CHƯƠNG 12

TỰ TƯƠNG QUAN

Không có một phương cách hiệu quả toàn năng nào giúp tránh được sự đặc trưng sai do giải thích sai hàm hồi qui khi có sự hiện diện của các sai số tương quan chuỗi.*

Một giả định quan trọng của mô hình tuyến tính cổ điển đã trình bày trong Phần I là không có quan hệ tự tương quan và tương quan chuỗi giữa các nhiễu u_i đã đưa vào hàm hồi qui tổng thể. Trong chương này, chúng ta hãy xem xét một cách có suy xét giả định này bằng cách đi tìm các câu trả lời cho các câu hỏi sau :

1. Bản chất của tự tương quan là gì ?
2. Các hậu quả về lý thuyết và thực tiễn của tự tương quan là gì ?
3. Do giả định về sự không tự tương quan có liên quan tới các nhiễu không thể quan sát được u_i, làm thế nào ta biết được rằng có quan hệ tự tương quan trong bất kỳ một tình thế đã được cho trước ?

Người đọc sẽ thấy chương này, theo nhiều cách, sẽ tương tự như chương trước về phương sai thay đổi trong đó **khi có cả quan hệ tự tương quan và phương sai thay đổi, các hàm ước lượng thông thường OLS, mặc dù không thiên lệch, không còn có các phương sai nhỏ nhất giữa tất cả các hàm tuyến tính không thiên lệch. Nói tóm lại, chúng không còn là ước lượng không thiên lệch tuyến tính tốt nhất (Best Linear Unbiased Estimation, BLUE) nữa.**

12.1 BẢN CHẤT CỦA VẤN ĐỀ

Thuật ngữ **tự tương quan** có thể được định nghĩa như là "quan hệ tương quan giữa các thành viên của chuỗi của các quan sát được sắp xếp theo thời gian [như trong dữ liệu chuỗi thời gian] hoặc không gian [như trong dữ liệu chéo]."[1] Trong ngữ cảnh hồi qui, mô hình hồi qui tuyến tính cổ điển giả định rằng quan hệ tự tương quan như vậy không tồn tại trong các nhiễu u_i. Viết theo ký hiệu là

[1] Maurince G. Kendall và William R. Buckland, Từ điển thuật ngữ thống kê, Hafner Publishing Company, NewYork, 1971, trang 8.

$$E(u_i u_j) = 0 \qquad i \neq j \qquad (3.2.5)$$

Đơn giản là mô hình cổ điển giả định rằng số hạng nhiễu liên quan tới bất cứ một quan sát nào đều không bị ảnh hưởng bởi số hạng nhiễu liên quan tới bất cứ một quan sát nào khác. Ví dụ, nếu chúng ta đang xử lý dữ liệu chuỗi thời gian theo quí có liên quan tới phép hồi qui sản lượng theo nhập lượng nhân công và vốn và nếu có xảy ra đình công tác động tới sản lượng trong một quí, không có lý do gì để tin rằng việc gián đoạn này sẽ kéo dài sang quí sau. Tức là nếu sản lượng là thấp hơn trong quí này, không có lý do gì để kỳ vọng nó sẽ thấp hơn trong quí sau. Tương tự, nếu chúng ta xử lý dữ liệu chéo có liên quan tới phép hồi qui của chi tiêu tiêu dùng gia đình theo thu nhập gia đình, tác động của gia tăng thu nhập của một gia đình tới chi tiêu tiêu dùng của gia đình đó không được kỳ vọng là tác động lên chi tiêu tiêu dùng của một gia đình khác.

Tuy nhiên, nếu có một sự phụ thuộc như vậy, chúng ta có quan hệ tự tương quan. Theo ký hiệu là

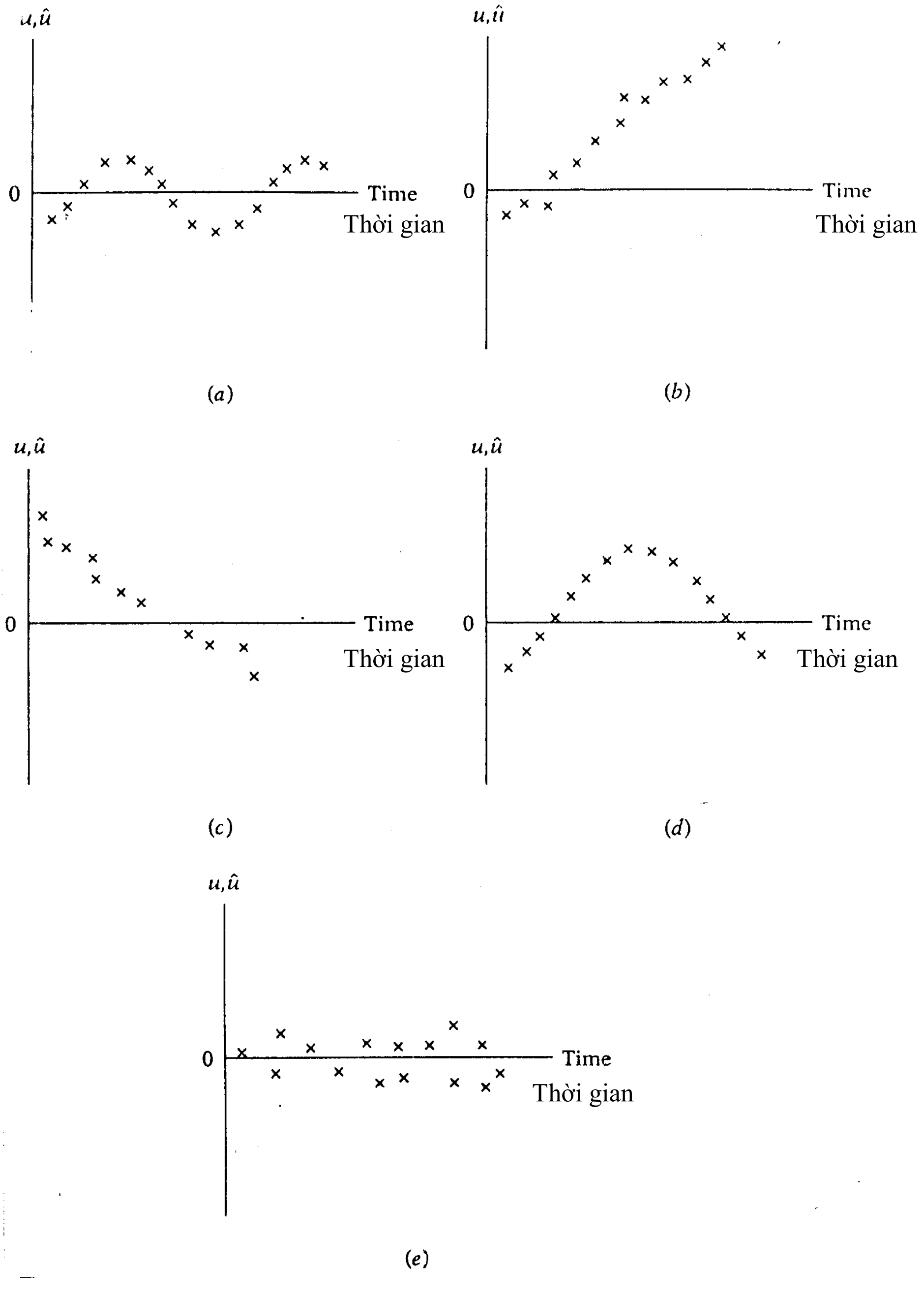
$$E(u_i u_j) \neq 0 \qquad i \neq j \qquad (12.1.1)$$

Trong tình thế này, sự gián đoạn xảy ra bởi đình công trong quí này có thể tác động rất nhiều tới sản lượng của quí sau, hoặc các gia tăng trong chi tiêu tiêu dùng của một gia đình có thể tạo ra cho một gia đình khác các gia tăng trong chi tiêu tiêu dùng của mình nếu nó muốn tuân theo Joneses.

Trước khi chúng ta tìm ra vì sao quan hệ tự tương quan tồn tại, điều cần thiết là làm rõ một số vấn đề thuộc về thuật ngữ. Mặc dù hiện nay trên thực tế thường coi các từ **tự tương quan** và **tương quan chuỗi** là đồng nghĩa, một số tác giả vẫn muốn phân biệt hai từ này. Ví dụ, Tintner định nghĩa tự tương quan như là "tương quan trễ của một chuỗi đã cho với chính nó, bị chậm lại bởi một số đơn vị thời gian", trong khi ông ta bảo tồn từ quan hệ chuỗi là "tương quan trễ giữa hai chuỗi khác nhau."[2] Do đó, tương quan giữa hai chuỗi thời gian như là $u_1, u_2, \ldots, u_{10}$ và $u_2, u_3, \ldots, u_{11}$, trong đó chuỗi thứ nhất là chuỗi thứ hai chậm lại một giai đoạn, được gọi là **tự tương quan**, trong khi tương quan giữa các chuỗi thời gian như là $u_1, u_2, \ldots, u_{10}$ và $v_2, v_3, \ldots, v_{11}$, trong đó u và v là hai chuỗi thời gian khác nhau, được gọi là **tương quan chuỗi**. Mặc dù sự khác biệt giữa hai từ này có thể là hữu ích, trong cuốn sách này chúng ta sẽ coi chúng là đồng nghĩa.

Chúng ta hãy xem xét một số các dạng dễ hiểu của *tự tương quan* và *không tự tương quan* được cho trong Hình 12.1. Hình 12.1a tới d cho thấy rằng có một dạng giữa các u. Hình 12.1a cho thấy dạng chu kỳ; Hình 12.1b và c cho thấy các xu hướng tuyến tính đi lên hay đi xuống của các nhiễu; trong khi Hình 12.1d chỉ ra cả hai từ xu hướng tuyến tính và bình phương đều có mặt trong các nhiễu. Chỉ có Hình 12.1e là cho thấy dạng không có hệ thống, ủng hộ cho giả định không có tự tương quan của mô hình hồi qui tuyến tính cổ điển.

[2] Gerhard Tintner, Kinh tế lượng, ấn bản nghiên cứu, John Wiley & Sons, New York, 1965, trang 187.



HÌNH 12.1 Các dạng của quan hệ tự tương quan

Câu hỏi quen thuộc là : Vì sao có tương quan chuỗi ? Có nhiều nguyên nhân, một số trong chúng là:

Tính ì. Một nét nổi bật của đa số chuỗi thời gian kinh tế là tính ì, hoặc tính chậm chạp. Như ta đã biết rõ, các chuỗi thời gian như GNP, chỉ số giá, sản xuất, việc làm và các chu kỳ xảy ra thất nghiệp (kinh doanh). Bắt đầu từ đáy của sự suy thoái, khi sự phục hồi kinh tế bắt đầu, đa số các chuỗi này bắt đầu chuyển động lên trên. Trong nhánh đi lên này, giá trị của một chuỗi tại một thời điểm lớn hơn giá trị trước đó của nó. Do đó có một "động lượng" được tạo nên trong chúng, và nó tiếp tục cho tới khi có xảy ra điều gì đó (nghĩa là gia tăng trong lãi suất hoặc thuế hoặc cả hai) để làm chậm chúng lại. Vì vậy, trong các phép hồi qui có liên quan tới dữ liệu chuỗi thời gian, các quan sát liên tiếp có khả năng là nội phụ thuộc.

Các thiên lệch trong xác định đặc trưng: trường hợp các biến bị loại ra. Trong phân tích theo kinh nghiệm, nhà nghiên cứu thường bắt đầu bằng một mô hình hồi qui có vẻ hợp lý có thể không phải là một mô hình "hoàn hảo" nhất. Sau khi phân tích hồi qui, nhà nghiên cứu mới mổ xẻ để tìm ra có phải các kết quả phù hợp với các kỳ vọng ban đầu hay không. Nếu không, cuộc giải phẫu bắt đầu. Ví dụ, nhà nghiên cứu có thể vẽ các phần dư $\hat{u}_i$ đã thu được từ phép hồi qui thích hợp và có thể thu được các dạng như là trong Hình 12.1a tới d. Các phần dư này (là các thay thế cho u_i) có thể đề xuất rằng một số biến tuy đã được tiến cử lúc đầu nhưng chưa được đưa vào mô hình này do nhiều lý do khác nhau sẽ cần được đưa vào. Đây là trường hợp các thiên lệch của đặc trưng mô hình do một số biến bị loại ra. Thông thường việc đưa vào các biến như vậy sẽ làm biến đổi dạng tương quan đã quan sát giữa các phần dư. Ví dụ, giả sử chúng ta có mô hình cầu sau đây :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t \qquad (12.1.2)$$

trong đó Y = lượng cầu thịt bò, X_2 = giá thịt bò, X_3 = thu nhập của người tiêu dùng, X_4 = giá thịt lợn, và t = thời gian.[3] Tuy nhiên, do một số lý do chúng ta thực hiện phép hồi qui sau :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + v_t \qquad (12.1.3)$$

Bây giờ, nếu (12.1.2) là mô hình "đúng" hoặc "thực sự" hoặc quan hệ thực sự, việc thực hiện (12.1.3) là tương đương với việc cho $v_t = \beta_4 X_{4t} + u_t$. Và với nghĩa là giá thịt lợn ảnh hưởng lên tiêu dùng thịt bò, số hạng sai số hoặc nhiễu v sẽ phản ánh một dạng có hệ thống, do đó tạo ra quan hệ tự tương quan (sai). Một kiểm định đơn giản của điều này có thể là thực hiện cả (12.1.2) lẫn (12.1.3) và xem có phải tự tương quan, nếu có, đã quan sát thấy trong mô hình (12.1.3) có biến mất khi thực hiện (12.1.2) hay không.[4] Các cơ chế thực tế của việc khám phá tự tương quan sẽ được thảo luận trong Phần 12.5, trong đó chúng ta sẽ chỉ ra rằng đồ thị các phần dư từ các phép hồi qui (12.1.2) và (12.1.3) sẽ thường làm rõ một cách đáng kể tương quan chuỗi.

Các thiên lệch trong xác định đặc trưng: dạng hàm không đúng. Giả sử mô hình "thực" hay đúng trong nghiên cứu về quan hệ chi phí-sản lượng là như sau :

$$\text{Chi phí biên}_i = \beta_1 + \beta_2 \text{ sản lượng}_i + \beta_3 \text{ sản lượng}_i^2 + u_i \qquad (12.1.4)$$

nhưng chúng ta thích hợp bằng mô hình sau :

[3] Do qui ước, chúng ta sẽ sử dụng chỉ số t để ký hiệu dữ liệu chuỗi thời gian và chỉ số i thông thường cho các dữ liệu chéo.

