

Bài giảng 2:

Một số vấn đề mở rộng khi sử dụng OLS với chuỗi thời gian



Chuỗi thời gian dừng (Stationary Time Series)

- **Những giả định áp dụng cho đến hiện tại đều khá khắt khe**
 - Ngoại sinh nghiêm ngặt, phương sai không đổi, và không có tương quan chuỗi đều là những yêu cầu khắt khe, đặc biệt trong hoàn cảnh của dữ liệu chuỗi thời gian
 - Các suy luận thống kê phụ thuộc vào tính hợp lý của giả định phân phối chuẩn
 - Những giả định yếu hơn vẫn cần thiết nếu mẫu nghiên cứu lớn
 - Một yêu cầu quan trọng là đối với phân tích dữ liệu chuỗi thời gian ở những mẫu nghiên cứu lớn, chuỗi thời gian đó phải dừng và có tính phụ thuộc yếu
- **Chuỗi thời gian dừng**
 - Nói sơ lược, một quá trình chuỗi thời gian dừng nếu các yếu tố ngẫu nhiên và sự phụ thuộc vào cấu trúc thời gian không thay đổi qua thời gian

Chuỗi thời gian dừng (tt.)

- **Quá trình ngẫu nhiên dừng**

Quá trình ngẫu nhiên $\{x_t : t = 1, 2, \dots\}$ được gọi là dừng, nếu với mỗi tập chỉ số thời gian $1 \leq t_1 \leq t_2 \leq \dots \leq t_m$, thì phân phối đồng thời của $(x_{t_1}, x_{t_2}, \dots, x_{t_m})$ cũng chính là phân phối đồng thời của $(x_{t_1+h}, x_{t_2+h}, \dots, x_{t_m+h})$ với mọi số nguyên $h \geq 1$

- **Quá trình dừng hiệp phương sai**

Quá trình ngẫu nhiên $\{x_t : t = 1, 2, \dots\}$ là dừng hiệp phương sai, nếu giá trị kỳ vọng, phương sai và hiệp phương là hằng số qua thời gian:

1) $E(x_t) = \mu$, 2) $Var(x_t) = \sigma^2$, và 3) $Cov(x_t, x_{t+h}) = f(h)$.

Chuỗi thời gian phụ thuộc yếu (**Weakly dependent time series**)

- **Chuỗi thời gian phụ thuộc yếu**

Quá trình ngẫu nhiên $\{x_t : t = 1, 2, \dots\}$ là phụ thuộc yếu, nếu x_t và x_{t+h} “độc lập hầu hết” khi h tăng không giới hạn (đối với tất cả t)

- **Thảo luận đặc điểm phụ thuộc yếu**

- Một hàm ý của phụ thuộc yếu là tương quan giữa x_t và x_{t+h} tiến về 0 nếu h tăng không giới hạn
- (Lưu ý chuỗi thời gian có thể không có tính dừng nhưng sẽ có tính phụ thuộc yếu)

Ví dụ về chuỗi thời gian phụ thuộc yếu

- **Quá trình trung bình trượt bậc nhất (MA(1))**

$$x_t = e_t + \alpha_1 e_{t-1} \quad \leftarrow \text{Quá trình là trung bình trượt ngắn hạn của chuỗi } e_t \text{ i.i.d.}$$

Quá trình này là phụ thuộc yếu vì các quan sát cách nhau nhiều hơn một giai đoạn không có điểm chung nào và do đó không tương quan với nhau.

- **Quá trình tự hồi quy bậc nhất (AR(1))**

$$y_t = \rho_1 y_{t-1} + e_t \quad \leftarrow \text{Quá trình này kéo dài giá trị của giai đoạn trước ở một mức độ nhất định (cộng thêm những cú sốc ngẫu nhiên từ chuỗi } e_t \text{ i.i.d.)}$$

$$\Rightarrow \text{Corr}(y_t, y_{t+h}) = \rho_1^h$$

Nếu điều kiện ổn định $|\rho_1| < 1$ vẫn đúng, quá trình này có tính phụ thuộc yếu vì tương quan chuỗi sẽ tiến đến 0 khi khoảng cách giữa các lần quan sát tăng vô hạn.

