CHƯƠNG 7.
TỰ TƯƠNG QUAN

Các vấn đề cần xem xét

- Định nghĩa loại khuyết tật của mô hình (Mô hình vi phạm giả thiết nào của phương pháp OLS)
- Hậu quả của khuyết tật đối với các ước lượng OLS
- Nguyên nhân của khuyết tật
- Cách phát hiện
- Giải pháp khắc phục

I. Bản chất của hiện tượng tự tương quan

I.1. Khái niệm

Thuật ngữ tự tương quan (autocorrelation) có thể hiểu là sự tương quan giữa các thành phần của chuỗi các quan sát được sắp xếp theo thứ tự thời gian (trong các số liệu chuỗi thời gian) hoặc không gian (trong số liệu chéo).

$$\mathrm{cov}\,(u_i, u_j) \neq 0$$

I.1. Khái niệm (tiếp)

- Khi có tự tương quan, giả thiết 4 của phương pháp OLS bị vi phạm.
- Giả thiết 4 của phương pháp OLS:

$$\mathrm{cov}(u_i, u_j) = 0$$

Giả thiết này có nghĩa là yếu tố ngẫu nhiên của bất kỳ quan sát nào cũng không bị ảnh hưởng bởi yếu tố ngẫu nhiên của các quan sát khác

I.1. Khái niệm (tiếp)

- Xét mô hình: $Y_t = \beta_1 + \beta_2 X_{2t} + u_t$ (7.1)
- Giả sử: $cov(u_t, u_{t-1}) \neq 0$ và:

$$u_t = \rho u_{t-1} + \varepsilon_t \quad (7.2)$$

- trong đó:

 ρ được gọi là hệ số tự hiệp phương sai (autocovariance)

ε_t là nhiễu ngẫu nhiên thoả mãn các giả thiết của OLS.

I.1. Khái niệm (tiếp)

Các giả thiết của OLS.

- $E(\varepsilon_t) = 0$
- $Var(\varepsilon_t) = \sigma^2_{\varepsilon_t}$
- $Cov(\varepsilon_t, \varepsilon_{t+s}) = 0$ (S $\neq$ 0)

I.1. Khái niệm (tiếp)

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad (7.2)$$

u_{t-1}: trễ một thời kỳ của u_t, ta nói, có tự tương quan bậc nhất trong mô hình

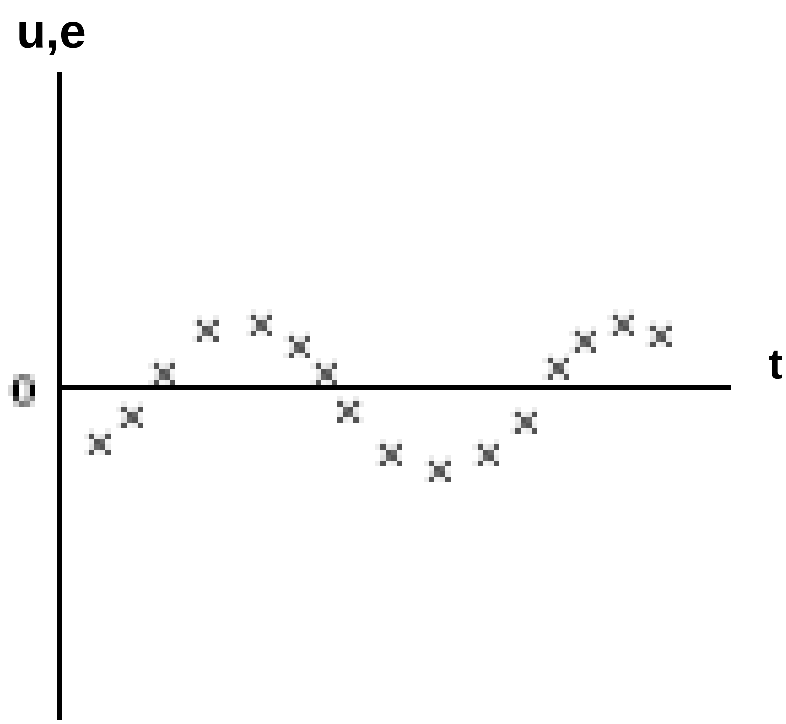
$\Rightarrow$(7.2) được ký hiệu là AR(1)

