

Chương 6

TỰ TƯƠNG QUAN

I. Bản chất của tự tương quan

Tự tương quan là hiện tượng có sự tương quan giữa các quan sát trong cùng bảng số liệu

Hiện tượng này thường xảy ra đối với dữ liệu chuỗi thời gian và còn được gọi bằng các thuật ngữ sau

- Serial Correlation – tương quan chuỗi
- Autocorrelation – tự tương quan
- AutoRegression – tự hồi quy

I. Bản chất của tự tương quan

Trong các giả thiết của mô hình hồi quy cổ điển, ta giả định rằng giữa các sai số U_i không tương quan với nhau. Nếu giả thiết này bị vi phạm thì mô hình hồi quy sẽ bị bệnh “Tự tương quan”

Vì tự tương quan thường xảy ra với số liệu theo thời gian nên phương trình hồi quy trong chương này ta viết là :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + U_t$$

by Tuấn Anh

I. Bản chất của tự tương quan

Nếu sai số U_t chỉ tương quan với U_{t-1} (sai số một kỳ trước đó) thì ta có hiện tượng tự tương quan bậc nhất, ký hiệu là **AR(1)**

Phương trình tự tương quan bậc nhất như sau :

$$U_t = \rho U_{t-1} + \varepsilon_t \quad \text{với} \quad -1 < \rho < 1 \quad (*)$$

- ρ : hệ số tự tương quan
- ε_t : Sai số ngẫu nhiên không còn tự tương quan

by Tuấn Anh

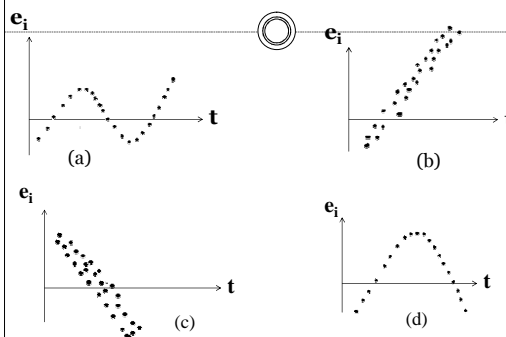
I. Bản chất của tự tương quan

Nếu U_t tương quan với m kỳ trước đó thì ta có hiện tượng tự tương quan bậc m , ký hiệu là **AR(m)** :

$$U_t = \rho_1 U_{t-1} + \rho_2 U_{t-2} + \dots + \rho_m U_{t-m} + \varepsilon_t$$

by Tuấn Anh

Một số dạng đồ thị có tự tương quan



by Tuấn Anh

II. Nguyên nhân của tự tương quan

1. Nguyên nhân khách quan

- Do tính “quán tính” của số liệu
- Do hiện tượng “mạng nhện”
- Do độ trễ của số liệu

bý Tuấn Anh

II. Nguyên nhân của tự tương quan

1. Nguyên nhân chủ quan

- Do việc xử lý số liệu (phương pháp trung bình trượt, làm trơn số liệu)
- Do việc nội suy số liệu (số liệu dân số, sản lượng bán trung thu .v.v...)
- Do lập mô hình (bỏ sót biến, do dạng hàm v.v...)
- Và các nguyên nhân khác

bý Tuấn Anh

IV. Hậu quả của tự tương quan

- Các hệ số hồi quy ước lượng được không còn tính **BLUE**.

- $\sigma^2 = \frac{RSS}{n-k}$ là ước lượng chệch dưới của σ^2

- Phương sai ước lượng được của các ước lượng thường bị chệch dưới (thấp hơn giá trị thực)
 - ❖ Làm cho tỷ số t lớn
 - ❖ Kiểm định t và F không còn ý nghĩa nữa

bý Tuấn Anh

IV. Hậu quả của tự tương quan

- R^2 cao hơn so với thực tế (vì RSS thấp hơn => ESS cao hơn => R^2 cao)

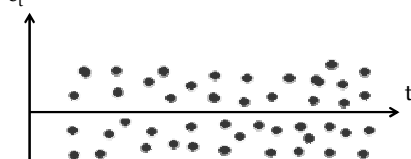
- Việc dự báo không có hiệu quả và ít được tin cậy (vì là các ước lượng chệch)

bý Tuấn Anh

V. Phát hiện tự tương quan

1. Phương pháp đồ thị:

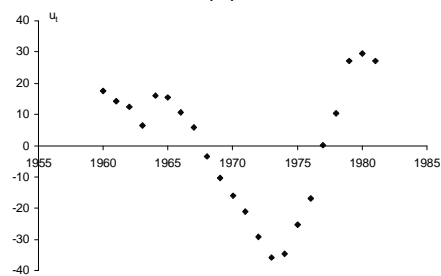
- Hồi qui mô hình gốc → thu phần dư e_t .
- Vẽ đồ thị phần dư e_t theo thời gian.



bý Tuấn Anh

V. Phát hiện tự tương quan

1. Phương pháp đồ thị:



V. Phát hiện tự tương quan

2. Phương pháp Durbin - Watson:

Phương pháp này dùng để kiểm định tự tương quan bậc nhất với giả thiết

$H_0 : \rho = 0$ (không có tự tương quan bậc nhất)

$H_1 : \rho \neq 0$ (có tự tương quan bậc nhất)

Với độ tin cậy $(1-\alpha)$

Các bước kiểm định như sau :

V. Phát hiện tự tương quan

2. Phương pháp Durbin - Watson:

Bước 1 : tính trị thống kê Durbin - Watson theo công thức

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

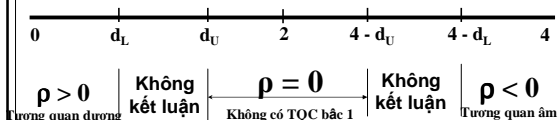
Bước 2 : tra bảng thống kê Durbin - Watson với mức ý nghĩa α , số quan sát n và số biến độc lập k' để tìm d_U và d_L

by Tuấn Anh

V. Phát hiện tự tương quan

2. Phương pháp Durbin - Watson:

Bước 3 : Kê thang kiểm định



Ví dụ : $n = 20$, $k' = 2$, $\alpha = 5\%$ và $d = 0,9$

Mô hình có bị tự tương quan bậc nhất không?