Tính chất tiệm cận của OLS

(Asymptotic properties of OLS)

- **Giả thiết TS.1' (Tuyến tính theo tham số) (Linear in parameters)**
 - Cũng như giả thiết TS.1 nhưng bây giờ biến phụ thuộc và biến độc lập được cho là có tính dừng và phụ thuộc yếu.
- **Giả thiết TS.2' (Không có đa cộng tuyến hoàn hảo) (No perfect collinearity)**
 - Tương tự như giả thiết TS.2
- **Giả thiết TS.3' (Trung bình có điều kiện bằng 0) (Zero conditional mean)**
 - Bây giờ biến giải thích được cho là chỉ có ngoại sinh *cùng kỳ* thay vì ngoại sinh nghiêm ngặt, nghĩa là

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

← Các biến giải thích cùng một giai đoạn không cung cấp thông tin về trung bình của sai số

Định lý 11.1 (Tính vững của OLS)

■ Định lý 11.1 (Tính vững của OLS) (**Consistency of OLS**)

$$TS.1' - TS.3' \Rightarrow \text{plim } \hat{\beta}_j = \beta_j, \quad j = 0, 1, \dots, k$$

Lưu ý quan trọng: Đối với tính vững, có thể giả định là các biến giải thích *không tương quan* cùng ký với sai số.

- **Vì sao phải nói lỏng giả thiết ngoại sinh nghiêm ngặt (**strict exogeneity assumption**)?**
 - Ngoại sinh nghiêm ngặt là một điều kiện chặt chẽ vì loại trừ hết tất cả các mối quan hệ động giữa biến giải thích và sai số
 - Cụ thể, giả thiết này loại bỏ ảnh hưởng của biến phụ thuộc lên giá trị trong tương lai của biến giải thích (đây là điều rất hay xảy ra trong kinh tế)
 - Ngoại sinh nghiêm ngặt không cho phép sử dụng biến phụ thuộc có độ trễ làm biến độc lập

Định lý 11.1 (Tính vững của OLS) (tt.)

- **Vì sao biến phụ thuộc có độ trễ vi phạm ngoại sinh nghiêm ngặt?**

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t$$

← Đây là mô hình hồi quy đơn giản nhất với biến phụ thuộc có độ trễ trở thành biến độc lập

Ngoại sinh cùng kỳ:

$$E(u_t | y_{t-1}) = 0$$

Ngoại sinh nghiêm ngặt: $E(u_t | y_0, y_1, \dots, y_{n-1}) = 0$

← Ngoại sinh nghiêm ngặt nghĩa là sai số không có tương quan với tất cả $y_t, t=1, \dots, n-1$

Điều này dẫn đến một mâu thuẫn:

$$Cov(y_t, u_t) = \beta_1 Cov(y_{t-1}, u_t) + Var(u_t) > 0$$

- **Ước tính OLS khi có sự hiện diện của biến phụ thuộc có độ trễ**

- Dưới giả thiết về ngoại sinh cùng kỳ, OLS có tính vững nhưng bị thiên vị

Tính tiệm cận của OLS (tt.)

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

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

← Sai số có phương sai không đổi cùng kỳ

- **Giả thiết TS.5' (Không có tương quan chuỗi)**

$$Corr(u_t, u_s | \mathbf{x}_t, \mathbf{x}_s) = 0, \quad t \neq s$$

← Điều kiện đối với biến giải thích trong giai đoạn t và s, các sai số không tương quan với nhau

- **Định lý 11.2 (Tính tiệm cận chuẩn của OLS)**

- Dưới các giả thiết TS.1' – TS.5', các ước lượng OLS có phân phối tiệm cận chuẩn. Ngoài ra, các sai số chuẩn OLS, thống kê t và thống kê F đều có tiệm cận hợp lý.