$\Rightarrow$ Tự tương quan bậc 2, AR(2)

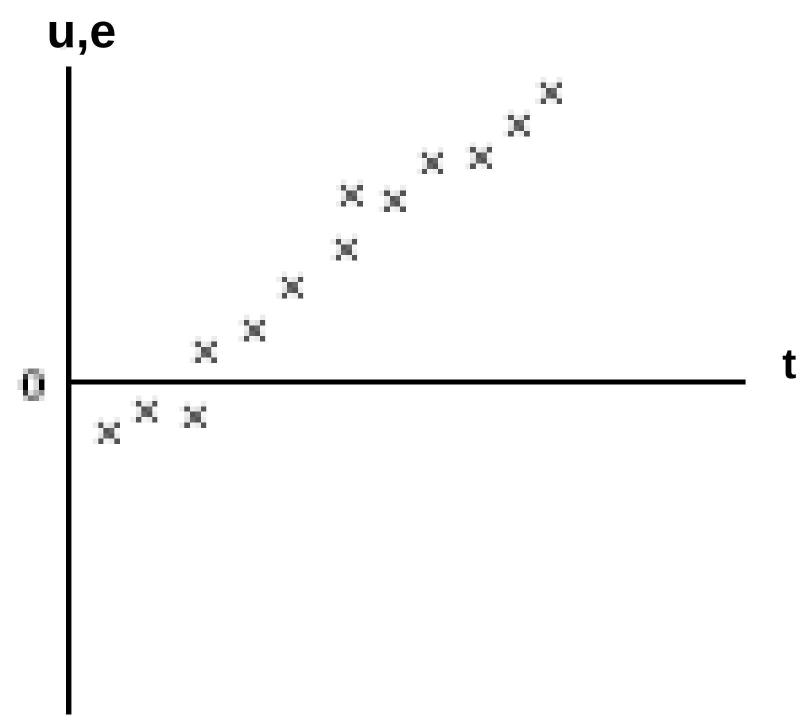
$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$$

$\Rightarrow$Tự tương quan bậc p, AR(p)

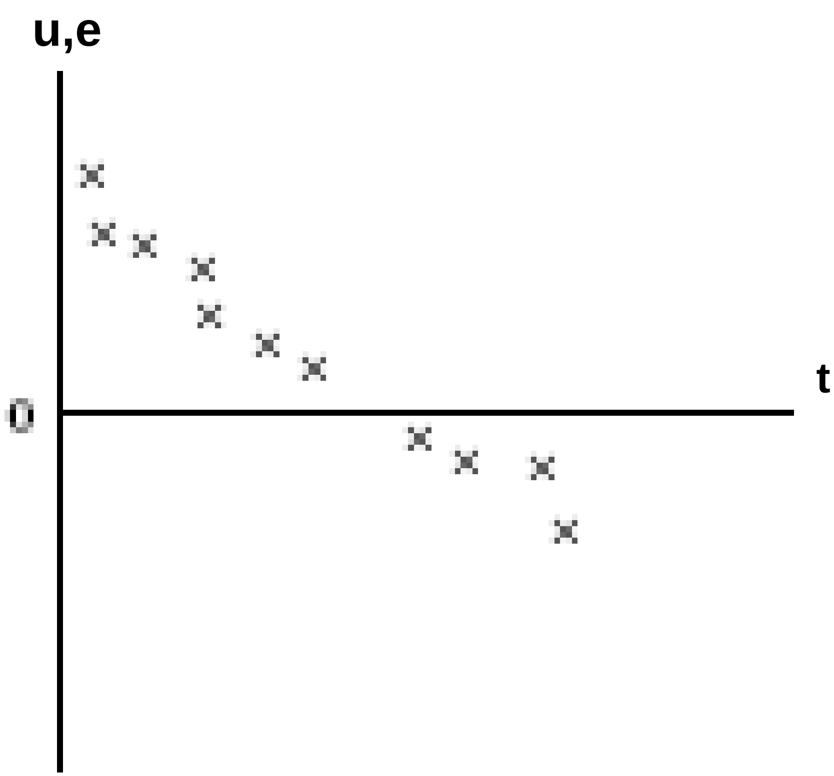
$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_p u_{t-p} + \varepsilon_t$$



