

Bài giảng 1:

Phân tích hồi quy cơ bản với dữ
liệu chuỗi thời gian

Bản chất của dữ liệu chuỗi thời gian



- Các quan sát có thứ tự xuất hiện theo thời gian; không thể sắp xếp lại thứ tự các quan sát này một cách ngẫu nhiên
- Những đặc điểm tiêu biểu: tương quan chuỗi/các quan sát không độc lập với nhau (**serial correlation/nonindependence of observations**)
- Tính ngẫu nhiên của dữ liệu chuỗi thời gian thể hiện như thế nào?
 - Kết quả của các biến kinh tế (vd. GNP) không thể đoán trước; vì vậy phải xem các biến này là biến ngẫu nhiên (**random variables**)
 - Chuỗi thời gian là thứ tự của các biến ngẫu nhiên (= quá trình ngẫu nhiên) (**stochastic processes**)
 - Tính ngẫu nhiên không phát sinh từ quá trình chọn mẫu (**sampling**)
 - “Mẫu” (**sample**)= đồ thị thực của dữ liệu chuỗi trong nhiều đồ thị có thể xảy ra do quá trình ngẫu nhiên

Nhận xét



- Đối với dữ liệu chuỗi thời gian, thứ tự có tính tự nhiên; các quan sát phải xuất hiện theo thứ tự thời gian. Cụ thể, dữ liệu chuỗi thời gian là một chuỗi các quan sát diễn ra theo thời gian.
- Mặc dù thời gian có tính liên tục (**continuous**), khoảng cách (**interval**) giữa các quan sát lại rời rạc (**discrete**). Thông thường khoảng cách giữa những lần quan sát được chia đều (**equally spaced**). Khoảng cách này có thể là theo từng năm, từng quý, từng tháng, từng ngày, từng giờ, v.v.
- Ví dụ:
 - (a) Tổng Sản phẩm Nội địa (GDP): dữ liệu theo quý hoặc theo năm
 - (b) Tỷ lệ lạm phát hoặc thất nghiệp: dữ liệu hàng tháng, hàng quý, hàng năm
 - (c) Chỉ số chứng khoán, tỷ giá hối đoái: dữ liệu hàng ngày, hàng tuần, hàng tháng, hàng quý, hàng năm
 - (d) Lợi nhuận doanh nghiệp, cổ tức, v.v.: dữ liệu hàng quý, hàng năm
 - (e) Nhiệt độ: dữ liệu hàng giờ

Nhận xét



- (i) Tần suất (**frequency**) của dữ liệu của tất cả các biến phải giống nhau khi chạy mô hình hồi quy.
- (ii) Cộng dữ liệu của 4 quý trong một năm bất kỳ = dữ liệu năm
- (iii) Tương tự, nếu có dữ liệu ngày, có thể đổi tần suất của dữ liệu
 - Ví dụ, có dữ liệu tỷ giá hối đoái theo ngày có thể đổi thành dữ liệu theo tháng
- (iv) Giả sử bạn có dữ liệu theo quý. Làm thế nào để đổi thành dữ liệu theo tháng?
 - Nội suy bậc 3 (**cubic spline** interpolation)

Ví dụ



■ Ví dụ: Tỷ lệ lạm phát và thất nghiệp Hoa Kỳ giai đoạn 1948-2003

TABLE 10.1 Partial Listing of Data on U.S. Inflation and Unemployment Rates, 1948–2003

Year	Inflation	Unemployment
1948	8.1	3.8
1949	−1.2	5.9
1950	1.3	5.3
1951	7.9	3.3
.	.	.
.	.	.
.	.	.
1998	1.6	4.5
1999	2.2	4.2
2000	3.4	4.0
2001	2.8	4.7
2002	1.6	5.8
2003	2.3	6.0

© Cengage Learning, 2013

← Ở đây chỉ số hai chuỗi thời gian. Có nhiều biến mà đồ thị theo thời gian của các biến này có thể quan sát cùng lúc.

Phân tích chuỗi dữ liệu thời gian là mô hình thể hiện sự phụ thuộc của một biến vào quá khứ của biến đó và giá trị hiện tại và quá khứ của những biến khác

Ví dụ về các mô hình hồi quy chuỗi thời gian



■ Mô hình tĩnh (**static models**)

- Trong mô hình chuỗi thời gian tĩnh, giá trị hiện tại của một biến là kết quả của giá trị hiện tại của các biến giải thích

■ Ví dụ về mô hình tĩnh

$$inf_t = \beta_0 + \beta_1 unem_t + u_t$$

Tỉ lệ thất nghiệp và lạm phát có mối quan hệ cùng kỳ (= Đường Phillips)

$$mrdrte_t = \beta_0 + \beta_1 convrte_t + \beta_2 unem_t + \beta_3 yngmle_t + u_t$$

Tỉ lệ giết người hiện tại sẽ được quyết định bởi tỉ lệ kết án, tỉ lệ thất nghiệp và tỉ lệ thanh niên nam trên tổng số dân hiện tại.