[4] Nếu đã tìm ra rằng vấn đề thực tế là một trong các thiên lệch trong xác định đặc trưng mô hình chứ không phải tự tương quan, thì như đã chỉ ra trong Phần 7.7, các hàm ước lượng OLS của các thông số (12.1.3) có thể bị thiên lệch cũng như không nhất quán. Về chi tiết, xin đọc Davidson và MacKinnon, op.cit., trang 327-328. Đồng thời đọc trích dẫn của họ cho trong phần đầu của chương này.

$$\text{Chi phí biên}_i = \alpha_1 + \alpha_2 \text{ sản lượng}_i + v_i \qquad (12.1.5)$$

Đường chi phí biên tương ứng với mô hình "thực" được nêu trong Hình 12.2 cùng với đường chi phí tuyến tính "không đúng".

Như Hình 12.2 cho thấy, ở giữa hai điểm A và B đường chi phí biên tuyến tính sẽ ước lượng cao hơn chi phí biên thực một cách nhất quán, trong khi ở ngoài hai điểm này sẽ ước lượng thấp hơn chi phí biên thực một cách nhất quán. Kết quả này cần được kỳ vọng, vì số hạng nhiễu v_i thực tế sẽ bằng sản lượng^2 + u_i, và vì vậy số hạng sản lượng^2 sẽ tác động có hệ thống lên chi phí biên. Trong trường hợp này, v_i sẽ phản ánh tự tương quan do sử dụng dạng hàm số không đúng. Trong Chương 13 chúng ta sẽ xem xét nhiều phương pháp phát hiện các thiên lệch trong xác định đặc trưng.

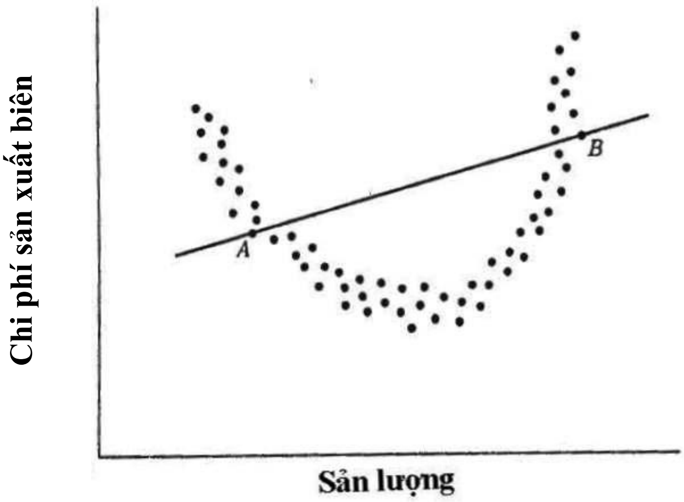
Hiện tượng Cobweb. Sự cung cấp nhiều mặt hàng nông sản phản ánh cái gọi là hiện tượng Cobweb, trong đó lượng cung phản ứng lại giá với một chậm trễ một thời đoạn vì các quyết định cung cần có thời gian để thực hiện (giai đoạn thai nghén). Do đó, vào lúc bắt đầu giao trồng vụ mùa măm nay, các nông dân bị ảnh hưởng bởi giá phổ biến trong năm trước, nên hàm cung của họ là :

$$\text{Lượng cung}_t = \beta_1 + \beta_2 P_{t-1} + u_t \qquad (12.1.6)$$

Giả sử vào cuối giai đoạn t, giá P_t trở nên thấp hơn P_{t-1}. Vì vậy, trong giai doạn t +1 các nông dân có thể quyết định rất rõ là sản xuất ít hơn họ đã làm trong giai đoạn t. Rõ ràng là trong tình hình này, các nhiễu u_t không được kỳ vọng là ngẫu nhiên bởi vì nếu các nông dân sản xuất vượt quá trong năm t, họ có khả năng giảm sản xuất của mình trong t +1, và tiếp tục như vậy, dẫn tới dạng Cobweb.

Các độ trễ. Trong hồi qui chuỗi thời gian của chi tiêu tiêu dùng lên thu nhập, không phải là bất thường khi nhận thấy rằng chi tiêu tiêu dùng trong giai đoạn hiện tại phụ thuộc vào, giữa các cái khác, chi tiêu tiêu dùng của giai đoạn trước đó. Tức là,

$$\text{Tiêu dùng}_t = \beta_1 + \beta_2 \text{ thu nhập}_t + \beta_3 \text{ tiêu dùng}_{t-1} + u_t \qquad (12.1.7)$$



HÌNH 12.2 Thiên lệch trong xác định đặc trưng: dạng hàm không đúng

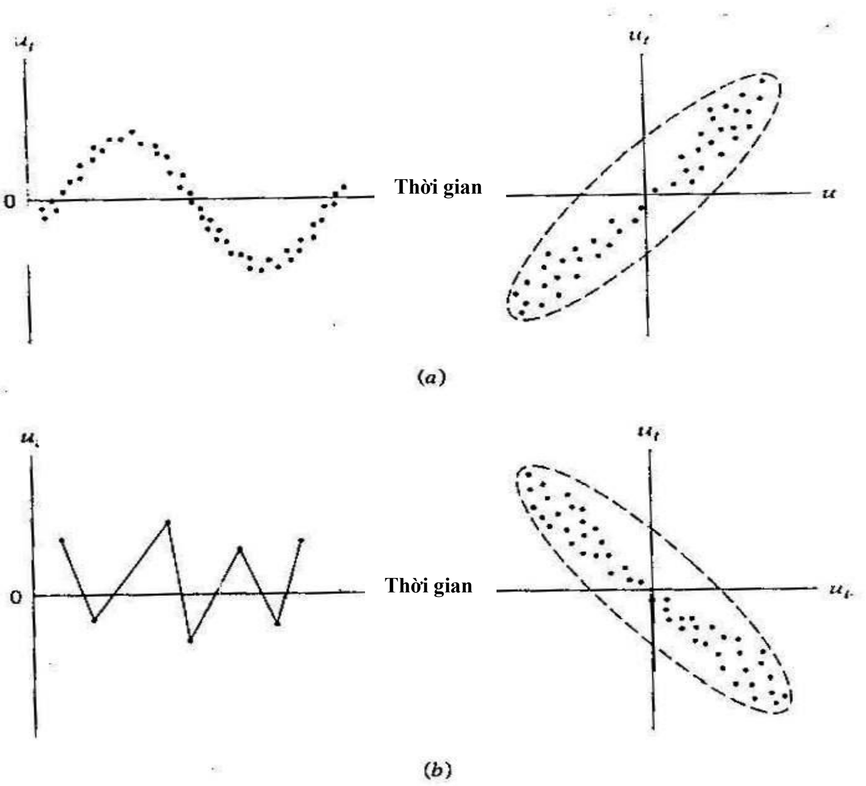
Một phép hồi qui như (12.1.7) được biết tới như là **tự hồi qui** bởi vì một trong các biến giải thích là giá trị chậm trễ của biến phụ thuộc. (Chúng ta sẽ nghiên cứu các mô hình như vậy trong Chương 17). Nguyên do của mô hình như là (12.1.7) đơn giản. Các người tiêu dùng không thay đổi thói quen tiêu dùng của mình do các nguyên nhân tâm lý, kỹ thuật hoặc thể chế. Bây giờ, nếu chúng ta bỏ qua số hạng chậm trễ trong (12.1.7), số hạng sai số kết quả sẽ phản ánh một dạng có hệ thống do sự ảnh hưởng của tiêu dùng chậm trễ lên tiêu dùng hiện tại.

"Nhào nặn" dữ liệu. Trong phân tích theo kinh nghiệm, dữ liệu thô thường được "nhào nặn". Ví dụ, trong các phép hồi qui chuỗi thời gian có liên quan tới dữ liệu từng quí, các dữ liệu như vậy thường được rút ra từ dữ liệu từng tháng bằng cách đơn giản cộng các quan sát của 3 tháng và chia tổng này cho 3. Cách lấy trung bình như vậy đưa vào dữ liệu một sự làm trơn nào đó bằng cách dàn đều các dao động trong dữ liệu hàng tháng. Vì vậy, đồ thị vẽ dữ liệu theo quí trông trơn hơn là dữ liệu quí, và sự làm trơn này có thể tự nó cho ra một dạng có hệ thống trong các nhiễu, bằng cách đưa tự tương quan vào. Một nguồn gốc khác của nhào nặn là **nội suy** và **ngoại suy** dữ liệu. Ví dụ, Điều tra dân số thực hiện từng 10 năm trong một nước này, lần cuối cùng trong năm 1990 và lần trước đó vào năm 1980. Bây giờ nếu cần thu dữ liệu cho một năm nào đó trog giai đoạn giữa các kỳ điều tra 1980-1990, thông thương trên thực tế người ta nội suy trên cơ sở các giả định đặc biệt nào đó. Tất cả mọi kỹ thuật "xoa bóp" dữ liệu như vậy có thể gắn vào dữ liệu một dạng có hệ thống mà không thể tồn tại trong dữ liệu gốc.[5]

Trước khi kết luận phần này, nên lưu ý rằng vấn đề tự tương quan thường là phổ biến hơn trong dữ liệu chuỗi thời gian, mặc dù nó có thể và có xảy ra trong dữ liệu chéo. Trong dữ liệu chuỗi thời gian, các quan sát được sắp xếp theo trật tự thời gian. Vì vậy, có khả năng có các

[5] Về vấn đề này, xin đọc William H. Greene, Phân tích kinh tế lượng, Mac Millan, in lần thứ 2, NewYork, 1993, trang 413.

tương quan nội tại giữa các quan sát liên tiếp đặc biệt là khi khoảng thời gian giữa các quan sát liên tiếp là ngắn, như là một ngày, một tuần, hoặc một tháng chứ không phải là một năm. Nói chung không có thứ tự thời gian như vậy trong dữ liệu chéo, mặc dù trong một số trường hợp có thể tồn tại một thứ tự tương tự. Do trong hồi qui chéo của chi tiêu tiêu dùng theo thu nhập trong đó các đơn vị của các quan sát là 50 bang của Hoa kỳ, có thể là dữ liệu được bố trí sao cho nó rơi vào các nhóm như là Phía Nam, Phía Tây Nam, Phía Bắc v.v... Do dạng tiêu dùng có khả năng khác nhau giữa các khu vực địa lý, mặc dù là tương tự về cơ bản trong bất cứ một khu vực nào, các phần dư đã được ước lượng từ hồi qui có thể biểu lộ một dạng có hệ thống kèm theo các khác biệt của khu vực. Điểm cần ghi nhận là, mặc dù việc xảy ra tự tương quan là hay có với dữ liệu chuỗi thời gian, nó vẫn có thể xảy ra trong dữ liệu chéo. Một số tác giả gọi tự tương quan trong dữ liệu chéo là **tự tương quan không gian**, tức là tương quan theo không gian chứ không phải là theo thời gian. Tuy nhiên, vấn đề quan trọng là cần nhớ rằng trong phân tích chéo việc sắp xếp thứ tự dữ liệu cần theo lô gich, hoặc lợi ích kinh tế nào đó, để làm cho bất cứ việc xác định xem có tồn tại tự tương quan tồn tại hay không là có ý nghĩa.



HÌNH 12.3 (a) Tự tương quan thuận (b) nghịch

Cũng cần phải lưu ý rằng tự tương quan có thể là đồng biến mà cũng có thể là nghịch biến, mặc dù hầu hết chuỗi thời gian kinh tế nói chung cho thấy tự tương quan đồng biến vì hầu hết chúng hoặc là hướng lên trên và xuống dưới theo các thời đoạn kéo dài và không cho thấy một sự chuyển động lên xuống không đổi như trong Hình 12.3b.

12.2 ƯỚC LƯỢNG OLS KHI TỒN TẠI TỰ TƯƠNG QUAN

Điều gì xảy ra với các hàm ước lượng và các phương sai của chúng nếu chúng ta đưa quan hệ tự tương quan vào các phần nhiễu bằng cách giả định rằng $E(u_iu_j) \neq 0$ $(i \neq j)$ nhưng vẫn giữ nguyên tất cả các giả định khác của mô hình cổ điển ? Chúng ta chuyển ngược lại một lần nữa về mô hình hồi qui hai biến để giải thích các ý tưởng căn bản có liên quan, cụ thể là $Y_t = \beta_1 + \beta_2 X_t + u_t$, trong đó, t ký hiệu cho dữ liệu hay quan sát vào thời đoạn t; nên nhớ rằng hiện nay chúng ta đang xử lý chuỗi thời gian.

Để làm bất kỳ điều gì tiếp tục, chúng ta cần giả định rằng cơ chế tạo ra u_t, đối với $E(u_t,u_{t+s}) \neq 0$ $(s \neq 0)$ là một giả định quá tổng quát để trở thành hữu dụng trong thực tiễn. Như một điểm xuất phát, hay là một phép xấp xỉ đầu tiên, người ta có thể giả định rằng các nhiễu được tạo ra như sau :

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad -1 < \rho < 1 \tag{12.2.1}$$

trong đó ρ được biết tới như **hệ số tự đồng phương sai** (coefficient of autocovariance) và trong đó ε_t là nhiễu ngẫu nhiên sao cho nó thoả mãn các giả định OLS chuẩn, cụ thể là,

$$\begin{aligned} E(\varepsilon_t) &= 0 \\ \text{var}(\varepsilon_t) &= \sigma^2 \\ \text{cov}(\varepsilon_t, \varepsilon_{t+s}) &= 0 \qquad s \neq 0 \end{aligned} \tag{12.2.2}$$

Sơ đồ (12.2.1) được gọi là **sơ đồ tự hồi qui bậc nhất Markov** hay còn gọi một cách đơn giản là **sơ đồ tự hồi qui bậc nhất**, thường ký hiệu là AR(1). Tên *tự tương quan* là phù hợp vì (12.2.1) có thể được giải thích như là phép hồi qui của u_t với chính nó sau khi trễ một thời đoạn. Nó là bậc nhất vì chỉ có u_t và giá trị ngay trước đó là có liên quan, tức là, độ trễ tối đa là 1. Nếu mô hình là $u_t = p_1u_{t-1} + p_2u_{t-2} + \varepsilon_t$, nó sẽ là AR(2) hoặc sơ đồ tự đồng phương sai bậc hai, và tương tự. Nhân đây, lưu ý rằng p, hệ số tự hồi qui, cũng có thể được giải thích như là *hệ số tự tương quan bậc nhất*, hoặc, chính xác hơn, là *hệ số tự tương quan có độ trễ 1*.[6]

Điều mà (12.2.1) đưa ra là sự vận động hoặc chuyển dịch của u_t bao gồm hai phần: một phần ρu_{t-1}, nó giải thích cho một dịch chuyển có hệ thống, và một phần khác ε_t đơn thuần là ngẫu nhiên.