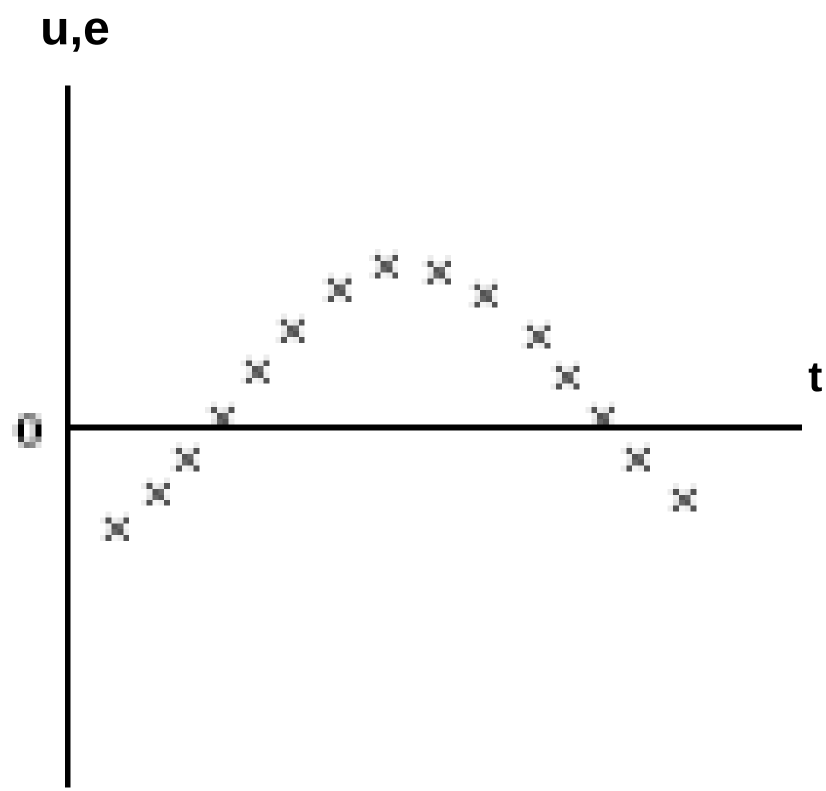
(a)



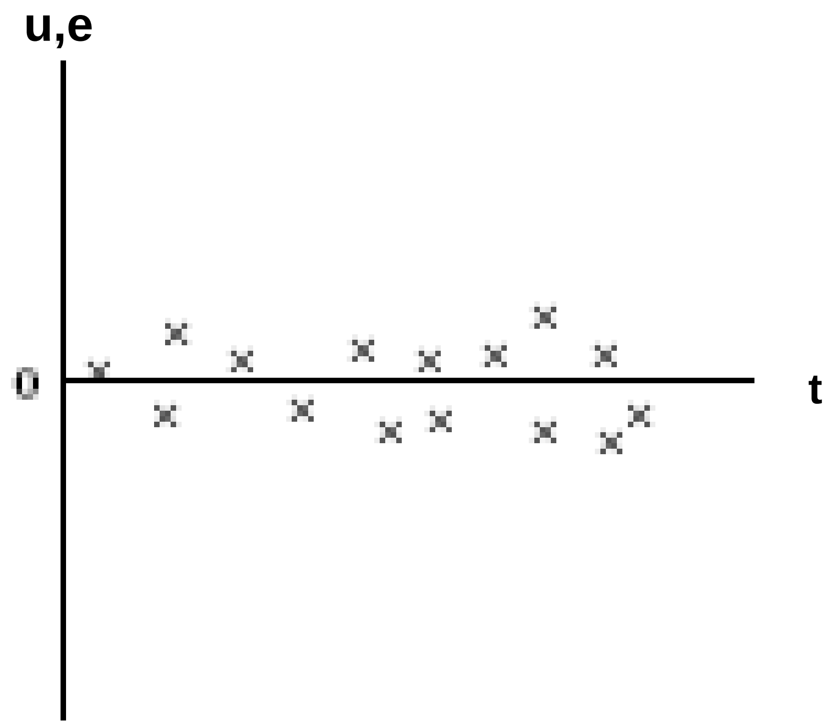
(b)