Ví dụ về các mô hình hồi quy chuỗi thời gian (tt.)



- **Mô hình phân phối trễ hữu hạn (Finite distributed lag models)**
 - Trong mô hình phân phối trễ hữu hạn, biến giải thích có thể tác động lên biến phụ thuộc với độ trễ về thời gian (a time lag)
- **Ví dụ về mô hình phân phối trễ hữu hạn**
 - Tỷ lệ sinh sản sẽ phụ thuộc vào giá trị thuế của một đứa trẻ, nhưng vì nguyên nhân sinh học và hành vi, tác động này sẽ có độ trễ

$$gfr_t = \alpha_0 + \delta_0 pe_t + \delta_1 pe_{t-1} + \delta_2 pe_{t-2} + u_t$$

Số trẻ em sinh ra trên
1000 phụ nữ trong
năm t

Tỷ lệ miễn
thuế trong
năm t

Tỷ lệ miễn thuế
trong năm t-1

Tỷ lệ miễn thuế
trong năm t-2

Ví dụ về các mô hình hồi quy chuỗi thời gian (tt.)



- **Giải thích hiệu ứng trong mô hình phân phối trễ hữu hạn**

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \cdots + \delta_q z_{t-q} + u_t$$

- **Ảnh hưởng của một cú sốc trong quá khứ lên giá trị hiện tại của biến phụ thuộc**

$$\frac{\partial y_t}{\partial z_{t-s}} = \delta_s$$

Ảnh hưởng của cú sốc tạm thời (**transitory shock**):

Nếu xảy ra cú sốc tạm thời một lần duy nhất trong quá khứ, biến phụ thuộc sẽ thay đổi tạm thời, mức độ thay đổi này bằng hệ số của biến giải thích tương ứng.

$$\frac{\partial y_t}{\partial z_{t-q}} + \cdots + \frac{\partial y_t}{\partial z_t} = \delta_1 + \cdots + \delta_q$$

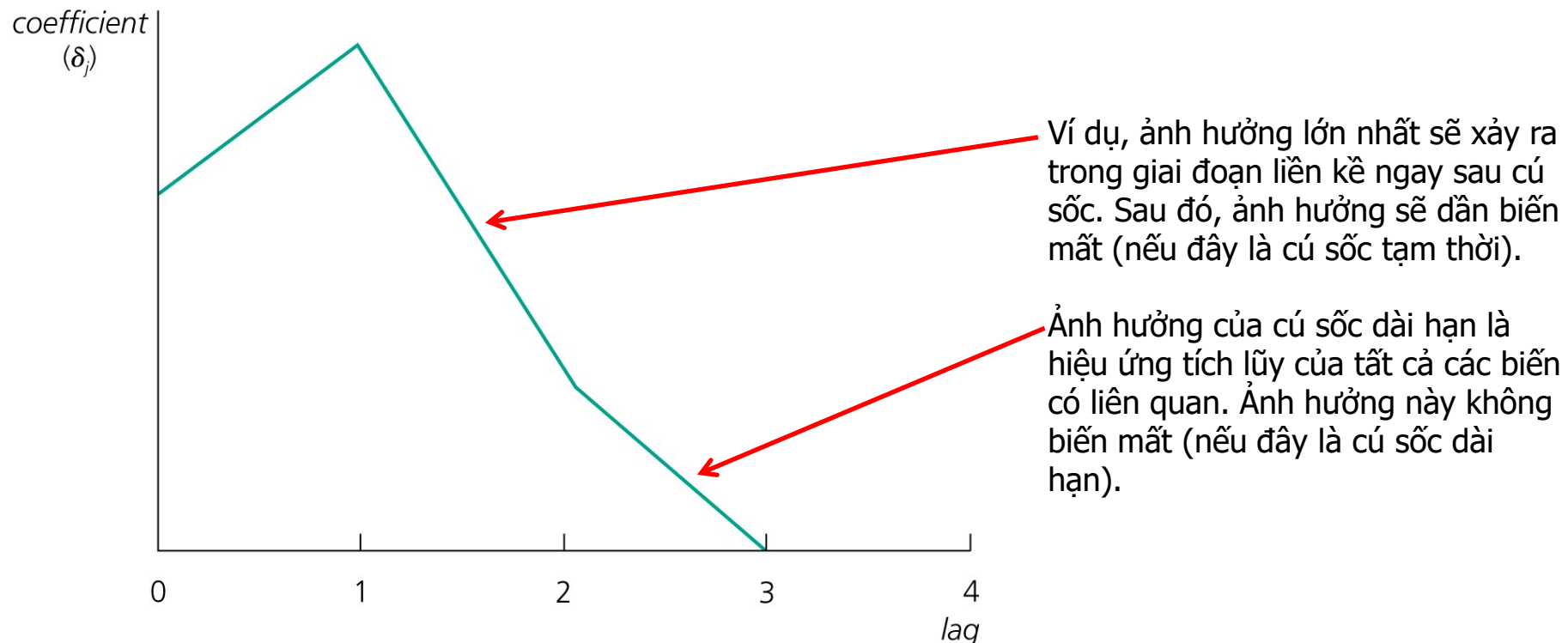
Ảnh hưởng của cú sốc vĩnh viễn (**permanent shock**):