Trước khi tiếp tục, lưu ý rằng có một tiên nghiệm là không có nguyên nhân vì sao chúng ta không thể chấp nhận AR(2) hoặc AR(3) hoặc bất cứ sơ đồ tự hồi qui có bậc cao hơn 1 trong (12.2.1). Trên thực tế, người ta có thể đã giả định rằng u_t được tạo ra bởi cơ chế như sau:

$$u_t = v_t + \lambda v_{t-1} \tag{12.2.3}$$

Trong đó v là một số hạng nhiễu ngẫu nhiên với giá trị trung bình bằng 0 và phương sai không đổi, λ là một hằng số sao cho $|\lambda| < 1$. Sơ đồ tạo sai số (12.2.3) được gọi là **trung bình trượt bậc nhất** hoặc **sơ đồ MA(1)** bởi vì nó có liên quan tới việc lấy trung bình của 2 biến ngẫu nhiên kế tiếp. Người ta cũng có thể xem xét các sơ đồ MA có bậc cao hơn.

Không chỉ có thế, người ta có thể giả định rằng u_t được tạo ra bởi một hỗn hợp của các quá trình tự hồi qui và trung bình trượt. Ví dụ, người ta có thể xem xét:

---

[6] Tên gọi này có thể được chứng tỏ dễ dàng. Theo định nghĩa, hệ số tương quan (tổng thể) giữa u_t và u_{t-1} là

$$\rho = \frac{E\{[u_t - E(u_t)][u_{t-1} - E(u_{t-1})]\}}{\sqrt{\text{var}(u_t)}\sqrt{\text{var}(u_{t-1})}}$$
$$= \frac{E(u_t u_{t-1})}{\text{var}(u_{t-1})}$$

Vì $E(u_t)=0$ đối với từng t và $\text{var}(u_t)=\text{var}(u_{t-1})$ do chúng ta đang giữ giả định về phương sai không thay đổi. Người đọc có thể thấy rằng ρ cũng là hệ số độ dốc trong phép hồi qui của u_t theo u_{t-1}.

$$u_t = \rho\, u_{t-1} + v_t + \lambda\, v_{t-1} \tag{12.2.4}$$

biểu thức được gọi, một cách thích hợp, là **sơ đồ ARMA (1,1)** do nó là một kết hợp của các sơ đồ tự hồi qui bậc nhất và trung bình trượt bậc I. Tất nhiên, các sơ đồ ARMA bậc cao hơn cũng có thể được xem xét tới. Trong chương về kinh tế lượng chuỗi thời gian (Chương 22) chúng ta sẽ trở lại chủ đề này.[7]

Hiện thời, chúng ta sử dụng sơ đồ AR(1) được cho trong (12.2.1) không chỉ vì tính đơn giản của nó mà cũng vì trong nhiều áp dụng, nó đã chứng tỏ được là hoàn toàn hữu ích. Ngoài ra, một số lượng đáng kể của nghiên cứu lý thuyết và thực nghiệm đã được thực hiện trên sơ đồ AR(1).

Bây giờ hàm ước lượng OLS của β_2, như thường lệ, là:

$$\hat{\beta}_2 = \frac{\sum x_t y_t}{\sum x_t^2} \tag{12.2.5}$$

nhưng phương sai của nó cho trong sơ đồ AR(1), bây giờ là

$$\operatorname{var}(\hat{\beta}_2)_{\text{AR1}} = \frac{\sigma^2}{\sum x_t^2} + \frac{2\sigma^2}{\sum x_t^2}\left(\rho \frac{\sum_{t=n}^{n-1} x_t x_{t+1}}{\sum_{t=1}^{n} x_t^2} + \rho^2 \frac{\sum_{t=n}^{n-2} x_t x_{t+2}}{\sum_{t=1}^{n} x_t^2} + \ldots + \rho^{n-1} \frac{x_1 x_n}{\sum_{t=1}^{n} x_t^2} \right) \tag{12.2.6}$$

trong đó $\operatorname{var}(\hat{\beta}_2)$AR1 có nghĩa là phương sai của $\hat{\beta}_2$ theo sơ đồ tự hồi qui bậc nhất. Đối chiếu công thức này với công thức thông thường khi không có tự tương quan:

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_t^2} \tag{12.2.7}$$

Một sự so sánh của (12.2.6) với (12.2.7) cho thấy rằng biểu thức trước là bằng biểu thức sau cộng với một số hạng phụ thuộc vào ρ và các đồng phương sai mẫu giữa các giá trị X đã chọn. Và nói chung chúng ta không thể nói có phải $\operatorname{var}(\hat{\beta}_2)$ nhỏ hơn hay lớn hơn $\operatorname{var}(\hat{\beta}_2)$AR1 hay không [nhưng hãy xem phương trình (12.4.1) dưới đây]. Tất nhiên, nếu ρ là 0, thì hai công thức sẽ trùng nhau, như chúng cần phải thế. (Vì sao?)

---

[7] Những gì được biết như là phương pháp Box-Jenkins trong việc lập mô hình chuỗi thời gian là dựa trên các cơ chế tạo sai số AR, MA, và ARMA.

Giả sử chúng ta tiếp tục sử dụng hàm ước lượng OLS β_2 và điều chỉnh công thức phương sai thông thường bằng cách chú ý tới sơ đồ AR(1). Nghĩa là, chúng ta sử dụng β_2 được cho bởi (12.2.5) nhưng sử dụng công thức phương sai cho bởi (12.2.6). Các tính chất của β_2 bây giờ là gì? Dễ dàng chứng minh được rằng β_2 vẫn là tuyến tính và không thiên lệch. Trên thực tế, như đã nêu ra trong Phụ lục 3A, Phần 3A.2, giả định không có tương quan chuỗi, như là giả định không có phương sai thay đổi, không được đặt ra để chứng minh rằng β_2 là không thiên lệch. β_2 vẫn là ước lượng không thiên lệch tuyến tính tốt nhất (BLUE)? Không may, điều đó không đúng; trong lớp các hàm không thiên lệch tuyến tính, nó không có phương sai cực tiểu.

Nói tóm lại, β_2, mặc dù không thiên lệch tuyến tính, không phải là hiệu quả (nói một cách tương đối, tất nhiên). Người đọc sẽ nhận thấy rằng phát hiện này là hoàn toàn tương tự với phát hiện cho rằng β_2 là kém hiệu quả hơn khi có tồn tại phương sai thay đổi. Ở đó chúng ta đã thấy rằng nó là hàm ước lượng β^*_2 bình phương tối thiểu được cho trong (11.3.8), một trường hợp đặc biệt của hàm ước lượng bình phương tối thiểu (GLS) tổng quát, đó là hàm có hiệu quả. Trong trường hợp tự tương quan, chúng ta có thể tìm được một hàm ước lượng BLUE hay không? Câu trả lời là có, như có thể thấy từ thảo luận trong phần tiếp theo.

12.3 HÀM ƯỚC LƯỢNG KHÔNG THIÊN LỆCH TUYẾN TÍNH TỐT NHẤT (BLUE) KHI CÓ TỰ TƯƠNG QUAN

Tiếp tục với mô hình hai biến và giả định quá trình AR(1), chúng ta có thể chỉ ra rằng hàm ước lượng BLUE của $\hat{\beta}_2$ được cho bởi biểu thức sau:[8]

$$\hat{\beta}_2^{GLS} = \frac{\sum_{t=2}^{n}(x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{\sum_{t=2}^{n}((x_t - \rho x_{t-1})^2} + C \tag{12.3.1}$$

trong đó C là hệ số hiệu chỉnh có thể bỏ qua trong thực tế. Nên nhớ là chỉ số t bây giờ thực hiện từ t = 2 tới t = n. Và phương sai của nó được cho bởi:

$$\operatorname{var}\hat{\beta}_2^{GLS} = \frac{\sigma^2}{\sum_{t=2}^{n}((x_t - \rho x_{t-1})^2} + D \tag{12.3.2}$$

trong đó D cũng là hệ số hiệu chỉnh có thể bỏ qua trên thực tế. (Xem bài tập 12.18.)

[8] Để biết các chứng minh, hãy xem Jan Kmenta, Elements of Econometrics, Macmillan, NewYork, 1971, trang 274-275. Hệ số hiệu chỉnh C liên quan tới quan sát thứ nhất (Y_1, X_1). Về điểm này xin xem bài tập 12.18.

Hàm ước lượng β_2^{GLS}, như là chỉ số cho thấy, là giá trị thu được bởi phương pháp GLS. Như đã lưu ý trong Chương 11, trong GLS chúng ta kết hợp bất cứ thông tin bổ sung nào mà ta có (tức là, về bản chất của phương sai thay đổi hoặc của tự hồi qui) một cách trực tiếp vào quá trình ước lượng bằng cách biến đổi các biến; trong khi đó ở trong OLS thông tin bên lề như vậy không được xem xét tới một cách trực tiếp. Như là người đọc có thể thấy, hàm ước lượng GLS của β_2 được cho trong (12.3.1) kết hợp thông số tự tương quan ρ trong công thức đang ước lượng, trong khi công thức OLS cho trong (12.2.5) bỏ qua nó một cách đơn giản. Về mặt trực giác, đây là nguyên nhân vì sao hàm ước lượng GLS là BLUE mà không phải là hàm ước lượng OLS – hàm ước lượng GLS làm cho thông tin đang có trở nên hữu ích nhất.[9] Rất cần bổ sung thêm rằng nếu $\rho = 0$, không có thông tin bổ sung cần được xem xét và vì vậy cả hai hàm ước lượng GLS và OLS là như nhau.

Tóm lại, khi có tự tương quan, hàm ước lượng GLS được cho trong (12.3.1) là BLUE, và phương sai cực tiểu bây giờ được cho bởi (12.3.2) chứ không phải bởi (12.2.6) và hiển nhiên là không phải bởi (12.2.7).

Điều gì xảy ra nếu chúng ta cứ vô tư tiếp tục công việc với qui trình OLS thông thường mà không xem xét đến tự tương quan? Câu trả lời được cho trong phần sau đây.

12.4 CÁC HẬU QUẢ CỦA VIỆC SỬ DỤNG OLS KHI CÓ TỰ TƯƠNG QUAN.

Như trong trường hợp về phương sai thay đổi, khi có tự tương quan, các hàm ước lượng OLS vẫn là tuyến tính không thiên lệch và nhất quán, nhưng chúng không còn là hiệu quả (tức là có phương sai nhỏ nhất). Điều gì sau đó xảy ra cho các qui trình kiểm định giả thiết thông thường của chúng ta nếu chúng ta tiếp tục sử dụng các hàm ước lượng OLS?. Một lần nữa, như trong trường hợp phương sai thay đổi, chúng ta phân biệt 2 trường hợp. Với mục đích sư phạm, chúng ta vẫn tiếp tục làm việc với mô hình hai biến, mặc dù thảo luận sau đây có thể được mở rộng sang các phép hồi qui đa biến mà không cần lo lắng gì nhiều.[10]

Ước lượng OLS có xét đến Tự Hồi qui.

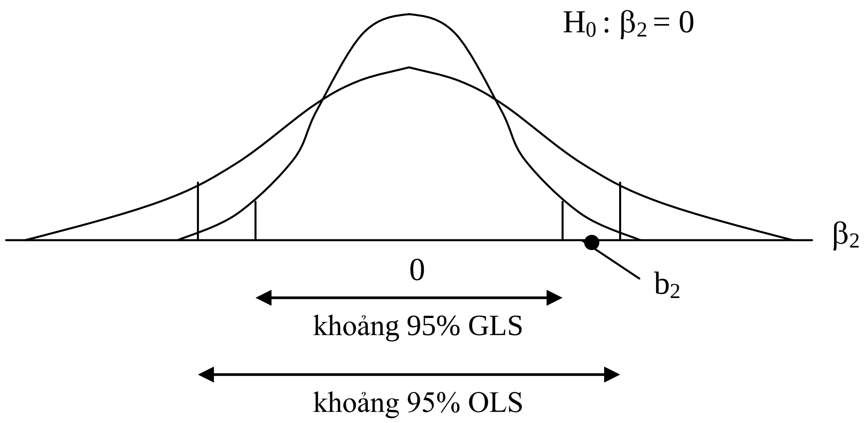
Như đã lưu ý, β_2 không là BLUE, và thậm chí nếu chúng ta sử dụng var$(\beta_2)_{AR1}$, các khoảng tin cậy được tìm ra từ đó có khả năng rộng hơn các khoảng dựa trên qui trình GLS. Như Kmenta chỉ ra, kết quả này có khả năng là đúng thậm chí ngay cả khi cỡ của mẫu tăng lên vô cùng[11]. Tức là, β_2 không phải là hiệu quả theo kiểu tiệm cận. Ý nghĩa của phát hiện này đối với việc kiểm định giả thiết là rõ ràng: chúng ta có khả năng tuyên bố rằng một hệ số là không có ý nghĩa về mặt thống kê (tức là, không khác không) thậm chí khi trong thực tế (tức là, dựa trên qui trình GLS chính xác) nó có thể là như vậy. Sự khác biệt này có thể thấy rõ từ Hình 12.4. Trong hình này

[9] Chứng minh chính thức rằng β_2^{GLS} là BLUE có thể được thấy trong Kmenta, ibid. Nhưng chứng minh đại số tẻ nhạt có thể được rút gọn đáng kể khi sử dụng khái niệm ma trận. Xem J. Jonhson, Econometric Methods, in lần thứ ba, McGraw-Hill, NewYork, 1984, trang 291-293.

[10] Nhưng đại số ma trận trở nên gần như một sự cần thiết để tránh khỏi các biến đổi đại số tẻ nhạt.

[11] Xem Kmenta, op. cit., trang 277-278

chúng ta chỉ ra các khoảng tin cậy 95% của OLS [AR(1)] và GLS khi giả định rằng β_2 thực = 0. Hãy xem xét hàm ước lượng cụ thể β_2, coi như là b_2. Do b_2 nằm trong khoảng tin cậy OLS, chúng ta có thể chấp nhận giả thiết rằng β_2 thực là 0 với độ tin cậy 95%. Nhưng nếu chúng ta phải dùng khoảng tin cậy GLS (chính xác), chúng ta có thể bác bỏ *giả thiết không* rằng β_2 thực là 0, vì b_2 nằm trong vùng bác bỏ.



HÌNH 12.4. Các khoảng tin cậy 95% GLS và OLS

Thông điệp là: Để xác lập nên các khoảng tin cậy và kiểm định các giả thiết, người ta nên sử dụng GLS chứ không phải OLS, mặc dù các hàm ước lượng này được rút ra từ hàm sau là không thiên lệch và nhất quán.

Ước lượng OLS không quan tâm tới sự hồi qui.