(c)



(d)


u,e
0
t

I.2. Nguyên nhân của tự tương quan

a. Nguyên nhân khách quan

- Do tính quán tính của các số liệu trong kinh tế
- Trong phân tích hồi qui chuỗi thời gian, mô hình có thể chứa các biến phụ thuộc ở thời kỳ trễ là các biến độc lập. Nếu chúng ta bỏ qua các yếu tố trễ này sẽ là sai số ngẫu nhiên mang tính hệ thống và dẫn tới hiện tượng tự tương quan.

Nguyên nhân của tự tương quan (tiếp)

b. Do yếu tố chủ quan

- Do quá trình xử lý số liệu. Ví dụ quá trình lấy trung bình trượt để làm trơn số liệu, hay quá trình ngoại suy…
- Do mô hình đã bỏ sót 1 hay 1 số biến thích hợp

Mô hình đúng: $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t$

Mô hình hồi qui : $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + v_t$

$\Rightarrow v_t = \beta_4 X_{4t} + u_t$

II. Hậu quả của tự tương quan

- $\hat{\beta}$ vẫn là ước lượng không chệch, nhưng không phải là các ước lượng có phương sai nhỏ nhất.
- $\hat{\sigma}^2$ là ước lượng chệch của σ^2 (thông thường nó ước lượng nhỏ hơn giá trị thực của σ^2).
- Kiểm định T và kiểm định F mất ý nghĩa.
- R^2 tính toán được có thể là độ đo không đáng tin cậy cho R^2 thực
- Không phải các ước lượng thu được từ phương pháp OLS mà các ước lượng thu được từ phương pháp GLS mới là BLUE

III. Cách pháp hiện

1. Dùng đồ thị phần dư

- Ta sử dụng phần dư để đại diện cho nhiễu u_t.

☞ Vẽ đồ thị phần dư theo thời gian

☞ Vẽ đồ thị phần dư chuẩn hoá theo thời gian. Việc chuẩn hoá giúp triệt tiêu đơn vị, do đó, ta có thể so sánh các phần dư chuẩn hoá trong các hồi qui khác nhau với nhau.

$\frac{U_t}{\sigma} \sim N(0,1) \Rightarrow$ với các mẫu lớn, $\frac{e_t}{\hat{\sigma}_t}$ phân phối xấp xỉ N(0,1)

III. Cách pháp hiện (tiếp)

1. Dùng đồ thị phần dư

☞ Vẽ đồ thị phần dư theo các giá trị trễ

Dùng phương pháp OLS ước lượng mô hình xuất phát.

Để phát hiện AR(1): vẽ đồ thị e_t phụ thuộc e_{t-1}

Để phát hiện AR(p): vẽ đồ thị e_t phụ thuộc e_{t-p}

III. Cách pháp hiện (tiếp)

III.2. Kiểm định Durbin-Watson (DW)

Thống kê Durbin-Watson được kí hiệu là d:

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

III.2. Kiểm định Durbin-Watson (DW)

a. Giả thiết của Dubin-Watson

☞ Mô hình hồi qui phải có hệ số chặn

☞ Các biến giải thích là biến phi ngẫu nhiên. Giả thiết này rất khó đáp ứng trong mô hình kinh tế liên quan đến chuỗi thời gian.