V. Phát hiện tự tương quan

2. Phương pháp Durbin - Watson:

Nhược điểm của kiểm định Durbin - Watson là gì ?

V. Phát hiện tự tương quan

2. Phương pháp Durbin - Watson:

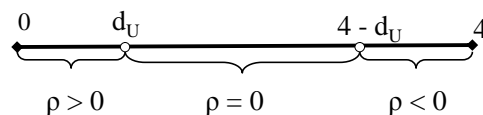
Dependent Variable: Y				
Method: Least Squares				
Date: 08/13/09 Time: 00:12				
Sample: 1 15				
Included observations: 15				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	45.28511	5.538866	8.175881	0.0000
X	5.236452	0.379310	13.80520	0.0000
Z	16.00976	2.058934	7.775752	0.0000
R-squared	0.952576	Mean dependent var	126.0667	
Adjusted R-squared	0.944672	S.D. dependent var	16.89238	
S.E. of regression	3.973393	Akaike info criterion	5.773974	
Sum squared resid	199.4542	Schwarz criterion	5.915584	
Log likelihood	-40.30481	Hannan-Quinn criter.	5.722466	
F-statistic	120.5192	Durbin-Watson stat	1.919503	
Prob(F-statistic)	0.000000			

V. Phát hiện tự tương quan

2. Phương pháp Durbin - Watson:

Kiểm định Durbin - Watson cải biên :

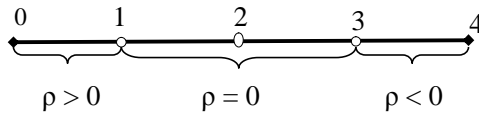
Tra bảng mức ý nghĩa 2α , số quan sát n và số biến độc lập k' , ta có d_U và d_L :



V. Phát hiện tự tương quan

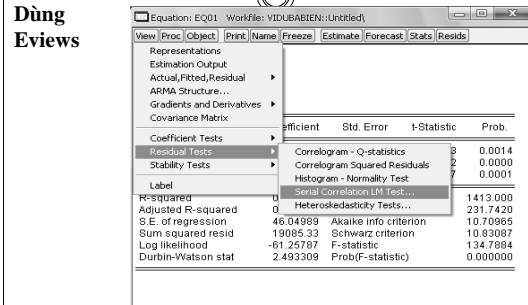
2. Phương pháp Durbin-Watson:

Kiểm định Durbin - Watson theo kinh nghiệm



V. Phát hiện tự tương quan

2. Phương pháp Breusch-Godfrey (BG test)



V. Phát hiện tự tương quan

2. Phương pháp Breusch-Godfrey (BG test)

The 'Lag Specification' dialog box has 'Lags to include:' set to 1. The 'OK' button is highlighted.

Đọc kết quả hồi quy như sau :

- Nếu p-value $\geq \alpha$: chấp nhận H_0
- Nếu p-value $< \alpha$: bác bỏ H_0

Breusch-Godfrey Serial Correlation LM Test			
F-statistic	0.082133	Prob. F(1,11)	0.7798
Obs*R-squared	0.111170	Prob. Chi-Square(1)	0.7388

VI. Khắc phục tự tương quan

1. Khi ρ đã biết.

Trong thực hành người ta thường dùng U_t theo mô hình tự tương quan bậc nhất:

$$U_t = \rho U_{t-1} + \varepsilon_t$$

Trong đó $|\rho| < 1$ và ε_t thỏa mãn các giả thiết của phương pháp OLS.

VI. Khắc phục tự tương quan

1. Khi ρ đã biết.

Ta xét hồi quy hai biến: $Y_t = \beta_1 + \beta_2 X_t + U_t$ (a)

Quan sát kỳ trước (t-1) $Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + U_{t-1}$ (b)

Nhân (b) cho ρ : $\rho Y_{t-1} = \rho \beta_1 + \rho \beta_2 X_{t-1} + \rho U_{t-1}$ (c)

Lấy (a) - (c):

$$Y_t - \rho Y_{t-1} = \beta_1(1 - \rho) + \beta_2(X_t - \rho X_{t-1}) + (U_t - \rho U_{t-1}) \quad (d)$$

$$\text{Đặt: } \begin{cases} \beta_1^* = \beta_1(1 - \rho); \beta_2^* = \beta_2 \\ Y_t^* = Y_t - \rho Y_{t-1}; X_t^* = X_t - \rho X_{t-1} \end{cases}$$

Khi đó (d) trở thành $Y_t^* = \beta_1^* + \beta_2^* X_t^* + \varepsilon_t$ (e)

Đây là phương trình hồi quy tuyến tính thông thường

VI. Khắc phục tự tương quan

1. Khi ρ chưa biết.

Bước 1: Ước lượng mô hình hai biến $Y_t = \beta_1 + \beta_2 X_t + U_t$ bằng phương pháp OLS và thu được các phần dư e_t .

Bước 2: Sử dụng các phần dư e_t để hồi quy dạng hàm:

$$e_t = \rho e_{t-1} + \varepsilon_t$$

Bước 3: Sử dụng ρ để khắc phục tự tương quan như trường hợp ρ đã biết