Tình thế này là rất nghiêm trọng về mặt tiềm năng nếu chúng ta không chỉ sử dụng β_2 mà lại còn tiếp tục sử dụng var (β_2) = $\sigma^2/ \Sigma x^2_t$, điều này hoàn toàn không quan tâm tới vấn đề tự tương quan. Nghĩa là, chúng ta tin tưởng một cách sai lầm rằng các giả định thông thường của mô hình cổ điển vẫn đúng. Các sai số sẽ xuất hiện do các nguyên nhân sau đây:

1. Phương sai phần dư $\sigma^2/ \Sigma u^2_t / (n – 2)$ có khả năng bị ước lượng thấp xuống so với σ^2 thực.
2. Kết quả là chúng ta có khả năng ước lượng quá cao R^2.
3. Ngay khi nếu σ^2 là không bị ước lượng nhỏ đi, var (β_2) có thể ước lượng var $(\beta_2)_{AR1}$ nhỏ đi [Phương trình 12.2.6], phương sai của nó khi có tự tương quan (bậc 1), mặc dù đại lượng cuối là không hiệu quả so với var $(\beta_2)^{GLS}$.
4. Vì vậy, các kiểm định về mức ý nghĩa t và F thông thường không còn hiệu lực nữa, và nếu áp dụng thì chúng có khả năng cho ta các kết luận sai lạc một cách nghiêm trọng về mức ý nghĩa thống kê của các hệ số hồi qui đã ước lượng.

Để xác lập một số trong các tỉ lệ này, chúng ta hãy quay về mô hình 2 biến. Chúng ta đã biết từ chương 3 rằng khi có giả định cổ điển.

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{(n-2)}$$

cho ta một hàm ước lượng không thiên lệch của $\hat{\sigma}^2$, tức là E ($\hat{\sigma}^2$) = σ^2. Nhưng nếu có quan hệ tự tương quan, được cho bởi AR (1), có thể chỉ ra rằng

$$E(\hat{\sigma}^2) = \frac{\hat{\sigma}^2\{n - [2/(1-\rho)] - 2\rho r\}}{n-2} \qquad (12.4.1)$$

trong đó $r = \sum_{t=1}^{n-t} x_t x_{t-1} / \sum_{t=1}^{n} x_t^2$, nó có thể được giải thích như là hệ số tương quan (mẫu) giữa các giá trị liên tục của x.[12] Nếu ρ và r đều là dương (không phải là một giả định có khả năng đối với đa số chuỗi thời gian kinh tế), nhìn bề ngoài, (12.4.1) có vẻ như E (σ^2) < σ^2; tức là, công thức phương sai phần dư thông thường, về mặt trung bình, sẽ ước lượng thấp đi σ^2 thực. Nói cách khác, $\hat{\sigma}^2$ sẽ là thiên lệch theo hướng đi xuống. Không cần phải nói gì, các thiên lệch này trong σ^2 sẽ được chuyển sang var (β_2) do trên thực tế chúng ta ước lượng đại lượng sau bằng công thức $\sigma^2 / \Sigma x_t^2$,

Nhưng thậm chí nếu $\hat{\sigma}^2$ không bị ước lượng nhỏ đi thì var (β_2) vẫn là một hàm ước lượng thiên lệch của var $(\beta_2)_{AR1}$, điều đó có thể thấy ngay bằng cách so sánh (12.2.6) với (12.2.7), [13] vì 2 công thức không giống như nhau. Thực tế là, nếu ρ dương (nó là có thực cho đa số chuỗi thời gian kinh tế) và các X có tương quan đồng biến (cũng là có thực cho đa số chuỗi thời gian kinh tế), thì rõ ràng là

$$\text{var}((\beta_2) < \text{var}(\beta_2)_{AR1} \qquad (12.4.2)$$

tức là, phương sai OLS thông thường của β_2 ước lượng nhỏ đi phương sai của nó khi có AR(1). Vì vậy, nếu chúng ta sử dụng var (β_2), chúng ta sẽ vi phạm sự chính xác hoặc sự đúng đắn (tức là, ước lượng nhỏ đi sai số chuẩn) của hàm ước lượng β_2. Kết quả là, khi tính toán tỉ số t như là t = β_2 /se (β_2) (dưới giả thiết là β_2 = 0), chúng ta sẽ ước lượng lớn lên giá trị của t, và vì vậy sẽ có ý nghĩa về mặt thống kê của β_2 đã ước lượng. Tình thế có khả năng xấu đi nếu σ^2 được ước lượng nhỏ đi một cách bổ sung nữa, như đã thấy trước đây.

Để thấy OLS có khả năng ước lượng nhỏ đi σ^2 và phương sai của β_2 như thế nào, chúng ta hãy thực hiện **thử nghiệm Monte Carlo** sau đây. Giả sử trong mô hình hai biến chúng ta "biết" rằng β_1 = 1 và β_2 = 0,8 trên thực tế. Vì vậy, PRF ngẫu nhiên là

$$Y_t = 1{,}0 + 0{,}8\ X_t + u_t \qquad (12.4.3)$$

Do đó,

$$E\ (Y_t / X_t) = 1{,}0 + 0{,}8\ X_t \qquad (12.4.4)$$

nó cho ta một đường hồi qui tổng thể thực. Chúng ta hãy giả sử rằng u_t được tạo ra bởi sơ đồ tự hồi qui bậc nhất như sau:

$$u_t = 0{,}7\ u_{t-1} + \varepsilon_t \qquad (12.4.5)$$

---

[12] Xem S.M. Goldfield và R.E. Quandt, Các phương pháp phi tuyến tính trong Kinh tế lượng, North Holland & Publishing Company, Amsterdam, 1972, trang 183. Nhân đây, lưu ý rằng nếu các sai số có tự tương quan đồng biến thì giá trị R^2 có xu hướng có các thiên lệch lên trên, tức là, nó hay trở nên lớn hơn là R^2 khi không có tương quan như vậy.

[13] Về chứng minh chính thức, xin xem Kmenta, op. cit., trang 281.

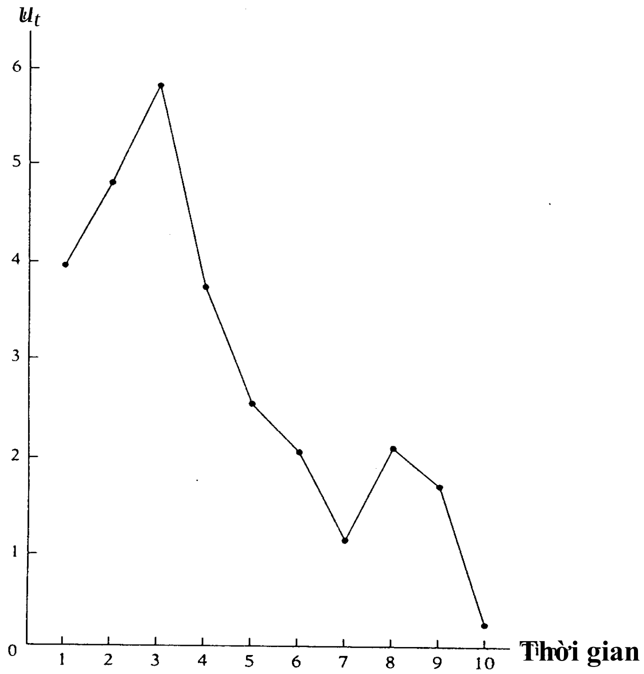
trong đó ε_t thỏa mãn mọi giả định OLS. Để thuận tiện, chúng ta giả định tiếp theo rằng ε_t có phân phối chuẩn với giá trị trung bình 0 và phương sai đơn vị (=1). Phương trình (12.4.5) yêu cầu một cách tất yếu rằng các nhiễu liên tục là tương quan đồng biến, với hệ số tự tương quan là 0,7, là một độ phụ thuộc hơi cao.

Bây giờ, bằng cách sử dụng một bảng các số chuẩn hóa ngẫu nhiên với giá trị trung bình 0 và phương sai đơn vị, chúng ta đã tạo ra 10 số ngẫu nhiên được nêu trong Bảng 12.1 bởi sơ đồ (12.4.5). Để bắt đầu theo sơ đồ, chúng ta cần xác định giá trị ban đầu của u, chẳng $u_0 = 5$.

BẢNG 12.1
Một ví dụ về giả thiết của các số hạng sai số tự tương quan đồng biến.

	ε^*_t	$u_t = 0,7\ u_{t-1} + \varepsilon_t$
0	0	u_0 = 5 (giả định)
1	0,464	u_1 = 0,7(5) + 0,464 = 3,964
2	2,026	u_2 = 0,7(3,964) + 2,0262 = 4,8008
3	2,455	u_3 = 0,7(4,8010) + 2,455 = 5,8157
4	-0,323	u_4 = 0,7(5,8157) – 0,323 = 3,7480
5	-0,068	u_5 = 0,7(3,7480) – 0,068 = 2,5556
6	0,296	u_6 = 0,7(2,5556) + 0,296 = 2,0849
7	-0,288	u_7 = 0,7(2,0849) – 0,288 = 1,1714
8	1,298	u_8 = 0,7(1,1714) + 1,298 = 2,1180
9	0,241	u_9 = 0,7(2,1180) + 0,241 = 1,7236
10	-0,957	u_{10} = 0,7(1,7236) – 0,957 = 0,2495

*Lấy từ A *Million Random Digits* (Một triệu chữ số ngẫu nhiên) *và One Hundred Thousand Deviates,* (Một trăm ngàn độ lệch), Rand Corporation, Santa Monica, Calif., 1950.



HÌNH 12.5 Mối tương quan tạo ra bởi sơ đồ $u_t = 0,7\ u_{t-1} + \varepsilon_t$ (Bảng 12.1)

Vẽ các giá trị u_t đã được tạo ra trong Bảng 12.1, chúng ta được Hình 12.5, nó cho thấy lúc đầu, mỗi u_t kế tiếp cao hơn giá trị trước đó của nó và sau đó, nói chung nó lại nhỏ hơn giá trị trước đó. Tất cả về tổng thể cho thấy một mối tự tương quan đồng biến.

Bây giờ giả sử các giá trị của X là cố định ở tại, 1,2,3,…., 10. Tiếp theo, với các X đã cho này, chúng ta có thể tạo ra một mẫu gồm 10 giá trị của Y từ (12.4.3) và các giá trị của u_t đã cho trong Bảng 12.1. Các chi tiết cho trong Bảng 12.2. Sử dụng dữ liệu trong Bảng 12.2, nếu chúng ta hồi qui Y theo X, chúng ta có phép hồi qui (mẫu) sau:

$$\begin{aligned} Y_t &= 6,5452 \quad + \quad 0,3051X_t \\ &\quad\ (0,6153) \qquad\quad (0,0992) \\ t &= (10,6366) \qquad\quad (3,0763) \qquad\qquad (12.4.6) \\ r^2 &= 0,5419 \qquad\quad \sigma^2 = 0,8114 \end{aligned}$$

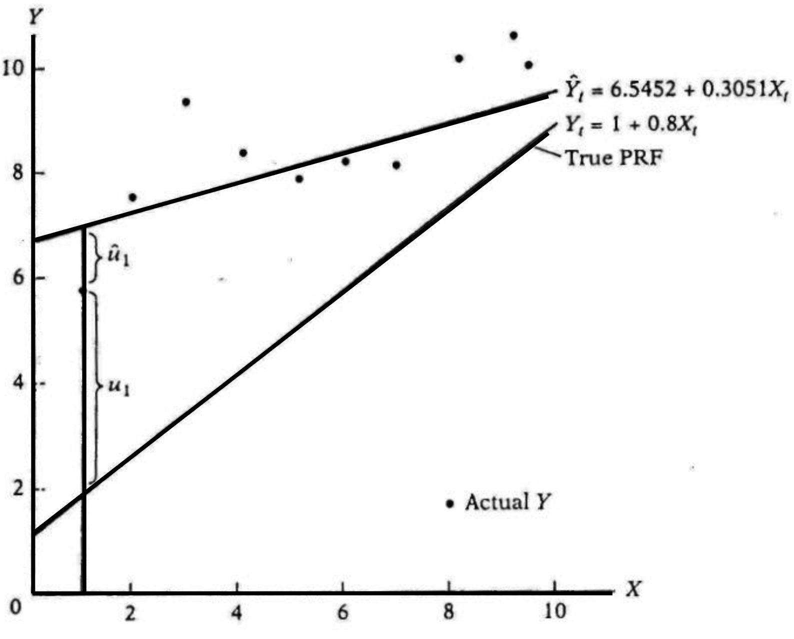
trong khi đường hồi qui thực được cho bởi (12.4.4). Cả hai đường hồi qui được cho trong Hình 12.6, chúng cho thấy rõ đường hồi qui thích hợp lệch khỏi đường hồi qui thực nhiều như thế nào; nó ước lượng ít đi một cách nghiêm trọng hệ số góc thực nhưng lại ước lượng tăng lên tung độ gốc thực. (Nhưng hãy lưu ý rằng các hàm ước lượng OLS vẫn không bị thiên lệch).

BẢNG 12.2
Tạo các giá trị mẫu Y.

X_t	u^*_t	$Y_t = 1{,}0 + 0{,}8\ X_t + u_t$
1	3,9640	$Y_1 = 1{,}0 + 0{,}8(1) + 3{,}9640 = 5{,}7640$
2	4,8010	$Y_2 = 1{,}0 + 0{,}8(2) + 4{,}8008 = 7{,}4008$
3	5,8157	$Y_3 = 1{,}0 + 0{,}8(3) + 5{,}8157 = 9{,}2157$
4	3,7480	$Y_4 = 1{,}0 + 0{,}8(4) + 3{,}7480 = 7{,}9480$
5	2,5556	$Y_5 = 1{,}0 + 0{,}8(5) + 2{,}5556 = 7{,}5556$
6	2,0849	$Y_6 = 1{,}0 + 0{,}8(6) + 2{,}0849 = 7{,}8849$
7	1,1714	$Y_7 = 1{,}0 + 0{,}8(7) + 1{,}1714 = 7{,}7714$
8	2,1180	$Y_8 = 1{,}0 + 0{,}8(8) + 2{,}1180 = 9{,}5180$
9	1,7236	$Y_9 = 1{,}0 + 0{,}8(9) + 1{,}7236 = 9{,}9236$
10	0,2495	$Y_{10} = 1{,}0 + 0{,}8(10) + 0{,}2495 = 9{,}2495$

* Thu từ Bảng 12.1

Hình 12.6 cũng cho thấy vì sao phương sai thực của u_i có khả năng bị ước lượng ít đi bởi hàm ước lượng σ^2, hàm này được tính toán từ u_i. u_i nói chung gần với đường thích hợp (là đường có được do qui trình OLS), nhưng chệch đáng kể so với PRF thực. Do đó, chúng không cho ta một bức tranh chính xác của u_i.