☞ Nhiễu u_t được hình thành từ quá trình tự hồi qui bậc 1 ($u_t = \rho u_{t-1} + \varepsilon_t$) nên không thể sử dụng thống kê d pháp hiện tự tương quan bậc cao.

III.2. Kiểm định Durbin-Watson (DW)

☞ Nhiễu u_t được giả thiết phân bố chuẩn. Nếu không, thống kê d sẽ không đáng tin cậy.

☞ Mô hình hồi qui gốc không chứa độc lập là biến trễ của biến phụ thuộc (mô hình tự hồi qui $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \ldots + \beta_k X_{kt} + \gamma Y_{t-1} + u_t$) . Trường hợp ngược lại, giá trị thống kê d thường xấp xỉ 2, chứng tỏ không có tự tương quan ngay cả khi hiện tượng này xuất hiện trong mô hình.

☞ Không có quan sát bị thiếu trong số liệu.

III.2. Kiểm định Durbin-Watson (DW)

b. Xây dựng thống kê d:

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

$\sum e_t^2$ và $\sum e_{t-1}^2$ chỉ khác nhau 1 quan sát nên coi chúng xấp xỉ nhau

III.2. Kiểm định Durbin-Watson (DW)

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

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

III.2. Kiểm định Durbin-Watson (DW)

$$d \approx 2(1-\hat{\rho})$$

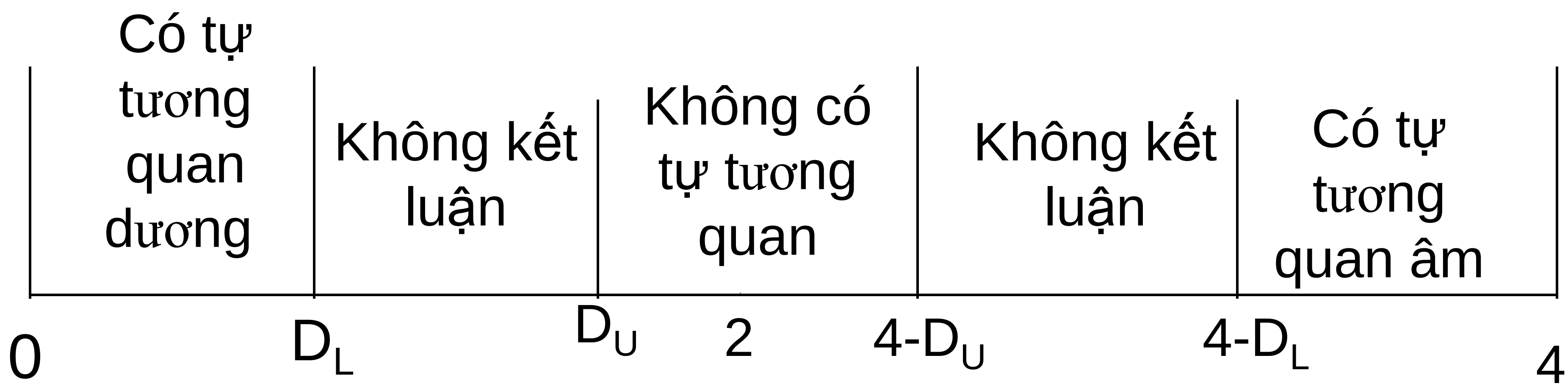
- $\hat{\rho}$ hệ số tương quan bậc nhất của mẫu, và là ước lượng của ρ
- Vì $-1 \leq \rho \leq 1$ nên ta suy ra rằng $0 \leq d \leq 4$

Nếu $\hat{\rho} = 0 \Rightarrow d = 2$: không có tự tương quan

Nếu $\hat{\rho} = 1 \Rightarrow d = 0$: tồn tại tự tương quan dương hoàn hảo

Nếu $\hat{\rho} = -1 \Rightarrow d = 4$: tồn tại tự tương quan âm hoàn hảo

III.2. Kiểm định Durbin-Watson (DW)



