lượng hoá được những thuộc tính như vậy.

Trong phân tích hồi quy người ta sử dụng kỹ thuật biến giả. Kỹ thuật này cho phép ta ước lượng hoá những thuộc tính như vậy. Biến giả được sử dụng trong mô hình hồi quy giống như biến số lượng thông thường.

Trước hết ta giả sử rằng, ở nước ta có hai loại hình xí nghiệp, đó là xí nghiệp quốc doanh và xí nghiệp tư nhân. Chúng ta có thể biểu thị điều này bằng phương pháp hồi quy sau:

$$W_i = \beta_1 + \beta_2 D_i + U_t$$

(3.1) trong đó W_i là tiền lương của công nhân i . D_i là biến giả một trong hai giá trị như sau: . . $D=0$ tư nhân; $D=1$ nhà nước

Điều này có thể thấy rằng bằng cách lấy giá trị kỳ vọng cả 2 vế của phương trình ứng với $D_t = 0$ và $D_t = 1$:

$$E(Y|D_i = 0) = \beta_1$$

$$E(Y|D_i = 1) = \beta_1 + \beta_2$$

Kiểm định giả thiết $H_0; \beta_2 = 0$ cung cấp kiểm định về giả thiết là không có sự khác nhau về tiền lương của công nhân làm việc trong các xí nghiệp tư nhân và nhà nước. Điều này dễ làm được như đã chỉ ra trước đây.

Chương 1. MỞ ĐẦU	
Chương 2. ƯỚC LƯỢNG VÀ KIỂM ĐỊNH MÔ HÌNH HỒI QUY ĐƠN	
Chương 3. MÔ HÌNH HỒI QUY BỘI	
Chương 4. ĐA CỘNG TUYẾN	
Chương 5. PHƯƠNG SAI SAI SỐ THAY ĐỔI	
Chương 6. TƯ TƯỢNG QUAN	
Chương 7. HỒI QUY VỚI BIẾN GIẢ	
Chương 8. LỰA CHỌN VÀ CHỈ ĐỊNH MÔ HÌNH	

Chương 1. MỞ ĐẦU	
Chương 2. ƯỚC LƯỢNG VÀ KIỂM ĐỊNH MÔ HÌNH HỒI QUY ĐƠN	
Chương 3. MÔ HÌNH HỒI QUY BỘI	
Chương 4. ĐA CỘNG TUYẾN	
Chương 5. PHƯƠNG SAI SAI SỐ THAY ĐỔI	
Chương 6. TỰ TƯƠNG QUAN	
Chương 7. HỒI QUY VỚI BIẾN GIẢ	
Chương 8. LỰA CHỌN VÀ CHỈ ĐỊNH MÔ HÌNH	

Chương 8. LỰA CHỌN VÀ CHỈ ĐỊNH MÔ HÌNH