HÌNH 12.6 PRF thực và đường hồi qui ước lượng đối với dữ liệu trong Bảng 12.2

BẢNG 12.3
Mẫu các giá trị của Y với tương quan chuỗi zéro

X_t	$\varepsilon_t = u^*_t$	$Y_t = 1{,}0 + 0{,}8\,X_t + \varepsilon_t$
1	0,464	2,264
2	2,026	4,626
3	2,455	5,855
4	-0,323	3,877
5	-0,068	4,932
6	0,296	6,096
7	-0,288	6,312
8	1,298	8,698
9	0,241	8,441
10	-0,957	8,043

* Do không có tự tương quan, u_t và ε_t như nhau. $\boldsymbol{\varepsilon_t}$ lấy từ Bảng 12.1

Để thu được bản chất nào đó về việc ước lượng ít đi cho σ^2 thực, giả sử chúng ta thực hiện một thử nghiệm lấy mẫu khác. Giữ X_t và ε_t như cho trong Bảng 12.1 và 12.2, chúng ta hãy giả định $\rho = 0$, tức là, không có tự tương quan. Mẫu mới gồm các giá trị của Y vì vậy được tạo ra như trong Bảng 12.3.

Phép hồi qui dựa trên Bảng 12.3 sẽ như sau:

$$\begin{aligned} Y_t &= 2{,}5345 \quad + \quad 0{,}6145X_t \\ &\quad (0{,}6796) \qquad (0{,}1087) \\ t &= (3{,}7910)\ (5{,}6541) \qquad\qquad\qquad (12.4.6) \\ r^2 &= 0{,}7997 \qquad\quad \sigma^2 = 0{,}9752 \end{aligned}$$

Phép hồi qui này gần hơn " giá trị thực" nhiều vì các Y bây giờ thực sự là ngẫu nhiên. Lưu ý rằng σ^2 đã tăng lên từ 0,8114 ($\rho = 0{,}7$) tới 0,9752 ($\rho = 0$). Đồng thời cũng lưu ý rằng các sai số chuẩn của β_1 và β_2 đã tăng lên. Kết quả này phù hợp với các kết quả lý thuyết đã được xem xét trước đây.

12.5 PHÁT HIỆN TỰ TƯƠNG QUAN

Như đã trình bày trong Phần 12.4, tự tương quan có khả năng là một vấn đề nghiêm trọng. Các số đo bổ sung vì vậy chắc chắn là phù hợp. Tất nhiên, trước khi người ta làm bất cứ điều gì, điều chính yếu là tìm xem có tồn tại quan hệ tự tương quan hay không trong tình huống đã cho. Trong phần này chúng ta sẽ xem xét một vài kiểm định thường được sử dụng đối với tương quan chuỗi.

Phương pháp Đồ thị

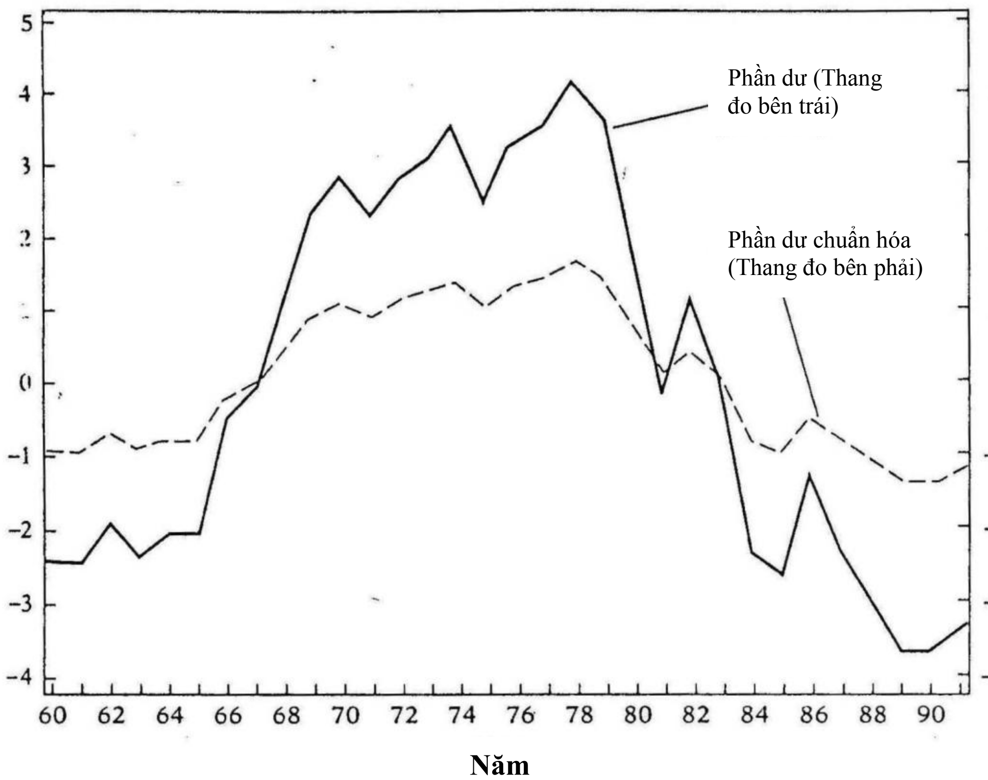
Hãy nhớ lại rằng giả định về mối quan hệ phi tự tương quan của mô hình cổ điển có liên quan tới các nhiễu tổng thể u_t, chúng không thể quan sát được một cách trực tiếp. Thay vào đó, điều chúng ta có là các biến thay thế của chúng, các phần dư u_t, chúng có thể thu được từ qui trình OLS thông thường. Mặc dù u_t không phải là u_t [14] rất thường xảy ra là một kiểm tra bằng mắt của các u cho chúng ta một vài hiểu biết nào đó về sự tồn tại có khả năng của tự tương

[14] Thậm chí nếu các nhiễu u_t có phương sai không đổi và không có tương quan, các hàm ước lượng, u_t, có phương sai thay đổi và tự tương quan . Về vấn đề này, xin xem G.S. Maddala, Introduction to Econometries (Nhập môn Kinh tế lượng), Macmillan, in lần thứ 2, New York, 1992,trang 480-481.

quan trong các u. Thực sự, một kểm tra bằng mắt của u_t (hoặc u^2_t có thể cho ta thông tin hữu ích không chỉ về tự tương quan, mà cả về phương sai thay đổi (như ra đã thấy trong chương trước), tính không đầy đủ của mô hình, hoặc các thiên lệch về đặc trưng, như chúng ta sẽ thấy trong chương sau. Như một tác giả ghi nhận:

> Tầm quan trọng của việc tạo ra và phân tích đồ thị (của các phần dư) như là một phần chuẩn của phân tích thống kê không thể được nhấn mạnh thái quá. Ngoài việc đôi khi cung cấp một sự dễ dàng để hiểu tóm lược của một vấn đề phức tạp, chúng cho phép xem xét đồng thời các dữ liệu về tổng thể trong khi bộc lộ rõ hành vi của các trường hợp riêng.[15].

Có nhiều cách xem xét các phần dư. Chúng ta có thể đơn giản vẽ chúng theo thời gian, **đồ thị theo thứ tự thời gian**, như chúng ta đã vẽ trong Hình 12.7, nó cho thấy các phần dư thu được từ phép hồi qui tiền công theo năng suất tại Mỹ trong giai đoạn 1960-1991 từ dữ liệu cho trong Phụ lục 12A.



HÌNH 12.7 Các phần dư và phần dư chuẩn hoá từ hồi qui tiền công theo năng suất: Xem Phụ lục 12A

Các giá trị của các phần dư này được cho trong Bảng 12.4 (Đồng thời xem Phụ lục 12A, Phần 12.A.1). Một cách khác, chúng ta có thể vẽ các phần dư chuẩn hóa theo thời gian, nó cũng được trình bày ở Hình 12.7 và Bảng 12.4. Các phần dư chuẩn hóa đơn giản là u_t chia cho σ, sai số chuẩn của ước lượng (= $\sqrt{\hat{\sigma}^2}$). Lưu ý rằng u_t cũng như σ được đo bằng các đơn vị mà Y được đo. Các giá trị u_t/σ sẽ là các số thuần túy (không có đơn vị đo) và vì vậy có thể được so sánh với các phần dư chuẩn hóa của các phép hồi qui khác. Ngoài ra, các phần dư chuẩn, giống như các

[15] Standford Weisberg, Applied Lineas Regression (Hồi qui tuyến tính ứng dụng), John Wiley & Sons, New york, 1980, trang 120.

u_t, có giá trị trung bình bằng 0 (vì sao?) và phương sai *xấp xỉ* bằng 1[16]. Trong các mẫu lớn $u_t/\hat{\sigma}$) có phân bố như chuẩn với giá trị trung bình bằng 0 và phương sai bằng 1.

Xem xét đồ thị theo thứ tự thời gian như cho trong Hình 12.7, chúng ta thấy rằng cả u_t và u_t chuẩn hóa đều cho ta một dạng tương tự với hình 12.1d, gợi cho ta rằng có lẽ u_t không phải là ngẫu nhiên.

BẢNG 12.4
Các phần dư u_t và phần dư chuẩn hóa (u_t/σ) từ phép hồi qui theo năng suất tại Mỹ 1960-1991

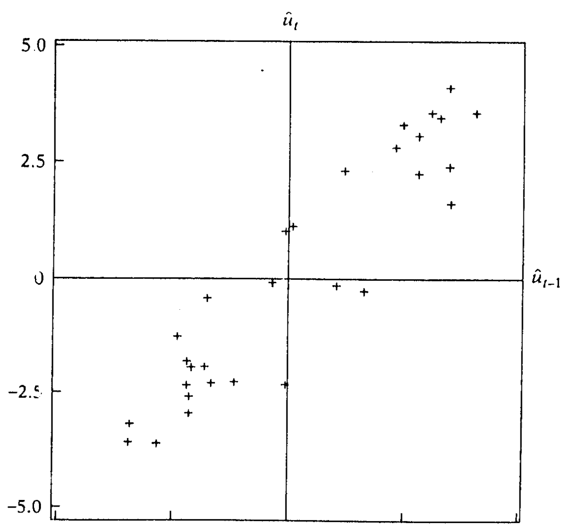
Năm	u_t	u_t/σ	u_{t-1}
1960	-2,409993	-0,922624	NA
1961	-2,433600	-0,931661	-2,409993
1962	-1,876264	-0,718295	-2,433600
1963	-2,342697	-0,896860	-1,876264
1964	-2,032917	-0,778266	-2,342697
1965	-2,032748	-0,778202	-2,032917
1966	-0,513517	-0,196591	-2,032748
1967	-0,132402	-0,050688	-0,513517
1968	1,063037	0,406965	-0,132402
1969	2,239265	0,857263	1,063037
1970	2,767930	1,059653	2,239265
1971	2,220547	0,850098	2,767930
1972	2,754114	1,054364	2,220547
1973	3,011447	1,152880	2,754114
1974	3,468447	1,327834	3,011447
1975	2,387666	0,914076	3,468447
1976	3,221236	1,233194	2,387666
1977	3,426122	1,311631	3,221236
1978	4,040456	1,546818	3,426122
1979	3,530841	1,351720	4,040456
1980	1,597454	0,611557	3,530841
1981	-0,254827	-0,097556	1,597454
1982	0,964233	0,369140	-0,254827
1983	-0,154652	-0,059?06	0,964233
1984	-2,359201	-0,903179	-0,154652
1985	-2,673363	-1,023450	-2,359201
1986	-1,354143	-0,518410	-2,673363
1987	-2,344527	-0,897561	-1,354143

[16] Thực sự là nó được gọi là các phần dư được **phân phối Student-t hóa** có phương sai bằng 1. Nhưng trên thực tế, các phần dư chuẩn hóa nói chung sẽ cho các bức tranh y như là các phần dư **phân phối Student-t hóa** và vì vậy, chúng ta có thể tin chúng được. Về điều này, xem Normal Draper và Harry Smith, Applied Regression Analysis (Phân tích Hồi qui Ứng dụng), in lần thứ 2, John Wiley & Sons, New York, 1981, trang 144.

1988	-3,053972	-1,169159	-2,344527
1989	-3,725473	-1,426232	-3,053972
1990	-3,687362	-1,411642	-3,725473
1991	-3,311136	-1,267610	-3,687362

Nguồn: u_t thu được từ hồi qui tiền công theo năng suất; xem Phụ lục 12A, Phần 12A.1.
Giá trị của σ = 2,6121.

Để nhìn vấn đề này một cách khác đi, chúng ta có thể vẽ u_t theo u_{t-1} , tức là, phần dư tại thời điểm t theo giá trị của chính nó tại thời điểm (t-1), một loại kiểm định theo kinh nghiệm của AR (1). Nếu các phần dư không có tính ngẫu nhiên, chúng ta sẽ thu được hình ảnh tương tự như trong Hình 12.3. Khi chúng ta vẽ u_t theo u_{t-1} đối với hồi qui tiền công-năng suất của chúng ta, chúng ta thu được hình ảnh như trong Hình 12.8; các dữ liệu cơ sở được cho trong Bảng 12.4. Như hình này cho thấy, đa số các phần dư cụm lại ở phần tử thứ nhất (đông bắc) và thứ 3 (tây nam), cho thấy rất rõ rằng có mối tương quan đồng biến trong các phần dư. Sau này, chúng ta sẽ thấy chúng ta có thể sử dụng hiểu biết này như thế nào để loại bỏ vấn đề tự tương quan (Xem Phần 12.6).



HÌNH 12.8 Các phần dư u_t theo u_{t-1} từ hồi qui tiền công - năng suất.

Phương pháp đồ thị mà chúng ta vừa thảo luận về bản chất thực sự mang tính chủ quan hoặc mang tính định tính. Nhưng có nhiều kiểm định mang tính định lượng có thể được sử dụng để bổ trợ cho phương pháp định tính thuần túy này. Bây giờ chúng ta xem xét một số các kiểm định này.

Kiểm định chạy.

Nếu chúng ta xem xét lại Hình 12.7, chúng ta nhận thấy có đặc tính kỳ lạ. Đầu tiên, chúng ta có nhiều phần dư có giá trị âm, sau đó là một loạt các phần dư có giá trị dương, và cuối cùng là nhiều phần dư lại có giá trị âm. Nếu các phần dư là đơn thuần ngẫu nhiên, liệu chúng ta có thể qua sát được một dạng như vậy? Một cách trực giác, nó có vẻ như không có khả năng. Trực giác này có thể được kiểm tra lại bởi cái gọi là **kiểm định chạy**, đôi khi còn được biết tới như là **kiểm định Geary**, một kiểm định phi tham số.[17] Để giải thích kiểm định này, chúng ta hãy đơn giản ghi ra các dấu (+ hay -) của các phần dư từ hồi qui tiền công - năng suất đã cho trong Bảng 12.4, Cột 1.

$$(--------)\ (+++++++++++++)\ (-)\ (+)(---------) \qquad (12.5.1)$$

Như vậy là 8 phần dư có giá trị âm, tiếp theo 13 phần dư có giá trị dương, tiếp theo là 1 phần dư có giá trị âm và1 phần dư có giá trị dương, sau đó là 9 phần dư có giá trị âm. Bây giờ chúng ta định nghĩa một sự kiện chạy như là một chuỗi của một ký hiệu hoặc một đặc tính, chẳng hạn như là + hoặc –. Tiếp theo, chúng ta định nghĩa **chiều dài của một sự kiện chạy** là số các phần tử trong nó. Trong chuỗi (12.5.1) có 5 cuộc chạy: một cuộc gồm 8 dấu trừ (nghĩa là chiều dài = 8), một cuộc chạy gồm 13 dấu cộng (tức là chiều dài bằng 13), một cuộc chạy gồm 1 dấu trừ (tức là chiều dài = 1), và một cuộc chạy gồm 1 dấu cộng (từ là chiều dài = 1), và một cuộc chạy gồm 9 dấu trừ (từ là chiều dài = 9). Để giúp nhìn cho rõ hơn, chúng ta đã trình bày các cuộc chạy khác nhau trong các dấu ngoặc.

