

TỰ TƯƠNG QUAN



I. NGUYÊN NHÂN VÀ HẬU QUẢ

I.1

Hiện tượng tự tương quan

I.2

Nguyên nhân hiện tượng tự tương quan

I.3

Hậu quả của hiện tượng tự tương quan

I.1

Hiện tượng tự tương quan

Mô hình tuyến tính cổ điển giả thiết:

$$Cov(U_i; U_j) = E(U_i U_j) = 0 \quad \forall i \neq j$$

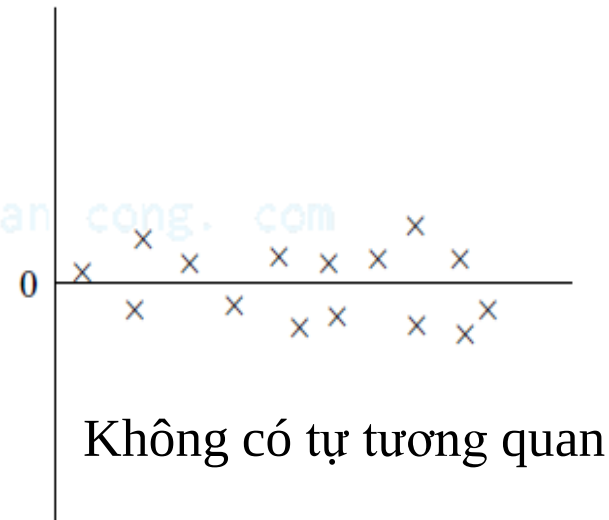
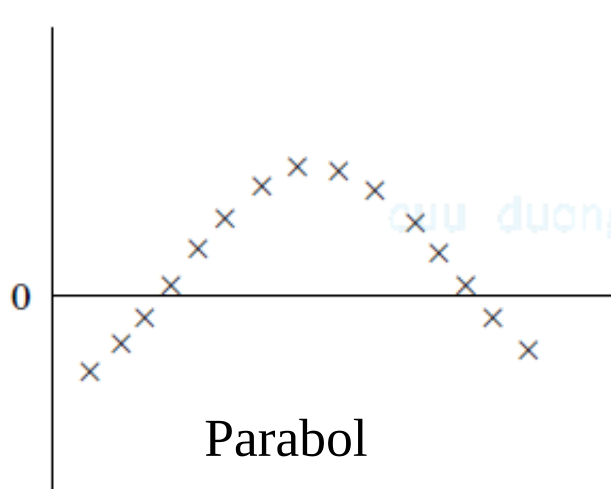
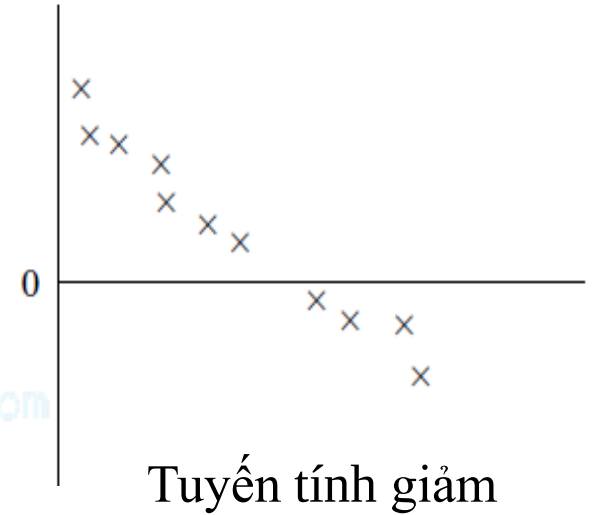
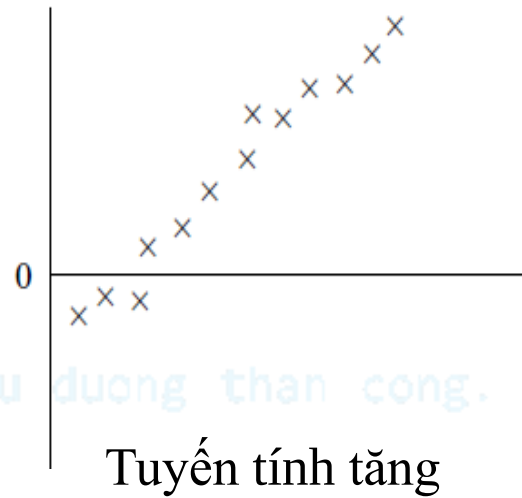
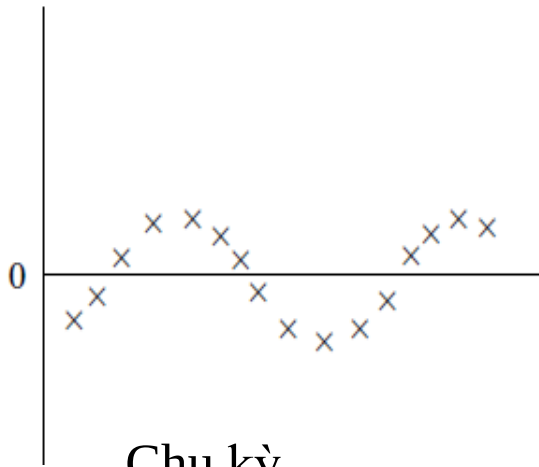
Nói cách khác, nhiễu gắn với một quan sát nào đó không bị ảnh hưởng bởi nhiễu gắn với một quan sát khác, nếu có giá trị cá biệt nào đó của $U_i > 0$ (< 0) thì không có nghĩa các giá trị khác cũng > 0 (< 0)