Ví dụ: Giả thuyết thị trường hiệu quả (EMH)

■ Ví dụ: Giả thuyết thị trường hiệu quả (EMH)

Theo giả thuyết EMH những thông tin có thể quan sát được trên thị trường trước tuần t không thể giúp ta dự đoán lợi nhuận của cổ phiếu trong tuần t . Phiên bản đơn giản của giả thuyết cho rằng chỉ thông tin về lợi nhuận trong quá khứ mới là thông tin có liên quan trong việc dự đoán lợi nhuận của tuần t . Điều này có nghĩa là

$$E(\text{return}_t | \text{return}_{t-1}, \text{return}_{t-2}, \dots) = E(\text{return}_t)$$

Cách đơn giản để kiểm chứng giả thuyết EMH là dựng mô hình AR(1). Dưới giả thiết EMH, giả thiết TS.3' vẫn đúng như vậy có thể sử dụng hồi quy OLS để kiểm tra xem lợi nhuận tuần này có phụ thuộc vào lợi nhuận của tuần trước hay không.

$$\widehat{\text{return}}_t = .180 + .059 \text{return}_{t-1}$$

(.081) (.038)

$$n = 689, R^2 = .0035, \bar{R}^2 = .0020$$

Không có chứng cứ chống lại giả thuyết EMH. Đưa thêm nhiều lợi nhuận có độ trễ cũng cho ra kết quả tương tự.

Sử dụng chuỗi thời gian bền trong phân tích hồi quy



- **Sử dụng quá trình dừng xu thế trong phân tích hồi quy**
 - Chuỗi thời gian có xu thế xác định (**deterministic time trends**) sẽ không có tính dừng
 - Nếu chuỗi thời gian có tính dừng quanh một xu thế và có thêm tính phụ thuộc yếu sẽ được gọi là quá trình dừng xu thế
 - Quá trình dừng xu thế (**trend-stationary processes**) cũng thỏa mãn giả thiết TS.1'
- **Sử dụng chuỗi thời gian bền trong phân tích hồi quy**
 - Không may nhiều chuỗi thời gian kinh tế vi phạm giả thiết phụ thuộc yếu vì đây là các chuỗi thời gian bền (= phụ thuộc mạnh)
 - Trong tình huống này, không thể sử dụng biện pháp OLS (trừ khi giả thiết CLM vẫn còn đúng)
 - Trong một số trường hợp, có thể chuyển dữ liệu thành chuỗi thời gian phụ thuộc yếu

Các bước ngẫu nhiên

■ Bước ngẫu nhiên (**random walks**)

$$y_t = y_{t-1} + e_t$$

Bước ngẫu nhiên có tên gọi như vậy vì giá trị tại thời điểm t sẽ bằng giá trị của thời điểm trước y_{t-1} cộng với một biến ngẫu nhiên e_t phân phối i.i.d.

$$\Rightarrow y_t = (y_{t-2} + e_{t-1}) + e_t = \dots = e_t + e_{t-1} + \dots + e_1 + y_0$$

Giá trị hiện tại là phép cộng của tất cả các cú sốc trong quá khứ với giá trị ban đầu. Đây là lý do vì sao bước ngẫu nhiên có tính bền: Hiệu ứng của một cú sốc kéo dài trong chuỗi thời gian trong dài hạn.