Nếu xảy ra cú sốc vĩnh viễn trong quá khứ, vd. biến giải thích tăng lên một đơn vị trong dài hạn, ảnh hưởng lên biến phụ thuộc sẽ là ảnh hưởng tích lũy của tất cả các biến giải thích có liên quan. Đây sẽ là ảnh hưởng dài hạn lên biến phụ thuộc.

Ví dụ về các mô hình hồi quy chuỗi thời gian (tt.)



■ Minh họa đồ thị của ảnh hưởng có độ trễ (**lagged effects**)



Một số điểm khác biệt giữa dữ liệu chéo và dữ liệu chuỗi thời gian



	Dữ liệu chéo (cross-sectional data)	Dữ liệu chuỗi thời gian (time series data)
Thời gian	Không có yếu tố thời gian	Yếu tố thời gian rất quan trọng.
Ký hiệu	$i = 1, 2, 3 \dots n$ i đại diện cho những cá nhân, người lao động, doanh nghiệp, thành phố, v.v. khác nhau	$t = 1, 2, 3 \dots n$ t đại diện cho những giai đoạn thời gian khác nhau
Tương quan (correlation)	Không có tương quan giữa các quan sát. Vì vậy, chúng ta có thể sắp xếp thứ tự của các quan sát một cách ngẫu nhiên.	Các quan sát của một biến thường có tương quan tự động (auto-correlation) với nhau. Không thể sắp xếp lại thứ tự của các quan sát, nếu không cấu trúc sẽ bị phá hủy.
Kết quả của mẫu	Mẫu ngẫu nhiên	Kết quả ngẫu nhiên
	Mẫu được thu thập	Kết quả thực tế

Câu hỏi: Xác định loại dữ liệu của các ví dụ bên dưới



- Dữ liệu về doanh số, doanh thu, số lượng khách hàng hàng ngày trong tháng vừa qua ở mỗi chi nhánh của Highlands Coffee trên TP.HCM.
- Dữ liệu về doanh thu và chi phí hàng ngày trong vòng 12 tháng ở chi nhánh Highlands Coffee ở Crescent Mall.
- Dữ liệu về doanh số, doanh thu, số lượng khách hàng trong tháng vừa qua của tất cả các chi nhánh Highlands Coffee trên TP.HCM.
- Dữ liệu về doanh thu và chi phí bán hàng vào ngày Giáng sinh năm 2019 ở tất cả chi nhánh của Highlands Coffee trên Việt Nam.

Các tính chất trên mẫu hữu hạn của các ước lượng OLS dưới các giả thiết cổ điển



- **Giả thiết TS.1 (Tuyến tính theo tham số) (**linear in parameter**)**

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + u_t$$

Chuỗi thời gian thể hiện quan hệ tuyến tính. Các quá trình ngẫu nhiên y_t , x_{t1}, \dots, x_{tk} đều có thể quan sát được, sai số u_t không quan sát được. Định nghĩa về biến giải thích mang tính tổng quát, vd. biến giải thích có thể có độ trễ hoặc là hàm số của các biến giải thích khác.

- **Giả thiết TS.2 (Không có đa cộng tuyến hoàn hảo) (**no perfect collinearity**)**

"Trong mẫu (và vì vậy trong quá trình chuỗi thời gian tổng thể), không có biến độc lập nào là hằng số cũng như không có biến độc lập nào là tổ hợp tuyến tính hoàn hảo của các biến độc lập còn lại."

Các tính chất trên mẫu hữu hạn của các ước lượng OLS dưới các giả thiết cổ điển (tt.)



■ Ký hiệu

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ \vdots & \vdots & & \vdots \\ x_{t1} & x_{t2} & \cdots & x_{tk} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

Ma trận này tập hợp tất cả thông tin về dữ liệu theo thời gian của tất cả biến giải thích

Giá trị của tất cả biến giải thích trong giai đoạn t

■ Giả thiết TS.3 (trung bình có điều kiện bằng 0) (**zero conditional mean**)

$$E(u_t | \mathbf{X}) = 0$$

Giá trị trung bình của các yếu tố không dự đoán được không tương quan với giá trị của các biến giải thích trong mọi giai đoạn

Các tính chất trên mẫu hữu hạn của các ước lượng OLS dưới các giả thiết cổ điển (tt.)