Hiện tượng tự tương quan xảy ra khi :

$$Cov(U_i; U_j) = E(U_i U_j) \neq 0 \quad (i \neq j)$$

ĐỒ THỊ TỰ TƯƠNG QUAN

Trục Ox: thời gian; Trục Oy: phần dư



Các giả định về cách sản sinh nhiễu (trong chuỗi thời gian)

Giả định 1: $U_t = \rho U_{t-1} + \epsilon_t$

Trong đó:

ρ : hệ số tự tương quan bậc 1 (hay hệ số tự hồi quy bậc 1)

ϵ_t : nhiễu ngẫu nhiên thỏa mãn mọi giả thiết của mô hình hồi quy tuyến tính cổ điển.

-> U_t tuân theo lược đồ tự hồi quy bậc 1, kí hiệu AR(1)

Các giả định về cách sản sinh nhiễu (trong chuỗi thời gian)

Giả định 2: $U_t = \rho_1 U_{t-1} + \rho_2 U_{t-2} + \dots + \rho_p U_{t-p} + \epsilon_t$

Trong đó:

ρ_j : hệ số tự hồi quy bậc j ($j=1..p$)

ϵ_t : nhiễu ngẫu nhiên thỏa mãn mọi giả thiết của mô hình hồi quy tuyến tính cổ điển.

-> U_t tuân theo lược đồ tự hồi quy bậc p , kí hiệu $AR(p)$



I.2

Nguyên nhân hiện tượng tự tương quan

- Quán tính – tính chất phổ biến của các đại lượng kinh tế quan sát theo thời gian
- Hiện tượng mạng nhện
- Tính chất “trễ” của các đại lượng kinh tế
- Phương pháp (kỹ thuật) thu thập và xử lý số liệu
- Sai lầm khi lập mô hình: bỏ biến (không đưa biến vào mô hình), dạng hàm sai...



I.3

Hậu quả của hiện tượng tự tương quan

- Các ước lượng BPNN là các ước lượng tuyến tính, không chệch nhưng không phải là ước lượng hiệu quả nữa.
- Các ước lượng của các phương sai là chệch và thông thường là thấp hơn giá trị thực, do đó giá trị của thống kê T được phóng đại lên nhiều lần so với giá trị thực của nó
- Thống kê T và F không còn có ý nghĩa về mặt thống kê nên việc kiểm định các giả thiết thống kê không còn đáng tin cậy nữa
- Các dự báo dựa trên các ước lượng BPNN không còn tin cậy nữa.