Bằng cách xem xét các cuộc chạy có hành vi ra sao trong một chuỗi ngẫu nhiên nghiêm ngặt của các quan sát, người ta có thể dẫn giải một thử nghiệm về độ ngẫu nhiên của các cuộc chạy. Chúng ta đặt câu hỏi này: 5 cuộc chạy được quan sát trong ví dụ minh họa của chúng ta bao gồm 32 quan sát là quá nhiều hay quá ít khi so sánh với các cuộc chạy được kỳ vọng trong một chuỗi ngẫu nhiên nghiêm ngặt của 32 quan sát? Nếu là quá nhiều cuộc chạy, điều đó có thể có nghĩa là trong ví dụ của chúng ta các u đổi dấu thường xuyên, vì vậy cho thấy tương quan chuỗi nghịch biến (Như Hình 12.3b).Tương tự, nếu là quá ít cuộc chạy, chúng có thể đề xuất mối tự tương quan đồng biến, như trong Hình 12.3a. Một tiên nghiệm, Hình 12.7 cho thấy tương quan đồng biến trong các phần dư.

Bây giờ gọi

n = tổng số quan sát = $n_1 + n_2$
n_1 = số ký hiệu + (tức là, các phần dư +)
n_2 = số ký hiệu - (tức là, các phần dư -)
k = số cuộc chạy

Sau đó theo *giả thiết không* cho rằng các kết quả liên tiếp (ở đây là các phần dư) là độc lập, và việc giả định là $n_1 > 10$ và $n_2 > 10$, số cuộc chạy có phân phối chuẩn (một cách gần đúng) với

[17] Trong các **kiểm định phi thông số** chúng ta không lập ra các giả định về phân bố xác suất mà từ đó rút ra các quan sát. Trong kiểm định Geary, xem R.C. Geary, "Hiệu quả tương đối của việc tính các thay đổi dấu khi đánh giá tự hồi qui phần dư trong Hồi qui bình phương nhỏ nhất", Biometrika, Tập 57, 1970, trang 123-127

$$\text{giá trị trung bình}: E(k) = \frac{2n_1 n_2}{n_1 + n_2} + 1$$
$$\text{phươngsai}: \sigma_k^2 = \frac{2n_1 n_2(2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)} \tag{12.5.2}$$

Nếu giả thiết về tính ngẫu nhiên là ổn định, chúng ta sẽ kỳ vọng k, số các cuộc chạy thu được trong bài tập, nằm giữa [E (k) ± 1,96 σ_k] với độ tin cậy 95% (Vì sao?) Vì vậy, chúng ta có qui tắc này:

> **Qui tắc quyết định.** Không bác bỏ *giả thiết không* về tính ngẫu nhiên với độ tin cậy 95% nếu [E (k) – 1,96 $\sigma_k \le k \le$ E (k) + 1,96 σ_k]; bác bỏ *giả thiết không* nếu k ước lượng nằm ngoài các giới hạn này.

Trong ví dụ của chúng ta, n_1 = 14 và n_2 = 18. Vì vậy chúng ta thu được

$$E(k) = 16{,}75 \tag{12.5.3}$$
$$\sigma_k^2 = 7{,}49395$$
$$\sigma_k = 2{,}7375$$

Vì khoảng tin cậy 95%[18] là

$$[16{,}75 \pm 1{,}96\ (2{,}7375)] = (11{,}3845;\ 12{,}1155)$$

Do số cuộc chạy là 5, nó rõ ràng nằm ngoài khoảng này. Vì vậy, chúng ta có thể bác bỏ giả thiết rằng chuỗi các phần dư quan sát được trong Hình 12.7 là ngẫu nhiên với độ tin cậy 95%.

Do số quan sát có thể nhỏ đối với kiểm định chuẩn trước đó, người đọc mong muốn chứng minh rằng trên cơ sở các cuộc chạy tới hạn, các giá trị cho trong Phụ lục D, Bảng D.6, chúng ta cũng đạt được kết luận như vậy, cụ thể là, chuỗi quan sát được không phải là ngẫu nhiên.[19]

Nếu n_1 và n_2 nhỏ hơn 20, Swed và Eisenhart đã phát triển các bảng đặc biệt cho các giá trị tới hạn của các cuộc chạy được kỳ vọng trong một chuỗi ngẫu nhiên của n quan sát. Các bảng này cho trong Phụ lục D, Bảng D.6.

---

[18] Người đọc nên kiểm tra lại các tính toán trước

[19] Sử dụng các giá trị tới hạn của các cuộc chạy cho trong bảng này, người đọc có thể chứng minh rằng đối với n_1 = 14 và n_2 = 18 các giá trị tới hạn trên và dưới của các cuộc chạy tương ứng là 23 và 10.

Kiểm định Durbin – Watson d [20]

Kiểm định nổi tiếng nhất để phát hiện tương quan chuỗi là kiểm định được phát triển bởi các nhà thống kê học Durbin và Watson. Nó được biết đến rộng rãi với tên **trị thống kê Durbin - Watson d**, được định nghĩa là

$$d = \frac{\sum_{t=2}^{t=n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=2}^{t=n} \hat{u}_t^2} \tag{12.5.4}$$

chỉ đơn giản là tỉ số giữa tổng các sai phần bình phương trong các phần dư liên tiếp và RSS. Lưu ý rằng trong tử số của trị thống kê d, số các quan sát là n –1 bởi vì 1 quan sát bị mất khi lấy các sai phần liên tiếp.

Một điểm mạnh lớn của trị thống kê d là nó dựa trên các phần dư ước lượng, chúng thường được tính toán trong phân tích hồi qui. Do điểm mạnh này, việc báo cáo Durbin - Watson d cùng với các trị thống kê tổng hợp như là R^2, R^2 đã hiệu chỉnh, các tỉ số t, v.v... bây giờ là một thực tế thông thường. Mặc dù bây giờ trị thống kê d thường được sử dụng, **vấn đề quan trọng là ghi nhớ các giả định ẩn chứa trong trị thống kê d:**

1. Mô hình hồi qui bao gồm một số hạng tung độ gốc. Nếu số hạng này không tồn tại, như trong trường hợp hồi qui qua gốc tọa độ, điều cần thiết là phải thực hiện lại hồi qui có bao gồm số hạng tung độ gốc để thu được RSS.[21]

2. Các biến giải thích X, là không ngẫu nhiên, hoặc cố định trong quá trình lấy mẫu lặp lại.

3. Các nhiễu u_t được tạo bởi sơ đồ tự hồi qui bậc 1: $u_t = \rho u_{t-1} + \varepsilon_t$

4. Mô hình hồi qui không bao gồm (các) giá trị trễ của biến phụ thuộc như là một trong các biến giải thích. Do đó, kiểm định này *không thể áp dụng* cho các mô hình có dạng sau:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \ldots + \beta_k X_{kt} + \gamma Y_{t-1} + u_t \tag{12.5.5}$$

   trong đó Y_{t-1} là giá trị trễ một thời đoạn của Y. Các mô hình như vậy được biết đến như là **các mô hình tự hồi qui**. Chúng ta sẽ xem xét chúng một cách đầy đủ trong Chương 17.

---

[20] J. Durbin và G.S. Watson, "Kiểm định tương quan chuỗi trong Hồi qui bình phương tối thiểu", Biometrika, Tập 38, 1951, trang 159-171.

[21] Tuy nhiên, R.W. Farebrother đã tính toán các giá trị d khi không có số hạng tung độ gốc trong mô hình. Xem cuốn "Kiểm định Durbin-Watson đối với tương quan chuỗi khi không có tung độ gốc trong hồi qui" của ông ta, Econometrica, Tập 48, 1980, trang 1553-1563.

5. Không có các quan sát bị thất lạc trong dữ liệu. Vì vậy, trong hồi qui tiền công-năng suất của chúng ta cho giai đoạn 1960-1991 nếu các quan sát của, chẳng hạn, năm 1963 và 1972 bị thất lạc vì lý do nào đó, trị thống kê d không chiếu cố các quan sát bị thất lạc như vậy.

Việc lấy mẫu chính xác hoặc phân phối xác suất của trị thống kê d được cho trong (12.5.4) là khó tìm được vì, như Durbin và Watson đã trình bày, nó phụ thuộc vào một con đường phức tạp dựa trên các giá trị của X có trong mẫu đã cho[22]. Điều khó khăn này có thể hiểu được vì d được tính từ u_t, mà u_t, tất nhiên, lại phụ thuộc vào các X đã cho. Vì vậy, khác với các kiểm định t, F, hoặc χ^2, không tồn tại giá trị tới hạn duy nhất dẫn tới việc bác bỏ hoặc chấp nhận *giả thiết không* cho rằng không có quan hệ chuỗi bậc nhất trong các nhiễu u_i. Tuy nhiên, Durbin và Watson đã thành công trong việc tính ra cận dưới d_L và cận trên d_U, để nếu d được tính từ (12.5.4) nằm ngoài các giá trị tới hạn này, ta có thể quyết định về việc có tồn tại tương quan chuỗi đồng biến hay nghịch biến. Hơn nữa, các giới hạn này phụ thuộc vào số quan sát n và số biến giải thích và không phụ thuộc vào các giá trị mà các biến giải thích này đã nhận. Các giới hạn này, đối với n có giá trị từ 6 tới 200 và tính tới 20 biến giải thích, đã được Durbin và Watson lập thành bảng trong Phụ lục D, Bảng D.5 (tính tới 20 biến giải thích).

Qui trình kiểm định thực sự có thể được giải thích tốt hơn với sự trợ giúp của Hình 12.9, nó cho thấy các giới hạn của d là 0 và 4. Chúng có thể được xác lập như sau.

Khai triển 12.5.4, ta được:

$$d = \frac{\sum \hat{u}_t^2 + \sum \hat{u}_{t-1}^2 - 2\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2} \tag{12.5.6}$$

Do $\sum u^2_t$ và $\sum u^2_{t-1}$ khác nhau chỉ ở nhất quan sát, chúng xấp xỉ bằng nhau. Vì vậy, đặt $\sum u^2_{t-1} = \sum u^2_t$ (12.5.6) có thể được viết là:

$$d = 2\left(1 - \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}\right) \tag{12.5.7}$$

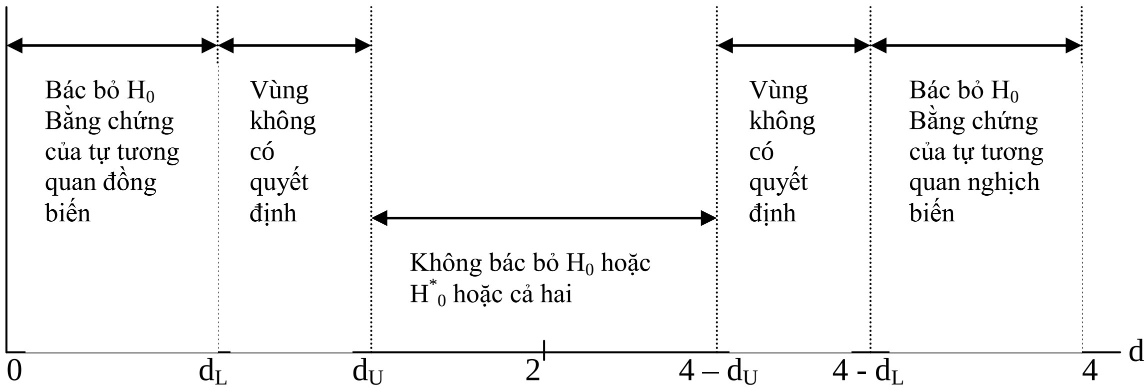
trong đó = có nghĩa là xấp xỉ.

Bây giờ ta hãy xác định:

$$\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2} \tag{12.5.8}$$

như là một hệ số tự hồi qui mẫu bậc 1, một hàm ước lượng của ρ. (Xem chú thích 6). Sử dụng (12.5.8), chúng ta có thể biểu diễn (12.5.7) như sau:

[22] Xem thảo luận về kiểm định Durbin-Watson "chính xác" trình bày trong phần sau của phần này.



Chú thích
H_0 : Không có tự tương quan đồng biến
H^*_0: Không có tự tương quan nghịch biến

HÌNH 12.9
Trị thống kê Durbin–Watson d

$$d = 2(1-\rho) \tag{12.5.9}$$

Nhưng do $-1 \leq \rho \leq 1$, (12.5.9) có nghĩa là

$$0 \leq d \leq 4 \tag{12.5.10}$$

Đây là các cận của d; bất kỳ giá trị ước lượng nào của d cũng phải nằm trong các giới hạn này.

Điều rõ ràng là từ Phương trình (12.5.9) nếu $\rho = 0$, $d = 2$; có nghĩa là, nếu có tồn tại tương quan chuỗi (bậc nhất), giá trị d được kỳ vọng sẽ gần bằng 2. Vì vậy, như là một qui tắc kinh nghiệm, nếu d tìm được có giá trị là 2 trong một ứng dụng, người ta có thể giả định rằng không tồn tại tự tương quan bậc nhất, bất kể là đồng hay nghịch biến. Nếu $\rho = +1$, cho thấy có tương quan đồng biến hoàn hảo trong các phần dư, $d = 0$. Vì vậy, d càng gần 0, bằng chứng của tương quan chuỗi đồng biến càng lớn. Mối quan hệ này cần được minh chứng từ (12.5.4) vì nếu không có tự tương quan đồng biến, các u_t sẽ cụm lại với nhau và các sai phân của chúng vì vậy có xu hướng sẽ nhỏ. Kết quả là, tổng của các bình phương ở tử số sẽ nhỏ hơn so với tổng của các bình phương ở mẫu số, chúng vẫn là một giá trị độc nhất đối với bất cứ một phép hồi qui đã cho nào.

Nếu $\rho = -1$, có nghĩa là, có tồn tại mối tương quan nghịch biến hoàn hảo giữa các phần dư liên tiếp, $d = 4$. Do đó, d càng gần 4, bằng chứng về quan hệ tương quan chuỗi nghịch biến càng lớn. Một lần nữa, khi xem xét (12.5.4), điều này có thể hiểu được. Vì nếu có tự tương quan nghịch biến, một u_t dương sẽ có xu hướng được nối tiếp bằng 1 u_t âm, và ngược lại, sao cho $|u_t - u_{t-1}|$ sẽ thường lớn hơn $|u_t|$. Vì vậy, tử số của d sẽ tương đối lớn hơn mẫu số.