■ Thảo luận giả thiết TS.3

Ngoại sinh cùng kỳ $E(u_t | \mathbf{x}_t) = 0$ ← Trung bình của sai số không liên quan đến giá trị của các biến giải thích trong cùng giai đoạn
(exogeneity):

Ngoại sinh nghiêm ngặt $E(u_t | \mathbf{X}) = 0$ ← Trung bình của sai số không liên quan đến giá trị của các biến giải thích trong tất cả các giai đoạn
(strict exogeneity):

■ Ngoại sinh nghiêm ngặt mạnh hơn ngoại sinh cùng kỳ

- TS.3 loại bỏ ảnh hưởng của biến phụ thuộc đối với giá trị tương lai của biến giải thích; điều này thường có nhiều nghi vấn đặc biệt nếu biến giải thích “điều chỉnh” trước những thay đổi trong quá khứ của biến phụ thuộc
- Nếu sai số (**the error term**) có liên quan đến giá trị trong quá khứ của biến giải thích, có thể đưa những giá trị này vào biến độc lập cùng kỳ

Các tính chất trên mẫu hữu hạn của các ước lượng OLS dưới các giả thiết cổ điển (tt.)



- **Định lý 10.1 (Tính không chệch của ước OLS) (Unbiasedness of OLS)**

$$TS.1-TS.3 \Rightarrow E(\hat{\beta}_j) = \beta_j, \quad j = 0, 1, \dots, k$$

- **Giả thiết TS.4 (Phương sai không đổi- Homoscedasticity)**

$$Var(u_t|\mathbf{X}) = Var(u_t) = \sigma^2$$

← Biến động của các sai số không liên quan đến biến giải thích trong bất kỳ giai đoạn nào

- Điều kiện đủ là biến động của sai số độc lập với biến giải thích và không phải là hằng số theo thời gian
- Trong phân tích hồi quy chuỗi thời gian, phương sai không đổi dễ bị vi phạm, vd. nếu biến động của biến phụ thuộc bị tác động bởi thay đổi mô hình

Các tính chất trên mẫu hữu hạn của các ước lượng OLS dưới các giả thiết cổ điển (tt.)



■ Giả thiết TS.5 (Không có tương quan chuỗi) (**No serial correlation**)

$$\text{Corr}(u_t, u_s | \mathbf{X}) = 0, \quad t \neq s$$

← Điều kiện đối với các biến giải thích, các yếu tố không dự đoán được không có tương quan với nhau theo thời gian

■ Thảo luận về giả định TS.5

- Vì sao giả thiết này không xuất hiện trong phân tích dữ liệu chéo?
- Giả định rất dễ bị vi phạm nếu với điều kiện đã biết giá trị của biến độc lập, những yếu tố không dự đoán được có tương quan theo thời gian
- Giả định này có thể thay thế cho giả định chọn mẫu ngẫu nhiên (**random sampling assumption**) nếu việc chọn mẫu trong phân tích dữ liệu chéo không ngẫu nhiên 100%
- Trong trường hợp này, nếu đã có giá trị của biến giải thích, sai số không được tương quan giữa các đơn vị dữ liệu chéo (vd. khu vực)

Các tính chất trên mẫu hữu hạn của các ước lượng OLS dưới các giả thiết cổ điển (tt.)



- **Định lý 10.2 (phương sai mẫu của OLS) (OLS sampling variances)**

Trong giả thiết TS.1 – TS.5:

Công thức tương tự như trong phân tích dữ liệu chéo

$$Var(\hat{\beta}_j|\mathbf{X}) = \frac{\sigma^2}{SST_j(1 - R_j^2)}, \quad j = 1, \dots, k$$

Điều kiện về giá trị của biến giải thích không dễ hiểu. Ta có thể hiểu nôm na như sau, trong một mẫu hữu hạn, chúng ta có thể bỏ qua biến đổi do lấy mẫu (sampling variability) phát sinh do tính ngẫu nhiên của các biến độc lập. Biến đổi do lấy mẫu dạng này thường không quá lớn (vì số tổng).

- **Định lý 10.3 (Ước lượng không thiên vị của phương sai sai số) (Unbiased estimation of the error variance)**

$$TS.1 - TS.5 \quad \Rightarrow \quad E(\hat{\sigma}^2) = \sigma^2$$

Các tính chất trên mẫu hữu hạn của các ước lượng OLS dưới các giả thiết cổ điển (tt.)