II. PHÁT HIỆN KHUYẾT TẬT

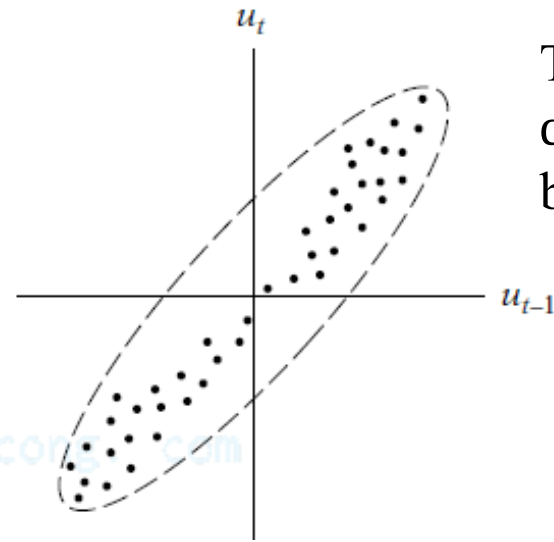
II.1 Phương pháp đồ thị

II.2 Kiểm định Durbin - Watson

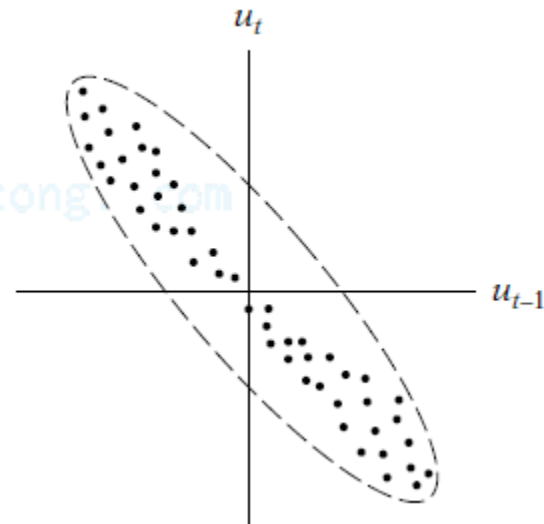
II.3 Kiểm định Breusch – Godfrey (BG)

II.1 Phương pháp đồ thị

- Quan sát đồ thị phần dư theo thời gian. Nếu phần dư biến đổi phụ thuộc vào thời gian thì tồn tại tự tương quan (xem lại một số đồ thị tự tương quan).
- Quan sát đồ thị e_t đối với e_{t-s} ($s \geq 1$). Nếu có mối liên hệ giữa 2 đại lượng này thì tồn tại tự tương quan bậc s .



Tự tương quan dương, bậc 1



Tự tương quan âm, bậc 1



II.2 Kiểm định Durbin - Watson

Đây là phương pháp kiểm định có ý nghĩa nhất để phát hiện ra tương quan chuỗi.

Các giả thiết cơ sở:

- Mô hình hồi quy gốc phải có hệ số chặn.
- Các biến giải thích có giá trị xác định.
- Các nhiễu được sản sinh ra từ lược đồ tự hồi quy bậc 1:

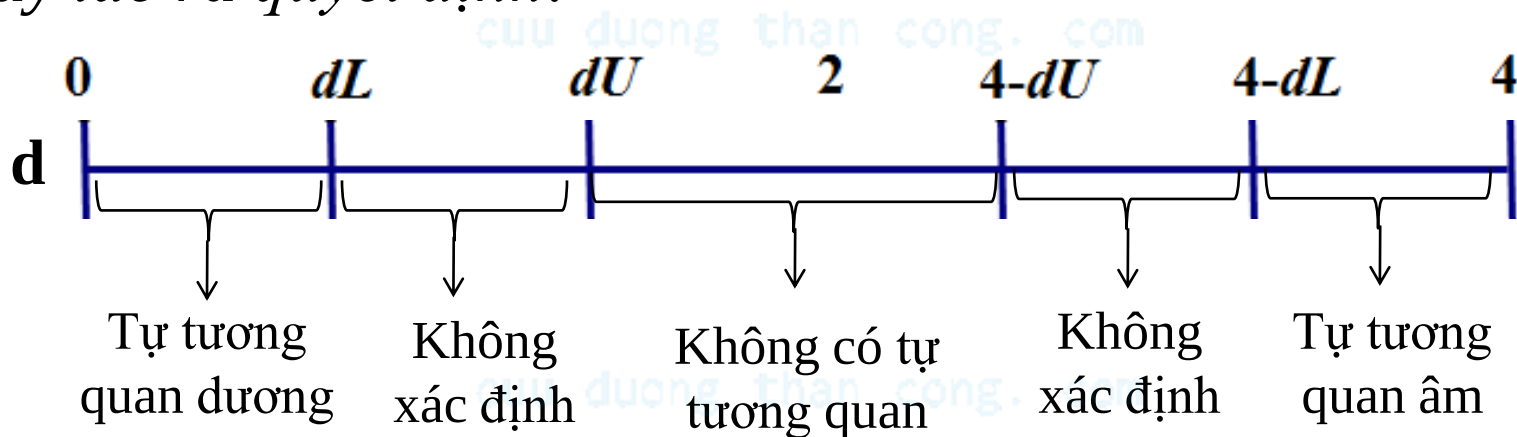
$$U_t = \rho U_{t-1} + \epsilon_t$$

- Mô hình hồi quy không chứa biến giải thích là giá trị trễ của biến phụ thuộc (không chứa biến Y_{t-s}).
- Không có các quan sát bị mất trong dữ liệu.

Thống kê d được định nghĩa:
$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \approx 2(1 - \hat{\rho})$$

trong đó: $\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2}$: ước lượng của ρ trong lược đồ tự hồi quy bậc 1. Vì $-1 \leq \hat{\rho} \leq 1 \rightarrow 0 \leq d \leq 4$

Quy tắc ra quyết định:



Các giá trị d_L , d_U được tra bảng, phụ thuộc α , n và số biến giải thích $(k-1)$ có trong mô hình.



II.3 Kiểm định Breusch – Godfrey (BG)

Xét mô hình: $Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + U_t$

Giả sử rằng: $U_t = \rho_1 U_{t-1} + \rho_2 U_{t-2} + \dots + \rho_p U_{t-p} + \epsilon_t$

Không tồn tại tự tương quan ở bất kỳ bậc nào $\leftrightarrow$ giả thuyết $H_0: \rho_1 = \rho_2 = \dots = \rho_p = 0$

Bước 1: Ước lượng mô hình gốc nhận được các phần dư.

Bước 2: Ước lượng mô hình sau thu được R^2 :

$$e_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \rho_1 e_{t-1} + \rho_2 e_{t-2} + \dots + \rho_p e_{t-p} + v_t$$

$$e_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \rho_1 e_{t-1} + \rho_2 e_{t-2} + \dots + \rho_p e_{t-p} + v_t$$

Bước 3: Kiểm định $H_0: \rho_1 = \rho_2 = \dots = \rho_p = 0$. 2 cách:

- Kiểm định χ^2 : Nếu H_0 đúng thì $\chi^2 = (n-p)R^2 \sim \chi^2(p)$. Miền bác bỏ $H_0: \chi^2 > \chi^2(p)$: ít nhất tồn tại tự tương quan ở một bậc nào đó.

- Kiểm định F: Ước lượng mô hình sau thu được R_1^2 :

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

Tiêu chuẩn kiểm định: $F = \frac{R^2 - R_1^2}{1 - R^2} \cdot \frac{n - k}{m}$

Nếu H_0 đúng thì $F \sim F(p, n-k)$.

Miền bác bỏ $H_0: F > F(p, n-k)$



III. BIỆN PHÁP KHẮC PHỤC

III.1 Khi cấu trúc tự tương quan đã biết

III.2 Khi cấu trúc tự tương quan chưa biết

cuu duong than cong. com

III.1 *Khi cấu trúc tự tương quan đã biết*

Xét mô hình hồi quy đơn: $Y_t = \beta_1 + \beta_2 X_t + U_t$ mà ta cho rằng có hiện tượng tự tương quan.

Giả sử rằng: $U_t = \rho U_{t-1} + \epsilon_t$ ($|\rho| < 1$); ϵ_t thỏa mãn các giả thiết OLS, ρ đã biết.