III.3. Kiểm định Breusch-Godfrey (BG)

Xét mô hình: $Y_t = \beta_1 + \beta_2 X_t + u_t$

$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_p u_{t-p} + \varepsilon_t$ AR(p)

ε_t thoả mãn các giả thiết của OLS

H_0: $\rho_1 = \rho_2 = \ldots = \rho_p = 0$

(không có tự tương quan)

H_1: $\exists \rho_i \neq 0$ (Có tự tương quan)

III.3. Kiểm định Breusch-Godfrey (BG)

- Bước 1: Dùng OLS ước lượng mô hình xuất phát, thu được e_t
- Bước 2: Ước lượng mô hình:

$$e_t = \beta_0 + \beta_1 X_t + \rho_1 e_{t-1} + \rho_2 e_{t-2} + \ldots + \rho_p e_{t-p} + v_t$$

Thu được R^2 (kích thước mẫu chỉ còn n-p).

- Bước 3: Với n đủ lớn, $(n-p)R^2 \sim \chi^2(p)$

Kiểm định giả thiết:

$(n-p)R^2 > \chi^2_\alpha(p)$: Bác bỏ giả thiết H_0

$(n-p)R^2 \le \chi(p)$: Không đủ cơ sở bác bỏ giả thiết H_0

III.3. Kiểm định Breusch-Godfrey (BG)

Chú ý trong thực hành kiểm định BG:

- Có thể sử dụng tiêu chuẩn F để kiểm định
- BG có thể áp dụng đối với mô hình gốc có chứa biến giải thích là biến trễ của biến phụ thuộc (Y_{t-1}, Y_{t-2}…)
- Có thể phát hiện tự tương quan bậc cao (p>1)
- Hạn chế của kiểm định BG là việc xác định độ dài của trễ (p)

IV. Khắc phục hiện tượng tự tương quan

IV.1.Trường hợp ρ đã biết

- Xét mô hình hồi qui 2 biến:

  $Y_t = \beta_1 + \beta_2 X_t + u_t$ (7.1)

- Giả sử có tương quan bậc 1:

  $u_t = \rho_1 u_{t-1} + \varepsilon_t$ (7.6)

  ε_t thoả mãn các giả thiết của OLS

- Ta có: $Y_{t-1} = \beta_1 + \beta_2 X_{2t-1} + u_{t-1}$ (7.7)

  Nhân 2 vế phương trình (7.7) với ρ:

IV. Khắc phục hiện tượng tự tương quan

$\rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_{t-1} + \rho u_{t-1}$ (7.8)

- Trừ (7.1) cho (7.8) ta được:

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

$$Y_t^* = \beta_1^* + \beta_2^* X_t^* + \varepsilon_t \qquad (7.9)$$

(7.9): phương trình sai phân tổng quát

IV. Khắc phục hiện tượng tự tương quan

(7.9) có sai số ngẫu nhiên thoả mãn các giả thiết của phương pháp OLS $\Rightarrow$ không có tự tương quan. Đây chính là việc áp dụng OLS cho hàm chuyển đổi hay phương pháp GLS.

III.2. Trường hợp ρ chưa biết

a. Dùng sai phân cấp 1

- Xét mô hình hồi qui 2 biến:

  $Y_t = \beta_1 + \beta_2 X_t + u_t$ (7.1)

- Giả sử có tương quan bậc 1:

  $u_t = \rho_1 u_{t-1} + \varepsilon_t$ (7.6)

  ε_t thoả mãn các giả thiết của OLS

III.2. Trường hợp ρ chưa biết

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \varepsilon_t \quad (7.10)$$

$$-1 \leq \rho \leq 1 \qquad \rho \neq 0$$

- Nếu ρ = 1 $\Rightarrow$ $Y_t - Y_{t-1} = 0 + \beta_2(X_t - X_{t-1}) + \varepsilon_t$

$$\Delta Y_t = \beta_2 \Delta X_t + \varepsilon_t \quad (7.11)$$

(7.11): phương trình sai phân không có hệ số chặn

III.2. Trường hợp ρ chưa biết

- Nếu ρ = -1 $\Rightarrow Y_t + Y_{t-1} = 2\beta_1 + \beta_2 \left(X_t + X_{t-1} \right) + \varepsilon_t$

$$(7.12) \quad \frac{Y_t + Y_{t-1}}{2} = \beta_1 + \beta_2 \frac{X_t + X_{t-1}}{2} + v_t$$

(7.12) hồi quy trung bình 2 thời kỳ liên tiếp của biến phụ thuộc.