- **Định lý 10.4 (Định lý Gauss-Markov) (Gauss-Markov Theorem)**

- Dưới giả thiết TS.1 – TS.5, các ước lượng OLS có phương sai nhỏ nhất trong tất cả các ước lượng tuyến tính không thiên vị của hệ số hồi quy
- Đây là điều kiện và cũng không phải là điều kiện đối với các biến độc lập

- **Giả thiết TS.6 (Phân phối chuẩn) (Normality)**

Giả thiết này bao hàm TS.3 – TS.5

$$u_t \sim N(0, \sigma^2) \quad \text{độc lập với} \quad \mathbf{X}$$

- **Định lý 10.5 (ước lượng mẫu có phân phối chuẩn) (Normal sampling distributions)**

- Dưới giả thiết TS.1 – TS.6, các ước lượng OLS có phân phối chuẩn (với điều kiện X. Có thể sử dụng thống kê F- và t- để kiểm định giả thuyết.

Các tính chất trên mẫu hữu hạn của các ước lượng OLS dưới các giả thiết cổ điển (tt.)



■ Ví dụ: Đường cong Phillips dạng tĩnh

$$\widehat{inf}_t = 1.42 + .468 unem_t$$

(1.72) (.289)

Trái với lý thuyết, phương trình đường cong Phillips không thể hiện sự đánh đổi giữa lạm phát và thất nghiệp

$$n = 49, R^2 = .053, \bar{R}^2 = .033$$

■ Thảo luận giả thiết CLM

TS.1: $inf_t = \beta_0 + \beta_1 unem_t + u_t$

Sai số là do những nhân tố như cú sốc tiền tệ, cú sốc thu nhập/cú sốc cầu, cú sốc giá dầu, cú sốc cung hoặc sốc tỷ giá hối đoái

TS.2: Quan hệ tuyến tính có giới hạn, và chỉ là ước lượng. Đa cộng tuyến hoàn hảo không phải là vấn đề đáng quan tâm miễn sao tỉ lệ thất nghiệp thay đổi theo thời gian.

Các tính chất trên mẫu hữu hạn của các ước lượng OLS dưới các giả thiết cổ điển (tt.)



■ Thảo luận giả thiết CLM (tt.)

TS.3: $E(u_t | unem_1, \dots, unem_n) = 0$ ← Dễ bị vi phạm

$unem_{t-1} \uparrow \rightarrow u_t \downarrow$ ← Ví dụ, cú sốc tỉ lệ thất nghiệp trong quá khứ có thể dẫn đến cú sốc cầu trong tương lai dẫn đến lạm phát tăng

$u_{t-1} \uparrow \rightarrow unem_t \uparrow$ ← Ví dụ, cú sốc giá dầu có nghĩa là lạm phát tăng và có thể dẫn đến thất nghiệp tăng trong tương lai

TS.4: $Var(u_t | unem_1, \dots, unem_n) = \sigma^2$ ← Giả thiết sẽ bị vi phạm nếu chính sách tiền tệ "căng" hơn trong giai đoạn thất nghiệp cao

TS.5: $Corr(u_t, u_s | unem_1, \dots, unem_n) = 0$ ← Giả thiết bị vi phạm nếu ảnh hưởng của tỷ giá hối đoái kéo dài theo thời gian (không thể giải thích bằng tỉ lệ thất nghiệp)

TS.6: $u_t \sim N(0, \sigma^2)$ ← Nhiều nghi vấn

Các tính chất trên mẫu hữu hạn của các ước lượng OLS dưới các giả thiết cổ điển (tt.)



■ Ví dụ: Ảnh hưởng của lạm phát và thâm hụt ngân sách lên lãi suất

Lãi suất T-bill 3 tháng

Tỉ lệ thâm hụt ngân sách trên GDP

$$\widehat{i3}_t = 1.73 + .606 \text{ inf}_t + .513 \text{ def}_t$$

(0.43) (.082) (.118)

$$n = 56, R^2 = .602, \bar{R}^2 = .587$$

Sai số thể hiện các yếu tố ảnh hưởng đến lãi suất khác ví dụ như chu kỳ kinh tế

■ Thảo luận về giả thiết CLM

TS.1: $i3_t = \beta_0 + \beta_1 \text{ inf}_t + \beta_2 \text{ def}_t + u_t$

TS.2: Quan hệ tuyến tính có giới hạn, và chỉ là ước lượng. Đa cộng tuyến hoàn hảo hiếm khi là vấn đề đáng quan tâm trong thực tế.

Các tính chất trên mẫu hữu hạn của các ước lượng OLS dưới các giả thiết cổ điển (tt.)



■ Thảo luận về giả thiết CLM (tt.)