Các cơ chế của kiểm định Durbin-Watson là như sau, giả định rằng các giả định ẩn chứa trong kiểm định được tuân thủ:

1. Thực hiện hồi qui OLS và thu được các phần dư
2. Tính d từ (12.5.4). (Bây giờ, hầu hết các chương trình trong máy điện toán thường đều có tính giá trị này).
3. Với cỡ mẫu cho trước và số các biến giải thích cho trước, tìm ra các giá trị tới hạn d_L và d_U
4. Bây giờ tuân theo các nguyên tắc quyết định cho trong Bảng 12.5. Để dễ tra cứu, các nguyên tắc quyết định này cũng được mô tả trong Hình 12.9.

Để minh họa các cơ chế, chúng ta hãy quay trở lại phép hồi qui tiền công-năng suất của chúng ta. Từ dữ liệu cho trong Bảng 12.4, giá trị ước lượng của d có thể được xác định là 0,1380, điều này kiến nghị rằng có tương quan chuỗi đồng biến trong các phần dư (Vì sao?) Từ các bảng Durbin-Watson chúng ta tìm thấy rằng đối với 32 quan sát và 1 biến giải thích (không kể tung độ gốc), d_L = 1,37 và d_U = 1,50 ở mức 5%. Do giá trị ước 5%. Do giá trị ước lượng 0,1380 nằm thấp hơn 1,37, chúng ta không thể bác bỏ giả thiết rằng không tồn tại tương quan chuỗi trong các phần dư.

BẢNG 12.5
Kiểm định Durbin-Watson d : các qui tắc kinh nghiệm.

Giả thiết không	**Quyết định**	**Nếu**
Không có tự tương quan đồng biến	Bác bỏ	$0 < d < d_L$
Không có tự tương quan đồng biến	Không quyết định	$d_L \le d \le d_U$
Không có tương quan nghịch biến	Bác bỏ	$4-d_L < d < 4$
Không có tương quan nghịch biến	Không quyết định	$4-d_U \le d \le 4-d_L$
Không có tự tương quan, đồng biến hoặc nghịch biến	Không bác bỏ	$d_U < d < 4-d_U$

Mặc dù cực kỳ thông dụng, kiểm định d có một mặt trái lớn là nó rơi vào *vùng không quyết định*, hoặc *khu vực bỏ qua*, người ta không thể kết luận được có phải là có tồn tại tự tương quan hay không. Để giải quyết vấn đề này, nhiều tác giả đã đưa ra cải biến cho kiểm định Durbin-Watson d nhưng chúng chỉ hơi có liên quan và nằm ngoài phạm vi của bài viết này[23].

Chương trình máy điện toán SHAZAM thực hiện một **kiểm định d chính xác** (nó cho giá trị p, xác suất chính xác, của giá trị d đã tính toán) và người sử dụng chương trình này có thể muốn sử dụng kiểm định đó trong trường hợp trị thống kê d thông thường nằm trong vùng không

[23] Về chi tiết hơn, xin đọc Thomas B. Fomby, R. Carter Hill, và Stanley R. Johnson, Các phương pháp kinh tế lượng tiên tiến, Springer - Verlag, New York, 1984, trang 225-228.

quyết định. Tuy nhiên, trong nhiều tình huống, người ta đã tìm ra rằng cận trên d_U là xấp xỉ giới hạn ở mức ý nghĩa thực[24], và vì vậy trong trường hợp giá trị ước lượng của d nằm trong vùng không quyết định, người ta có thể sử dụng qui **trình kiểm định d cải biến** như sau. Cho trước mức ý nghĩa α,

1. H_o: ρ = 0 so với H_1: ρ > 0: Nếu d ước lượng < d_U, bác bỏ H_o ở mức α, tức là có tương quan đồng biến đáng kể về mặt thống kê.
2. H_o: ρ = 0 so với H_1: ρ < 0: Nếu (4-d) ước lượng < d_U, bác bỏ H_o ở mức α; về mặt thống kê có tồn tại bằng chứng đáng kể của tự tương quan nghịch biến.
3. H_o: ρ = 0 so với H_1: ρ ≠ 0. Nếu d ước lượng < d_U hoặc (4-d) ước lượng < d_U, bác bỏ H_o ở mức 2 α; về mặt thống kê có tồn tại bằng chứng đáng kể của tự tương quan, đồng biến hoặc nghịch biến.

   Ví dụ: Giả sử trong hồi qui bao gồm 50 quan sát và 4 biến giải thích, d ước lượng là 1,43. Từ các bảng Durbin-Watson chúng ta tìm ra rằng với mức 5%, các giá trị tới hạn của d là d_L = 1,38 và d_U = 1,72. Trên cơ sở của kiểm định d thông thường chúng ta không thể nói liệu có tồn tại tương quan đồng biến hay không vì giá trị ước lượng d nằm trong khoảng không quyết định. Nhưng trên cơ sở của kiểm định d cải biến chúng ta có thể bác bỏ giả thiết của việc không tồn tại tương quan đồng biến (bậc nhất) bởi vì d < d_U.[25]

Nếu người ta không muốn sử dụng kiểm định d cải biến, họ có thể rơi ngược vào kiểm định các cuộc chạy phi thông số đã được thảo luận trước đây.

Trong khi sử dụng kiểm định Durbin-Watson, điều cốt yếu là cần ghi nhớ rằng nó không thể áp dụng khi vi phạm các giả định của nó. Đặc biệt là nó không thể được dùng để kiểm định đối với tương quan chuỗi trong các mô hình tự hồi qui, tức là, các mô hình có chứa (các) giá trị trễ của biến phụ thuộc được xem như (các) biến giải thích. Nếu áp dụng sai, giá trị của d trong các trường hợp này sẽ thường có giá trị xung quanh 2, đó là giá trị kỳ vọng của d khi không tồn tại tự tương quan bậc 1 [Xem (12.5.9)]. Do đó, sẽ phát sinh các thiên lệch gây khó khăn cho việc tìm ra tương quan chuỗi trong các mô hình như vậy. Kết quả này không có nghĩa là các mô hình tự hồi qui không chịu hậu quả từ vấn đề tự tương quan. Như chúng ta sẽ thấy ở cuối chương, Durbin đã phát triển cái gọi là **trị thống kê h** để kiểm định tương quan chuỗi trong các mô hình như vậy.

[24] Ví dụ, Theil và Nagar đã chỉ ra rằng cận trên d_U "là gần bằng giới hạn ở mức ý nghĩa thực trong tất cả mọi trường hợp mà trong đó hành vi của các biến giải thích là trơn theo nghĩa rằng các sai phân bậc 1 và bậc 2 của chúng là nhỏ so với phạm vi của biến tương ứng". Xem Henri Theil, Các Nguyên lý Kinh tế lượng, John Wiley & Sons, New York, 1971, trang 201. Đồng thời xem E.J. Hannon và R.D. Terrell, "Kiểm định tương quan chuỗi sau khi hồi qui bình phương tối thiểu" Econometrika, Tập 34, 1961, trang 646-660.

[25] Trong lời khuyên nào đó trên thực tế về việc sử dụng trị thống kê Durbin-Watson thế nào, xem Draper và Smith op-Cit, trang 162-169. Đồng thời xem G.S. Maddala, op-cit, Chương 6, về một vài sử dụng và lạm dụng trị thống kê Durbin-Watson.

Các kiểm định tự tương quan bổ sung

Kiểm định tiệm cận, hoặc mẫu lớn. Theo *giả thiết không* cho rằng ρ = 0 và giả định rằng cỡ mẫu n là lớn (nói theo thuật ngữ kỹ thuật là vô định), có thể chỉ ra rằng $\sqrt{n}.\rho$ tuân theo phân phối chuẩn với giá trị trung bình = 0 và phương sai = 1. Có nghĩa là, một cách tiệm cận

$$\sqrt{n}\cdot\hat{\rho} \sim N(0,1) \qquad (12.5.11)$$[26]

Như là một minh họa cho kiểm định này, đối với ví dụ tiền công–năng suất của chúng ta, ước lượng của ρ có thể được tìm ra là 0,8844. Cho trước cỡ của mẫu là 32, chúng ta tìm $\sqrt{32}\cdot(0{,}8844) = 5{,}003$. Một cách tiệm cận, nếu *giả thiết không* cho rằng ρ = 0 là đúng, xác suất có một giá trị khoảng 5,00 hoặc lớn hơn là cực nhỏ. Hãy nhớ lại rằng đối với phân bố xác suất chuẩn chuẩn hóa, Z tới hạn ở mức ý nghĩa 5% (2 đầu) (tức là, biến chuẩn chuẩn hóa) là 1,96 và giá trị của Z tới hạn ở mức ý nghĩa 1% là khoảng 2,58. Do đó, chúng ta bác bỏ H_o cho rằng ρ = 0.

Kiểm định Breusch–Godfrey (BG) của tự tương quan bậc cao hơn. Giả sử rằng số hạng nhiễu u_t được tạo bởi sơ đồ tự hồi qui bậc p như sau:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_p u_{t-p} + \varepsilon_t \qquad (12.5.12)$$

trong đó ε_t là nhiễu ngẫu nhiên thuần túy với giá trị trung bình bằng 0 và phương sai không đổi.

Giả thiết không H_o của chúng ta cho rằng $\rho_1 = \rho_2 = \ldots = \rho_p = 0$, nghĩa là tất cả các hệ số tự hồi qui đồng thời bằng 0, tức là không có tự tương quan của bất cứ bậc nào. Breusch và Godfrey đã chỉ ra rằng *giả thiết không* có thể được kiểm định như sau:[27]

1. Ước lượng mô hình hồi qui bằng qui trình OLS thông thường thu được các phần dư u_t.

2. Hồi qui u_t được xác định theo tất cả các biến độc lập trong mô hình hồi qui cộng với các biến độc lập bổ sung này, u_{t-1} u_{t-2}, …, u_{t-p}, trong đó các biến sau là các giá trị trễ của các phần dư đã ước lượng trong Bước 1. Do đó, nếu p = 4,chúng ta sẽ đưa vào 4 giá trị trễ của các phần dư như là các biến hồi qui bổ sung trong mô hình. Lưu ý rằng để thực hiện hồi qui này, chúng ta sẽ có chỉ (n-p) quan sát (vì sao?). Thu được giá trị R^2 từ phép hồi qui này, một phép hồi qui phụ.

3. Nếu cỡ củ mẫu lớn, Breusch và Godfrey đã chỉ ra rằng

$$(n-p).\ R^2 \sim X^2_p \qquad (12.5.13)$$

Tức là, một cách tiệm cận, (n-p) nhân với R^2 vừa thu được ở Bước 2 tuân theo kiểm định Chi-bình phương với bậc tự do là p. Nếu trong một ứng dụng (n-p). R^2 vượt quá giá trị Chi-bình phương tới hạn ở mức ý nghĩa đã chọn, chúng ta có thể bác bỏ *giả thiết không*, trong trường hợp này ít nhất có 1 ρ là khác 0 một cách đáng kể.

Các *điểm thực hành* sau đây về kiểm định BG có thể được ghi nhận:

[26] Xem George G. Judge, R. Carter Hill, Nillian E. Griffith, Helmut Luthepohl, và Tsoung – Chao Lee, Nhập môn Lý thuyết và Thực hành của kinh tế lượng, in lần thứ 2, John Wiley & Sons, New York, 1988, trang 394.
[27] L.G. Godfrey "Kiểm định dựa theo các mô hình tự hồi qui tổng quát và sai số trung bình trượt khi các biến độc lập bao gồm các biến phụ thuộc trễ", Econometrica, tập 46, 1978, trang 1293-1302; và T.S.Breusch, "Kiểm định tự tương quan trong các mô hình tuyến tính động", Bài viết Kinh tế Úc, tập 17,1978, trang 334-355.

1. Các biến độc lập có trong mô hình hồi qui có thể có chứa các giá trị trễ của biến hồi qui phụ thuộc Y, tức là, Y_{t-1}, Y_{t-2} v.v… có thể xuất hiện như các biến giải thích. Đối chiếu mô hình này với hạn chế trong kiểm định Durbin Watson cho rằng không có các giá trị trễ của biến phụ thuộc giữa các biến giải thích.

2. Kiểm định BG có thể áp dụng thậm chí nếu số hạng nhiễu tuân theo **quá trình MA** bậc p, tức là u_t được tạo bởi:

$$u_t = \varepsilon_t + \lambda_1 \varepsilon_{t-1} + \lambda_2 \varepsilon_{t-2} + \ldots + \lambda_p \varepsilon_{t-p} \tag{12.5.14}$$

   trong đó ε là số hạng nhiễu ngẫu nhiên với giá trị trung bình bằng 0 và phương sai không đổi.

3. Nếu trong (12.5.12) p = 1, có nghĩa là tự hồi qui bậc 1, thì kiểm định BG được gọi là **kiểm định Durbin m.**

4. Điểm yếu của kiểm định BG là giá trị của p, chiều dài của độ trễ, không thể được xác định một tiên nghiệm. Một vài thử nghiệm nào đó với giá trị của p là không thể tránh được. Chúng ta sẽ trở lại chủ đề này khi chúng ta thảo luận về kinh tế lượng chuỗi thời gian sau này.

   Ví dụ minh họa. Trở lại hồi qui tiền công–năng suất đã xem xét trước đây, chúng ta đã tuân theo qui trình BG, đưa vào 5 giá trị trễ của các phần dư OLS trong hồi qui phụ (tức là, hồi qui của tiền công theo năng suất và 5 giá trị trễ của các phần dư thu được từ hồi qui chỉ riêng của tiền công theo năng suất). Giá trị R^2 từ hồi qui (phụ) này là 0,8660. Trong toàn bộ, có 32 quan sát trong hồi qui ban đầu, nhưng do 5 độ trễ được sử dụng, chúng ta chỉ có 27 quan sát đối với hồi qui phụ. Vì vậy, (27)(0,8660) = 23,382, giá trị của p, hoặc xác suất chính xác, của việc thu được giá trị Chi-bình phương như vậy là khoảng 0,0003, nó hoàn toàn thấp. Vậy chúng ta có thể bác bỏ giả thiết cho rằng cả 5 hệ số trễ của u bằng 0. Ít nhất là một hệ số trễ cần phải khác không. Thực tế này không làm ta ngạc nhiên khi xem xét phát hiện trước đây của chúng ta rằng có tự tương quan AR (1) trong các phần dư.