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

Khi đó: $Y_t^* = \beta_1^* + \beta_2^* X_t^* + \epsilon_t$ thỏa mãn các điều kiện của OLS $\rightarrow$ Phương trình sai phân tổng quát

Các bước khắc phục

Bước 1: Lập các biến mới:

$$Y_t^* = Y_t - \rho Y_{t-1}; X_t^* = X_t - \rho X_{t-1}$$

Bước 2: Hồi quy mô hình sau, xác định $\hat{\beta}_1^*; \hat{\beta}_2^*$:

$$Y_t^* = \beta_1^* + \beta_2^* X_t^* + \epsilon_t \quad (*)$$

Bước 3: Tính

$$\hat{\beta}_1 = \frac{\hat{\beta}_1^*}{(1 - \rho)}; \hat{\beta}_2 = \hat{\beta}_2^*$$

Bước 4: SRF của mô hình gốc cho bởi: $\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_2 X_t$

Chú ý: - Việc có hay không hệ số chặn của hồi quy () phụ thuộc vào mô hình gốc có hệ số chặn hay không.*

- Mất 1 quan sát so với mô hình gốc.



III.2 *Khi ρ chưa biết*

Ý tưởng: ước lượng ρ rồi lặp lại các thao tác khắc phục như trường hợp đã biết cấu trúc tự tương quan. Áp dụng trên các mẫu lớn, cẩn thận khi áp dụng trên các mẫu bé.



Ước lượng ρ dựa trên [thống kê d](#) Durbin - Watson

[cuuduongthancong.com](#)



Phương pháp [sai phân](#) cấp 1



Thủ tục lặp [Cochrane – Orcutt](#) để ước lượng ρ

[cuuduongthancong.com](#)



Phương pháp [Durbin – Watson](#) 2 bước



Ước lượng ρ dựa trên thống kê d

Durbin - Watson

Mô hình gốc: $Y_t = \beta_1 + \beta_2 X_t + U_t$ ($U_t = \rho U_{t-1} + \epsilon_t$: AR(1))

Bước 1: Xác định $\hat{\rho} \approx 1 - \frac{d}{2}$ (d: thống kê của kiểm định Durbin – Watson)

Bước 2: Tương tự trường hợp ρ đã biết:

- Lập các biến mới: $Y_t^* = Y_t - \hat{\rho} Y_{t-1}$; $X_t^* = X_t - \hat{\rho} X_{t-1}$

- Hồi quy mô hình: $Y_t^* = \beta_1^* + \beta_2^* X_t^* + \epsilon_t \rightarrow \hat{\beta}_1^*; \hat{\beta}_2^*$

- Tính $\hat{\beta}_1 = \frac{\hat{\beta}_1^*}{(1 - \hat{\rho})}$; $\hat{\beta}_2 = \hat{\beta}_2^*$

-SRF của mô hình gốc cho bởi: $\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_2 X_t$





Phương pháp sai phân cấp 1

Điều kiện áp dụng: $\rho = 1 \Leftrightarrow d \approx 0$ (tương quan dương, phổ biến) (Madalad: $d < R^2$)

Mô hình gốc: $Y_t = \beta_1 + \beta_2 X_t + U_t$ (U_t theo sơ đồ AR(1))

Bước 1: Lập các biến mới: $\Delta Y_t = Y_t - Y_{t-1}$; $\Delta X_t = X_t - X_{t-1}$
(sai phân cấp 1 của Y_t, X_t)

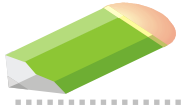
Bước 2: Hồi quy mô hình: $\Delta Y_t = \beta_2 \Delta X_t + \epsilon_t$ thu được $\hat{\beta}_2$

Bước 3: SRF của mô hình gốc cho bởi: $\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_2 X_t$

$$= \hat{Y}_0 - \hat{\beta}_2 X_0 + \hat{\beta}_2 X_t \approx Y_0 - \hat{\beta}_2 X_0 + \hat{\beta}_2 X_t$$

Chú ý: mô hình của sai phân cấp 1 không có hệ số chặn.