TS.3: $E(u_t | inf_1, \dots, inf_n, def_1, \dots, def_n) = 0$ ← Dễ bị vi phạm

$def_{t-1} \uparrow \rightarrow u_t \uparrow$ ← Ví dụ, thâm hụt chi tiêu trong quá khứ có thể thúc đẩy hoạt động kinh tế, từ đó dẫn đến tăng lãi suất

$u_{t-1} \uparrow \rightarrow inf_t \uparrow$ ← Ví dụ, những cú sốc cầu không quan sát được thường khiến lãi suất tăng và lạm phát tăng trong những giai đoạn sau

TS.4: $Var(u_t | inf_1, \dots, def_n) = \sigma^2$ ← Giả định bị vi phạm nếu thâm hụt cao dẫn đến nhiều bất ổn về tài chính và có thể dẫn đến thay đổi đột ngột về lãi suất

TS.5: $Corr(u_t, u_s | inf_1, \dots, def_n) = 0$ ← Giả thiết bị vi phạm nếu ảnh hưởng của chu kỳ kinh tế kéo dài qua nhiều năm (và chu kỳ kinh tế cũng không giải thích hoàn toàn thay đổi trong tỉ lệ lạm phát và thâm hụt ngân sách)

TS.6: $u_t \sim N(0, \sigma^2)$ ← Nhiều nghi vấn

Sử dụng biến giả trong phân tích chuỗi thời gian



Tỷ lệ số trẻ em được sinh ra trên 1000 phụ nữ trong năm t

Tỉ lệ miễn thuế cá nhân trong năm t

Biến giả cho những năm Thế chiến II (1941-45)

Biến giả cho những năm khi thuốc tránh thai bắt đầu được sử dụng (1963-nay)

$$\widehat{gfr}_t = 98.68 + .083 pe_t - 24.24 ww2_t - 31.59 pill_t$$

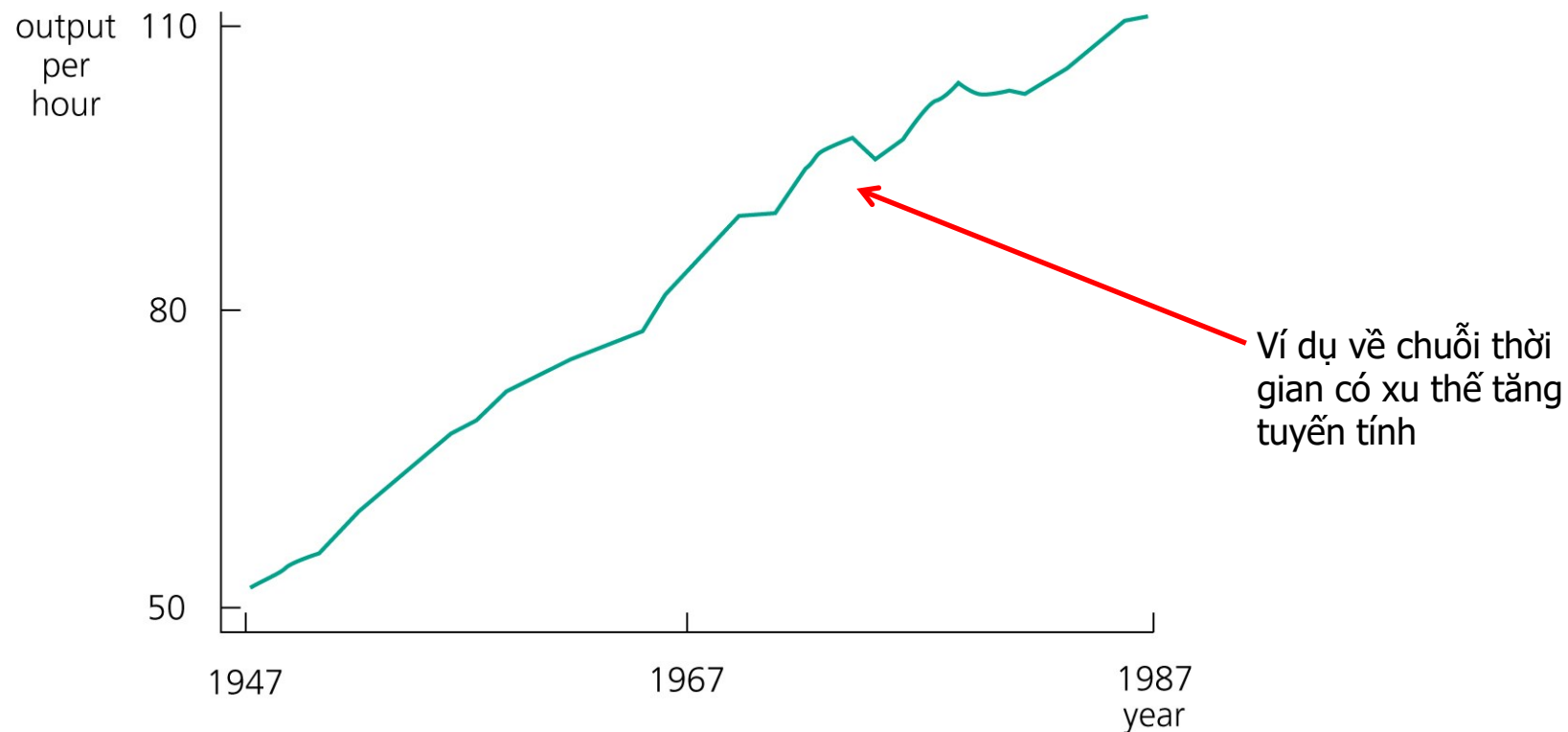
(3.68) (.030) (7.46) (4.08)