$$E(y_t) = E(y_0)$$

$$Var(y_t) = \sigma_e^2 t$$

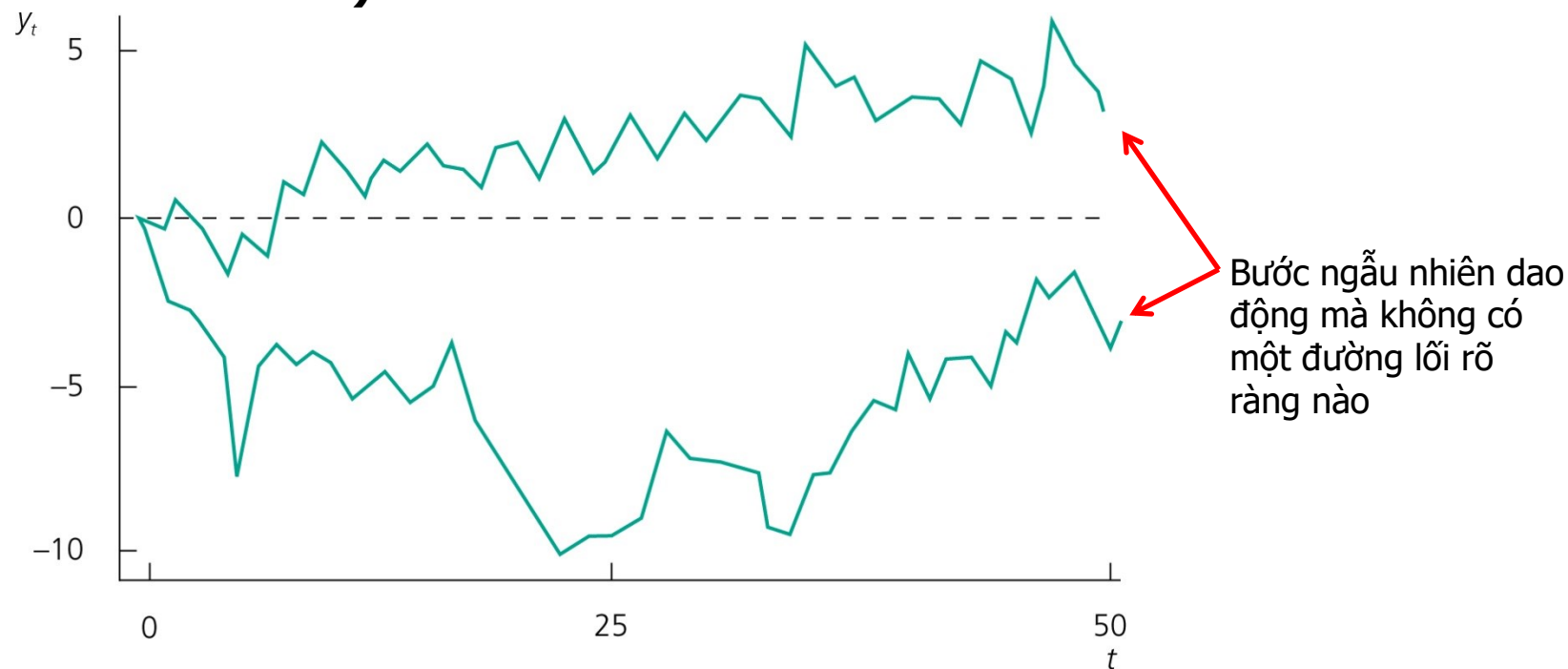
$$Corr(y_t, y_{t+h}) = \sqrt{t/(t+h)}$$

Bước ngẫu nhiên không có tính hiệp phương sai dừng (covariance stationary) vì phương sai và hiệp phương sai phụ thuộc theo thời gian.

Bước ngẫu nhiên cũng không có tính phụ thuộc yếu vì tương quan giữa các quan sát biến mất rất chậm và việc biến mất phụ thuộc vào độ lớn của t .

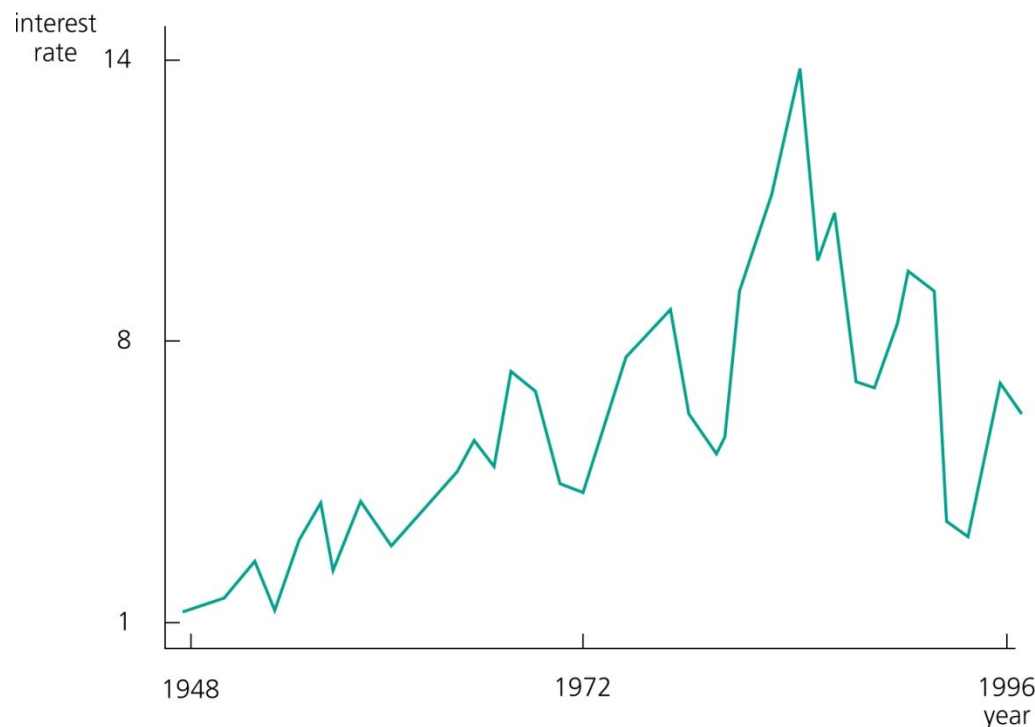
Ví dụ về chuỗi thực nghiệm của bước ngẫu nhiên