$\Rightarrow$ Từ các giả thiết về ρ, ta ước lượng được các hệ số mà không cần biết giá trị thực của ρ

III.2. Trường hợp ρ chưa biết

b. Dùng thống kê d

Sau hồi qui, các phần mềm thường cho giá trị của thống kê Durbin-Watson, d=2(1- $\hat{\rho}$)

$$\hat{\rho} = 1 - \frac{d}{2}$$

Sử dụng $\hat{\rho}$ như ước lượng của ρ và thay vào phương trình sai phân tổng quát

III.2. Trường hợp ρ chưa biết

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

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

Ước lượng phương trình sai phân tổng quát, ta được các ước lượng của các hệ số $\hat{\beta}_1^*, \hat{\beta}_2^*$

$$\hat{\beta}_1 = \frac{\hat{\beta}_1^*}{1-\hat{\rho}} \qquad \hat{\beta}_2 = \hat{\beta}_2^*$$

III.2. Trường hợp ρ chưa biết

c. Ước lượng ρ dựa trên phần dư

- $u_t = \rho_1 u_{t-1} + \varepsilon_t$
  Để thu được ước lượng của ρ, ta hồi qui e_t theo e_{t-1}
- Bước 1: ước lượng mô hình gốc bằng OLS $\Rightarrow e_t$
- Bước 2: hồi qui $e_t = \rho e_{t-1} + \varepsilon_t \Rightarrow \hat{\rho}$
- $\hat{\rho}$ thu được bằng phương pháp này không khác $\hat{\rho}$ thu được khi dựa trên thống kê d

III.2. Trường hợp ρ chưa biết

d. Dùng thủ tục Cochrane – Ocutt

Phương pháp này ước lượng ρ lặp lại nên còn được gọi là phương pháp lặp Cochrane – Ocutt

- Bước 1: Ước lượng mô hình

$$Y_t = \beta_1 + \beta_2 X_t + U_t \qquad (7.6)$$

$\Rightarrow e_t$

III.2. Trường hợp ρ chưa biết

- Bước 2: Ước lượng $e_t = \rho e_{t-1} + v_t \Rightarrow \hat{\rho}$
- Bước 3: Ước lượng phương trình sai phân tổng quát (7.7)

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

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

$$\Rightarrow \quad \hat{\beta}_1^*, \hat{\beta}_2^* \Rightarrow \hat{\beta}_1 = \frac{\hat{\beta}_1^*}{1-\hat{\rho}} \qquad \hat{\beta}_2 = \hat{\beta}_2^*$$

III.2. Trường hợp ρ chưa biết

- Bước 4. Vì chưa biết $\hat{\rho}$ thu được từ (7.7) có phải là ước lượng tốt nhất của ρ hay không nên ta thay $\hat{\rho}$ vào phương trình hồi qui gốc (7.6) $\Rightarrow e_t^*$

$$e_t^* = Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_t$$

III.2. Trường hợp ρ chưa biết

- Quay trở lại bước 2 và cứ tiếp tục cho đến khi hai ước lượng kế tiếp nhau của ρ khác nhau không đáng kể, chẳng hạn bé hơn 0,01 hoặc 0,05.
- Trong thực tế, dùng 3 đến 4 bước lặp là đủ.
- Chú ý: ở bước hai, mô hình hồi qui có thể là AR(1) hoặc tự hồi qui ở các bậc cao hơn AR(2), AR(3) …