$$n = 72, R^2 = .473, \bar{R}^2 = .450$$

■ Giải thích

- Trong giai đoạn Thế chiến II, tỉ lệ sinh tạm thời giảm xuống
- Tỉ lệ sinh giảm dài hạn từ khi thuốc tránh thai xuất hiện từ năm 1963

Phân tích chuỗi thời gian có xu thế



Phân tích chuỗi thời gian có xu thế (tt.)



■ Phương trình thể hiện xu thế thời gian (**a linear time trend**)

$$y_t = \alpha_0 + \alpha_1 t + e_t \quad \Leftrightarrow \quad E(\Delta y_t) = E(y_t - y_{t-1}) = \alpha_1$$

$$\partial y_t / \partial t = \alpha_1$$

← Không xét đến độ lệch ngẫu nhiên, lượng mà biến phụ thuộc tăng là hằng số trên mỗi đơn vị thời gian

$$E(y_t) = \alpha_0 + \alpha_1 t$$

← Hoặc, giá trị kỳ vọng của biến phụ thuộc là hàm số tuyến tính của thời gian

■ Phương trình thể hiện xu thế thời gian dạng mũ (**an exponential time trend**)

$$\log(y_t) = \alpha_0 + \alpha_1 t + e_t \quad \Leftrightarrow \quad E(\Delta \log(y_t)) = \alpha_1$$

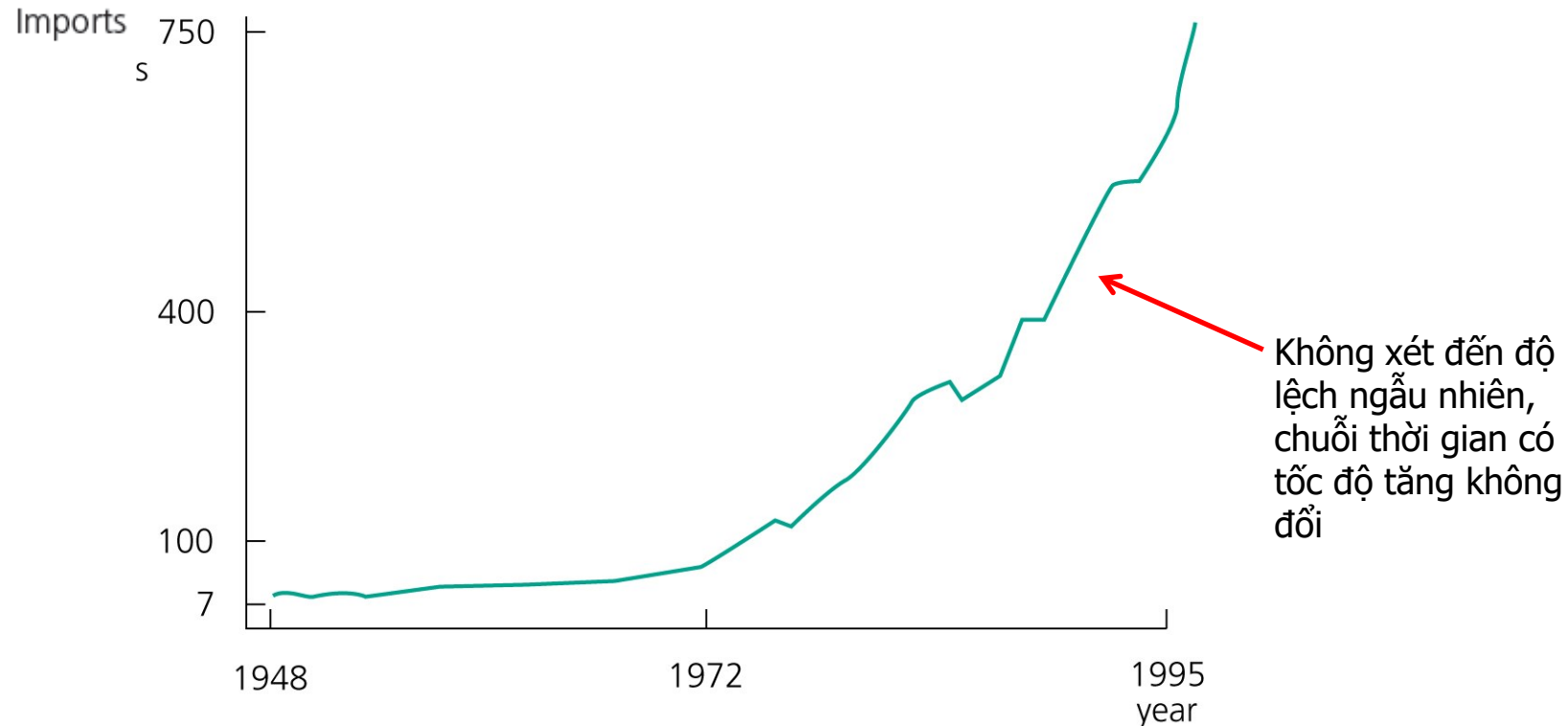
$$(\partial y_t / y_t) / \partial t = \alpha_1$$