- Ví dụ về chuỗi thực nghiệm của bước ngẫu nhiên (**random walk realizations**)



Ví dụ về chuỗi thực nghiệm của bước ngẫu nhiên (tt.)

■ Lãi suất 3 tháng tín phiếu kho bạc Mỹ, từ năm 1948-1996



Bước ngẫu nhiên là trường hợp đặc biệt của quá trình nghiệm đơn vị (unit root).

Quá trình nghiệm đơn vị được định nghĩa là bước ngẫu nhiên nhưng e_t có thể là một quá trình ngẫu nhiên phụ thuộc yếu.

Từ quan điểm kinh tế, phải xác định chuỗi thời gian có tính bền hay không. Trong những chuỗi thời gian bền, những cú sốc hay thay đổi chính sách sẽ có tác động kéo dài/vĩnh viễn, trong những quá trình phụ thuộc yếu, tác động chỉ mang tính tạm thời.

Bước ngẫu nhiên có trượt

■ Bước ngẫu nhiên có trượt (**Random walks with drift**)

$$y_t = \boxed{\alpha_0} + y_{t-1} + e_t$$

Ngoài bước ngẫu nhiên thông thường, còn có sự tăng/giảm xác định (= trượt) trong mỗi giai đoạn

$$\Rightarrow y_t = \boxed{\alpha_0 t} + e_t + e_{t-1} + \dots + e_1 + y_0$$

Điều này dẫn đến xu thế thời gian tuyến tính trong đó chuỗi thời gian thể hiện hành vi bước ngẫu nhiên. Vì không có rõ là bước ngẫu nhiên sẽ diễn ra theo chiều hướng nào, bước ngẫu nhiên có thể khác hoàn toàn với xu thế