Thủ tục lặp Cochrane – Orcutt để ước lượng ρ

Mô hình gốc: $Y_t = \beta_1 + \beta_2 X_t + U_t$ (với $U_t = \rho U_{t-1} + \epsilon_t$)

Bước 1: Ước lượng mô hình gốc $\rightarrow e_t$

Bước 2: Bắt đầu vòng lặp:

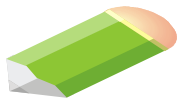
- Ước lượng hồi quy: $e_t = \rho e_{t-1} + V_t \rightarrow \hat{\rho}$
- Lập các biến mới: $Y_t^* = Y_t - \hat{\rho} Y_{t-1}$; $X_t^* = X_t - \hat{\rho} X_{t-1}$
- Hồi quy mô hình: $Y_t^* = \beta_1^* + \beta_2^* X_t^* + \epsilon_t$ thu được $\hat{\beta}_1^*$; $\hat{\beta}_2^*$
- Tính $\hat{\beta}_1 = \frac{\hat{\beta}_1^*}{(1 - \hat{\rho})}$; $\hat{\beta}_2 = \hat{\beta}_2^*$
- Tính $e_t^* = Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_t$ bắt đầu lại vòng lặp với e_t^*

Thoát khỏi vòng lặp khi các ước lượng kế tiếp nhau của khác nhau một lượng rất nhỏ ($<1\%$ hoặc 5%). Thực tế vòng lặp thường dừng sau 3-4 bước.

Bước 3: SRF của mô hình gốc cho bởi: $\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_2 X_t$ với $\hat{\beta}_1; \hat{\beta}_2$ là giá trị ước lượng cuối cùng của $\beta_1; \beta_2$ khi thoát ra khỏi vòng lặp.

Đôi khi trong thực hành, người ta chỉ tiến hành 1 lần vòng lặp cho kết quả hoàn toàn tương tự với kết quả thu từ thủ tục lặp kỹ lưỡng ở trên -> Thủ tục Cochrane – Orcutt hai bước.





Phương pháp Durbin – Watson 2 bước

Mô hình gốc: $Y_t = \beta_1 + \beta_2 X_t + U_t$ (với $U_t = \rho U_{t-1} + \epsilon_t$)

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

Bước 1: Ước lượng mô hình:

$$Y_t = \beta_1^* + \beta_2^* X_t - \beta_3^* X_{t-1} + \beta_4^* Y_{t-1} + \epsilon_t \rightarrow \hat{\rho} = \hat{\beta}_4^*$$

Bước 2: Tương tự trường hợp ρ đã biết:

- Lập các biến mới: $Y_t^* = Y_t - \hat{\rho} Y_{t-1}$; $X_t^* = X_t - \hat{\rho} X_{t-1}$
- Hồi quy mô hình: $Y_t^* = \beta_1^* + \beta_2^* X_t^* + \epsilon_t$ thu được $\hat{\beta}_1^*$; $\hat{\beta}_2^*$
- Tính $\hat{\beta}_1 = \frac{\hat{\beta}_1^*}{(1 - \hat{\rho})}$; $\hat{\beta}_2 = \hat{\beta}_2^*$
- SRF của mô hình gốc cho bởi: $\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_2 X_t$

Dự báo dựa vào phần dư

Giả sử chúng ta chỉ có dữ liệu cho đến thời điểm t .

+ Tại thời điểm $t+1$: $\hat{Y}_{t+1} = \hat{\beta}_1 + \hat{\beta}_2 X_{t+1} + \hat{\rho} e_t$

với $\hat{\beta}_1; \hat{\beta}_2$ là các ước lượng OLS, $\hat{\rho}$ thu được bằng các cách ở trên và $e_t = \hat{Y}_t - Y_t$

+ Tại thời điểm $t+2$: (do e_{t+1} chưa biết)

$$\hat{Y}_{t+2} = \hat{\beta}_1 + \hat{\beta}_2 X_{t+2} + \hat{\rho} e_{t+1} \approx \hat{\beta}_1 + \hat{\beta}_2 X_{t+2} + \hat{\rho}^2 e_t$$

+ Tại thời điểm $t+k$:

$$\hat{Y}_{t+k} = \hat{\beta}_1 + \hat{\beta}_2 X_{t+k} + \hat{\rho} e_{t+k-1} \approx \hat{\beta}_1 + \hat{\beta}_2 X_{t+k} + \hat{\rho}^k e_t$$