← Không xét đến độ lệch ngẫu nhiên, tỉ lệ mà biến phụ thuộc là hằng số trên mỗi đơn vị thời gian

Phân tích chuỗi thời gian có xu thế (tt.)



■ Ví dụ về chuỗi thời gian có xu thế dạng mũ



Phân tích chuỗi thời gian có xu thế (tt.)



- **Sử dụng biến xu thế trong phân tích hồi quy**
 - Nếu thực hiện hồi quy giữa các biến xu thế, tương quan giả mạo có thể xuất hiện nếu các biến đều có chung một xu thế
 - Trong trường hợp này, phải đưa xu thế vào mô hình hồi quy
- **Ví dụ: Đầu tư và giá nhà ở**

Đầu tư nhà ở trên đầu người

Chỉ số giá nhà

$$\widehat{\log(invpc)} = -\frac{.550}{(.043)} + \frac{1.241}{(.382)} \log(price)$$

$$n = 42, R^2 = .208, \bar{R}^2 = .189$$

Có vẻ như đầu tư và giá nhà có tương quan thuận với nhau

Phân tích chuỗi thời gian có xu thế (tt.)



■ Ví dụ: Đầu tư và giá nhà (tt.)

$$\widehat{\log(invpc)} = - .913 + .381 \log(price) + .0098 t$$

(.136) (.679) (.0035)

$$n = 42, R^2 = .341, \bar{R}^2 = .307$$

Không còn tương quan có ý nghĩa về mặt thống kê giữa giá nhà và đầu tư nhà ở

■ Khi nào nên đưa xu thế (**trend**) vào?

- Nếu biến phụ thuộc thể hiện đặc điểm xu thế rõ ràng
- Nếu cả biến phụ thuộc và biến độc lập đều có xu thế
- Nếu chỉ có một vài biến độc lập có xu thế; ảnh hưởng lên biến phụ thuộc chỉ rõ ràng sau khi loại trừ xu thế

Phân tích chuỗi thời gian có xu thế (tt.)



- **Diễn giải kết quả trong hàm hồi quy có biến xu thế**
 - Hóa ra hệ số OLS trong một hàm hồi quy có biến xu thế cũng tương tự như hệ số của một hàm hồi quy không có xu thế không tất cả các biến đã được khử tính xu thế trước khi thực hiện hồi quy
 - Điều này xuất phát từ diễn giải chung của hồi quy đa biến
- **Tính R^2 khi biến phụ thuộc có xu thế**
 - Do xu thế, phương sai của biến phụ thuộc sẽ bị ước lượng quá cao (overstated)
 - Đầu tiên nên khử tính xu thế (**detrend**) của biến phụ thuộc sau đó thực hiện hồi quy trên tất cả các biến độc lập (cộng thêm biến xu thế nếu các biến này cũng có tính xu thế)
 - R^2 của hồi quy này thể hiện mức độ phù hợp tốt hơn

Thể hiện tính mùa vụ trong hồi quy chuỗi thời gian



- Một phương pháp đơn giản là đưa vào các biến giả theo mùa:

$$y_t = \beta_0 + \delta_1 \text{feb}_t + \delta_2 \text{mar}_t + \delta_3 \text{apr}_t + \cdots + \delta_{11} \text{dec}_t + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + u_t$$

=1 nếu quan sát nằm trong tháng 12
=0 nếu là các tháng khác

- Một số nhận xét tương tự như trường hợp của xu thế thời gian
 - Hệ số hồi quy trên biến giả thích có thể xem là kết quả của việc khử tính mùa vụ ở biến phụ thuộc và biến giả thích
 - R^2 của hồi quy đã khử tính mùa vụ ở biến phụ thuộc có khả năng giải thích cao hơn



Chapter 10: Basic Regression Analysis with Time Series Data (Chương 10: Phân tích hồi quy cơ bản với dữ liệu chuỗi thời gian)

Trong sách: Wooldridge: Introduction to Econometrics: Europe, Middle East and Africa Edition