$$E(y_t) = \boxed{\alpha_0 t} + E(y_0)$$

$$Var(y_t) = \sigma_e^2 t$$

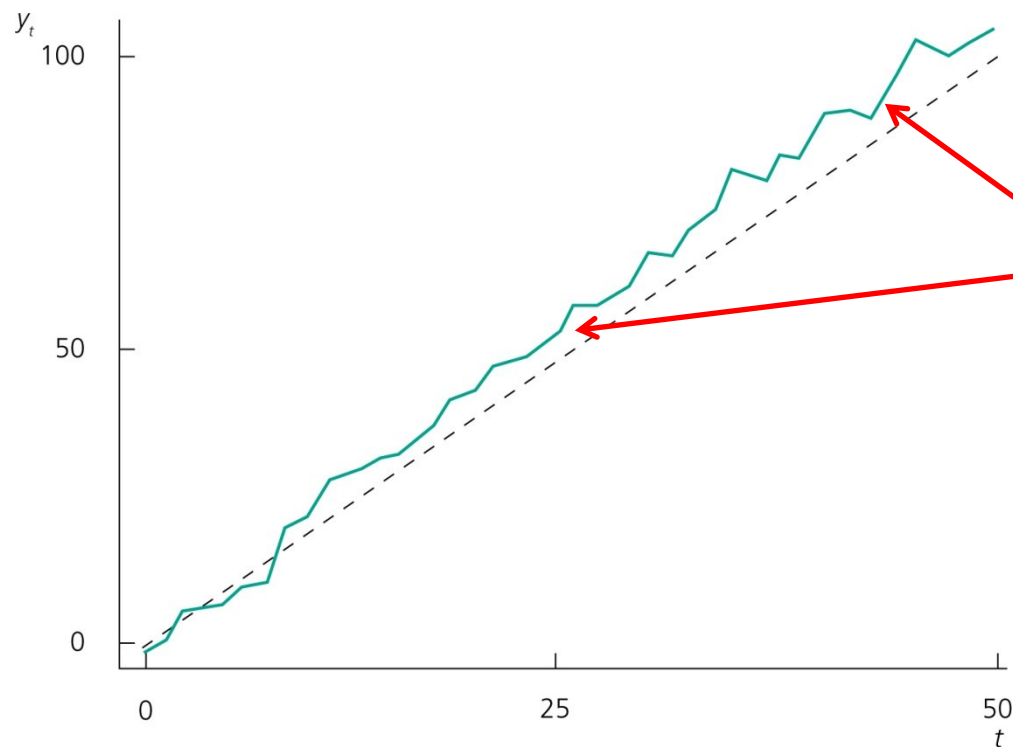
$$Corr(y_t, y_{t+h}) = \sqrt{t/(t+h)}$$

Mặt khác, bước ngẫu nhiên có trượt có những đặc điểm tương tự như bước ngẫu nhiên không có trượt.

Bước ngẫu nhiên có trượt không có hiệp phương sai dừng và cũng không có tính phụ thuộc yếu.

Ví dụ điển hình của bước ngẫu nhiên có trượt

■ Ví dụ mẫu của bước ngẫu nhiên có trượt



Lưu ý là chuỗi thường không trùng với đường xu thế.

Bước ngẫu nhiên có trượt có thể là mô hình mô phỏng tốt chuỗi thời gian có xu thế rõ ràng nhưng không có tính phụ thuộc yếu.

Các phép biến đổi trên chuỗi thời gian bền

- **Các phép biến đổi trên chuỗi thời gian bền (highly persistent time series)**

- **Quá trình liên kết (**order of integration**)**

- Chuỗi thời gian phụ thuộc yếu là chuỗi có quá trình liên kết bậc 0 (= I(0))
- Nếu chuỗi thời gian phải sai phân một lần để có chuỗi thời gian phụ thuộc yếu, đây được gọi là quá trình liên kết bậc nhất (= I(1))

- **Ví dụ về quá trình liên kết I(1)**

$$y_t = y_{t-1} + e_t \Rightarrow \Delta y_t = y_t - y_{t-1} = e_t$$

$$\log(y_t) = \log(y_{t-1}) + e_t \Rightarrow \Delta \log(y_t) = e_t$$

Sau khi sai phân, kết quả là ta có chuỗi thời gian phụ thuộc yếu (vì e_t có tính phụ thuộc yếu).

- **Sai phân là cách thông thường để có được chuỗi thời gian phụ thuộc yếu (**weak dependence**)**

Các phép biến đổi trên chuỗi thời gian bền (tt.)

- **Xác định chuỗi thời gian có phải là I(1) hay không**

- Có những loại kiểm định thống kê để kiểm tra xem một chuỗi thời gian có phải là I(1) hay không (=kiểm định nghiệm đơn vị); những chương sau sẽ nói về các kiểm định này
- Hoặc, nhìn vào tương quan tự động bậc 1 của mẫu:

$$\hat{\rho}_1 = \widehat{Corr}(y_t, y_{t-1})$$

← Đo lường các quan sát kế nhau trên chuỗi thời gian có mức độ tương quan với nhau như thế nào

- Nếu tự tương quan bậc nhất gần với 1, điều này có nghĩa là chuỗi thời gian có tính bền (= chứa nghiệm đơn vị)
- Hoặc, chuỗi thời gian có thể có xu hướng xác định
- Có thể khử cả nghiệm đơn vị và xu thế bằng sai phân

Ví dụ: Phương trình về tỉ lệ sinh đẻ

■ Ví dụ: Fertility equation

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

Có thể ước tính phương trình này bằng OLS nếu giả thiết CLM vẫn đúng. Nhưng điều này không chắc chắn, cho nên chúng ta buộc phải sử dụng phân tích mẫu lớn. Đối với phân tích mẫu lớn, chuỗi thời gian của tỉ lệ sinh và chuỗi thời gian của tỉ lệ miễn thuế cá nhân phải có tính dừng và phụ thuộc yếu. Đây là điều chưa chắc chắn vì hai chuỗi thời gian này có tính bền:

$$\hat{\rho}_{gfr} = .977, \hat{\rho}_{pe} = .964$$

Vì vậy, nên ước tính phương trình bằng sai phân bậc nhất. Điều này hợp lý hơn vì nếu phương trình đúng ở mọi bậc, phương trình cũng đúng ở bậc nhất:

$$\Delta \widehat{gfr} = - \frac{.964}{(.468)} - \frac{.036}{(.027)} \Delta pe - \frac{.014}{(.028)} \Delta pe_{-1} + \frac{.110}{(.027)} \Delta pe_{-2}$$

$$n = 69, R^2 = .233, \bar{R}^2 = .197$$

Ước tính δ_2

Ví dụ: Tiền lương và năng suất lao động

■ Ví dụ: Tiền lương và năng suất lao động

Có biến xu thế vì cả hai chuỗi để thể hiện tính xu thế rõ ràng.

$$\widehat{\log(hr\text{wage})} = -5.33 + 1.64 \log(outphr) - .018 t$$

(.37) (.09) (.002)

$$n = 41, R^2 = .971, \bar{R}^2 = .970$$

Độ co giãn của lương theo giờ so với sản lượng theo giờ (=năng suất) có vẻ lớn một cách khó tin.

Hóa ra thậm chí sau khi khử xu thế, cả hai chuỗi đều thể hiện tự tương quan mẫu gần bằng 1 như vậy ước tính phương trình trong phân sai bậc nhất có vẻ hợp lý hơn:

$$\Delta \widehat{\log(hr\text{wage})} = - .0036 + .809 \Delta \log(outphr)$$

(.0042) (.173)

Con số ước tính độ co giãn của lương theo giờ theo năng suất hợp lý hơn.

$$n = 40, R^2 = .364, \bar{R}^2 = .348$$

Mô hình động đầy đủ



- **Mô hình động đầy đủ (Dynamically complete models)**

- Một mô hình được cho là mô hình động đầy đủ nếu các biến có độ trễ được đưa đầy đủ vào làm biến giải thích như vậy những biến trễ khác không thể giải thích biến phụ thuộc:

$$E(y_t | \mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, y_{t-2}, \dots) = E(y_t | \mathbf{x}_t)$$

- **Mô hình động đầy đủ hàm ý rằng không có tương quan chuỗi trong mô hình**

- Nếu có biến trễ thực sự thuộc về mô hình hồi quy, việc bỏ đi các biến này sẽ tạo ra tương quan chuỗi (nếu các biến có tương quan chuỗi với nhau)

- **Có thể kiểm chứng tính động đầy đủ rất dễ dàng**

- Nếu không thể loại bỏ các biến trễ, điều này có nghĩa là có tương quan chuỗi

Ngoại sinh chuỗi (Sequential exogeneity)

- **Ngoại sinh chuỗi**

- Một loạt các biến giải thích được cho là có ngoại sinh chuỗi nếu “đủ” biến giải thích được đưa vào phân tích hồi quy:

$$E(u_t | \mathbf{x}_t, \mathbf{x}_{t-1}, \dots) = E(u_t) = 0$$

- Ngoại sinh chuỗi yếu hơn ngoại sinh nghiêm ngặt
- Ngoại sinh chuỗi tương đương với động đầy đủ nếu trong các biến giải thích có chứa biến phụ thuộc có độ trễ

- **Tất cả mô hình hồi quy có nên là mô hình động đầy đủ?**

- Không cần thiết: nếu ngoại sinh chuỗi vẫn đúng, vẫn có thể ước tính chính xác quan hệ nhân quả; việc không có tương quan chuỗi không quan trọng

Tài liệu tham khảo

Chapter 11: Further Issues Using OLS with Time Series Data. (Chương 11: Một số vấn đề mở rộng khi sử dụng OLS với chuỗi thời gian)

Trong sách: Wooldridge: Introductory Econometrics: A Modern Approach, 5e