12.6 CÁC BIỆN PHÁP SỬA CHỮA

Do khi có tương quan chuỗi các hàm ước lượng OLS sẽ không hiệu quả, điều cốt yếu là phải tìm các biện pháp sửa chữa. Tuy nhiên, biện pháp sửa chữa phụ thuộc vào kiến thức mà người ta có được về bản chất của mối liên phụ thuộc giữa các nhiễu. Chúng ta phân biệt 2 tình huống: khi đã biết cấu trúc của tự tương quan và khi không biết.

Khi đã biết Cấu trúc của Tự Tương quan

Do các nhiễu u_t là không thể quan sát được, bản chất của tương quan chuỗi thường thường là vấn đề của sự suy đoán hay các yêu cầu cấp thiết của thực tế. Trên thực tế, người ta thường giả định rằng u_t tuân theo sơ đồ tự hồi qui bậc 1, cụ thể là:

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{12.6.1}$$

trong đó $|\rho| < 1$ và ε_t tuân theo các giả định của OLS về giá trị kỳ vọng = 0, phương sai không đổi, và không tự tương quan, như trình bày trong (12.2.2).

Nếu chúng ta giả định tính hiệu lực của (12.6.1), vấn đề tương quan chuỗi có thể được giải quyết một cách thỏa đáng nếu đã biết hệ số tự tương quan ρ. Để thấy điều này, chúng ta hãy quay lại mô hình 2 biến.[28]

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad (12.6.2)$$

Nếu (12.6.2) đúng tại thời gian t, nó cũng sẽ đúng tại thời gian t – 1. Vì,

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \qquad (12.6.3)$$

Nhân cả hai vế của (12.6.3) với ρ, ta có

$$\rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_{t-1} + \rho u_{t-1} \qquad (12.6.4)$$

Lấy (12.6.2) trừ (12.6.4) ta được

$$\begin{aligned}(Y_t - \rho Y_{t-1}) &= \beta_1(1-\rho) + \beta_2 X_t - \rho\beta_2 X_{t-1} + (u_t - \rho u_{t-1}) \qquad (12.6.5)\\ &= \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \varepsilon_1\end{aligned}$$

trong đó bước cuối cùng là do (12.6.1)

Chúng ta có thể biểu thị (12.6.5) như sau

$$Y^*_t = \beta^*_1 + \beta^*_2 X^*_t + \varepsilon_t \qquad (12.6.6)$$

trong đó $\beta^*_1 = \beta_1(1-\rho)$, $Y^*_t = (Y_t - \rho Y_{t-1})$ và $X^*_t = (X_t - \rho X_{t-1})$

Do ε_t thỏa mãn mọi giả định OLS, người ta có thể tiếp tục áp dụng OLS cho các biến đã biến đổi Y^* và X^* và thu được các hàm ước lượng với mọi tính chất tối ưu, cụ thể là ước lượng không thiên lệch tuyến tính tốt nhất BLUE. Hiệu quả là, việc thực hiện (12.6.6) là tương đương với việc sử dụng các bình phương tối thiểu tổng quát (GLS) đã được thảo luận trong Phần 12.3 (Xem bài tập 12.19). Nhưng nên lưu ý rằng quan sát thứ nhất (Y_1, X_1) đã bị loại ra ngoài. (Vì sao?)

Hồi qui (12.6.5) được biết như là **phương trình sai phân hầu như tổng quát hóa**. Nó liên quan đến việc thực hiện hồi qui Y theo X, không dưới dạng ban đầu, nhưng dưới dạng sai phân, nó thu được bằng cách lấy giá trị của nó trong thời đoạn hiện tại trừ ra tỉ phần (= ρ) giá trị của một biến trong thời đoạn trước đó. Trong qui trình lấy sai phân này, chúng ta mất đi một quan sát vì quan sát đầu tiên không có quan sát trước nó. Để tránh được mất mát một quan sát

[28] Việc mô hình có nhiều hơn một biến giải thích hay không sẽ không thành vấn đề vì tự tương quan là một đặc tính của các u_t.

này, quan sát đầu của Y và X được biến đổi như sau: [29] $Y_1\sqrt{1-\rho^2}$ và $X_1\sqrt{1-\rho^2}$. Sự biến đổi này gọi là **biến đổi Prais – Winsten.**

Khi Không Biết ρ

Mặc dù là đơn giản khi áp dụng, hồi qui sai phân tổng quát hóa nói chung là khó thực hiện vì ρ rất ít khi biết được trên thực tế. Vì vậy, cần phải nghĩ ra các phương pháp thay thế. Một vài phương pháp này như sau:

Phương pháp sai phân thứ nhất. Do ρ nằm giữa 0 và ± 1, người ta có thể bắt đầu từ 2 vị trí thái cực. Tại một thái cực, chúng ta có thể giả định rằng ρ = 0, tức là, không có tương quan chuỗi, và tại thái cực khác chúng ta có thể cho ρ = ± 1, tức là, tự tương quan đồng biến hoặc nghịch biến hoàn hảo. Trên thực tế, khi một hồi qui được thực hiện, người ta giả định tổng quát rằng không có tự tương quan và sau đó để cho kiểm định Durbin-Watson hoặc các kiểm định khác chứng tỏ liệu có phải giả định này là xác đáng. Tuy nhiên, nếu ρ = ± 1, phương trình sai phân tổng quát hóa (12.6.5) giảm xuống thành phương trình sai phân thứ nhất như sau:

$$\begin{aligned} Y_t - Y_{t-1} &= \beta_2 (X_t - X_{t-1}) + (u_t - u_{t-1}) \\ &= \beta_2 (X_t - X_{t-1}) + \varepsilon_t \end{aligned}$$

hoặc

$$\Delta Y_t = \beta_2 \Delta X_t + \varepsilon_t \tag{12.6.7}$$

trong đó Δ, gọi là delta, là toán tử sai phân thứ nhất và là một ký hiệu hoặc toán tử (giống như toán tử giá trị kỳ vọng E) đối với các sai phân liên tiếp của hai giá trị. (Lưu ý: Nói chung một toán tử là một ký hiệu để biểu thị một phép toán). Trong khi thực hiện (12.6.7) tất cả những gì ta phải làm là tạo ra các sai phân thứ nhất của cả biến phụ thuộc và biến giải thích và sử dụng chúng như là các nhập lượng trong phân tích hồi qui.

Cần lưu ý một đặc điểm quan trọng của mô hình sai phân thứ nhất: Không có số hạng tung độ gốc. Do đó, để thực hiện (12.6.7), phải dùng mô hình hồi qui qua gốc tọa độ. Nhưng giả sử rằng mô hình ban đầu là:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 t + u_t \tag{12.6.8}$$

trong đó t là biến xu hướng và u_t tuân theo sơ đồ tự hồi qui bậc 1. Người đọc có thể chứng minh rằng việc biến đổi sai phân thứ nhất của (12.6.8) là như sau:

[29] Việc mất một quan sát có thể không phải là rất nghiêm trọng trong một mẫu lớn, nhưng có thể thay đổi kết quả đáng kể trong 1 mẫu nhỏ. Về điều này, xen xem J. Johnston, op.cit., Chương 8, trang 321-323, và cũng xem Phần 12.7. Xem Davidson và MacKinnon, op-cit, Bảng 10.1, trang 349 về một số kết quả Monte Carlo về tầm quan trọng của quan sát đầu tiên.

$$\Delta Y_t = \beta_2\, \Delta X_t + \beta_3 + \varepsilon_t \tag{12.6.9}$$

trong đó $\Delta Y_t = Y_t - Y_{t-1}$ và $\Delta X_t = X_t - X_{t-1}$. Phương trình (12.6.9) cho thấy rằng có số hạng tung độ gốc dưới dạng sai phân thứ nhất, nó ngược với 12.6.7). Nhưng tất nhiên, β_3 là một hệ số của biến xu hướng trong mô hình ban đầu. *Do đó, sự có mặt của số hạng tung độ gốc dưới dạng sai phân thứ nhất chứng thực rằng có một số hạng xu hướng tuyến tính trong mô hình ban đầu và số hạng tung độ gốc thực tế là hệ số của biến xu hướng*. Ví dụ, nếu β_3 có giá trị dương trong (12.6.9), thì nó có nghĩa là Y có xu hướng đi lên sau khi xem xét tác động của tất cả các biến khác.

Thay vì giả định $\rho = +1$, nếu chúng ta giả định rằng $\rho = -1$, tức là, tương quan chuỗi nghịch biến hoàn hảo (đó không phải là điển hình cho chuỗi thời gian kinh tế), phương trình sai phân tổng quát (12.6.5) bây giờ trở thành

$$Y_t + Y_{t-1} = 2\beta_1 + \beta_2 (X_t + X_{t-1}) + \varepsilon_t$$

hoặc

$$\frac{Y_t + Y_{t-1}}{2} = \beta_1 + \beta_2 \frac{X_t + X_{t-1}}{2} + \frac{\varepsilon_t}{2} \tag{12.6.10}$$

Mô hình trước đó đã biết đến như là mô hình **hồi qui trung bình trượt** (2 thời đoạn) bởi vì chúng ta đang hồi qui giá trị của một trung bình trượt theo một giá trị trung bình khác.[30]

Sự biến đổi sai phân thứ nhất được trình bày trên đây là hoàn toàn thông dụng trong kinh tế lượng ứng dụng do nó dễ thực hiện. Nhưng nên lưu ý rằng sự biến đổi này dựa trên giả định rằng $\rho = +1$; tức là các phần nhiễu có tương quan đồng biến hoàn hảo. Nếu không đúng thì sự cứu chữa còn tồi tệ hơn cả căn bệnh. Nhưng làm thế nào người ta tìm ra xem có phải giả định $\rho = +1$ là thích đáng trong một tình huống đã hay không? Điều này có thể kiểm định bằng **kiểm định Berenblutt – Webb.**

Kiểm định Berenblutt-Webb về giả định cho rằng $\rho = 1$. Để kiểm định giả thiết rằng $\rho = 1$ (tức là, tương quan chuỗi đồng biến hoàn hảo bậc 1). Berenblutt và Webb đã phát triển **trị thống kê (kiểm định) g** như sau.[31]

$$g = \frac{\sum_{t=2}^{n} \hat{e}_t^2}{\sum_{t=1}^{n} \hat{u}_t^2} \tag{12.6.11}$$

[30] Do $(Y_t + Y_{t-1})/2$ và $X_t + X_{t-1})/2$ là các trung bình của hai giá trị sát nhau, chúng được gọi là các trung bình 2 thời đoạn. Chúng là trượt vì khi tính toán các trung bình này trong các giai đoạn liên tiếp chúng ta bỏ đi một quan sát và thêm vào một quan sát khác. Do đó $(Y_{t-1} + Y_t)/2$ sẽ là trung bình 2 giai đoạn tiếp theo, v.v…

[31] I.I.Berenblutt và G.I. Webb, "Kiểm định mới đối với các sai số tự tương quan trong mô hình hồi qui tuyết tính", Journal of the Royal Statistical Society, Loạt bài B, Tập 35, No 1, 1973, trang 33-50.

trong đó u_t là các phần dư OLS từ mô hình ban đầu và $\hat{e}_t$ là các phần dư OLS từ hồi qui dựa trên sai phân bậc 1 của Y, ΔY (tức là, $Y_t - Y_{t-1}$) dựa trên sai phân thứ nhất của các biến hồi qui, ΔX (tức là, $[X_{2\,t} - X_{2\,(t-1)}]$, $[X_{3t} - X_{3\,(t-1)}]$, v.v…). Nhưng nên lưu ý là dưới dạng sai phân thứ nhất, không có tung độ gốc (Vì sao?)

Nếu mô hình ban đầu bao gồm một số hạng là hằng số, chúng ta có thể sử dụng các bảng của Durbin-Watson để kiểm định trị thống kê g, chỉ khác là *giả thiết không* bây giờ là $\rho = 1$ chứ không phải là giả thiết của Durbin – Watson $\rho = 0$.

Để minh họa kiểm định Berenblutt- Webb, chúng ta hãy trở lại ví dụ tiền công-năng suất và giả định rằng $H_0 : \rho = 1$. Hồi qui Y (tiền công) theo X (năng suất), chúng ta thu được RSS = 204,6934. Và hồi qui ΔY trên ΔX (Lưu ý: không có tung độc gốc trong hồi qui này), chúng ta được RSS = 28,1938. Vì vậy

$$g = \frac{28{,}1938}{294{,}6934} = 0{,}1377$$

Kiểm tra bảng Durbin-Watson đối với 31 quan sát và 1 biến giải thích, chúng ta tìm ra rằng $d_L = 1{,}363$ và $d_U = 1{,}237$ (mức ý nghĩa là 5%) và $d_L = 1{,}147$ và $d_U = 1{,}273$ (mức ý nghĩa là 1%). Do giá trị của g quan sát được nằm dưới giới hạn thấp hơn, chúng ta không bác bỏ *giả thiết không* rằng ρ thực = 1. Hãy nhớ trong đầu rằng mặc dù chúng ta sử dụng cùng các bảng Durbin – Watson, bây giờ *giả thiết không* là $\rho = 1$ và không phải $\rho = 0$. Theo phát hiện này, sự biến đổi sai phân thứ nhất đã thảo luận trước đây, dưới giả định cho rằng $\rho = 1$, có thể là phù hợp.

ρ dựa trên trị thống kê Durbin-Watson d. Hãy nhớ lại rằng trước đây chúng ta đã thiết lập mối quan hệ sau:

$$d = 2\,(1 - \rho) \qquad (12.5.9)$$

hoặc

$$\hat{\rho} = 1 - \frac{d}{2} \qquad (12.6.12)$$

nó cho ta một con đường đơn giản để tìm ước lượng của ρ từ trị thống kê d đã ước lượng. Rõ ràng là từ (12.6.12) giả định cho rằng $\rho = +1$ là hiệu lực chỉ khi $d = 0$ hoặc xấp xỉ như vậy. Cũng rõ ràng rằng khi $d = 2$, $\rho = 0$ và khi $d = 1$, $\rho = -1$. Vì vậy, trị thống kê d cho ta một phương pháp sẵn sàng để tìm một ước lượng của ρ. Nhưng cũng lưu ý rằng mối liên hệ (12.6.12) chỉ là một quan hệ xấp xỉ và có thể không đúng đối với các mẫu nhỏ. Đối với các mẫu nhỏ, ta có thể sử dụng **trị thống kê d cải biến Theil – Nagar**.[32]

Đối với ví dụ tiền công–năng suất của chúng ta, d = 0,1380. Do đó, ρ = 1-(0,1382)/2 = 0,931.

Khi ρ được ước lượng từ (12.6.12), ta có thể biến đổi dữ liệu như đã trình bày trong (12.6.6) và tiến hành phép ước lượng OLS thông thường. Chúng ta sẽ minh họa kỹ thuật này ngay sau đây. Nhưng trước đó, chúng ta đưa ra một câu hỏi quan trọng: các hệ số hồi qui ước lượng sẽ có các tính chất tối ưu thông thường của mô hình cổ điển không? Nên lưu ý rằng trong phương trình sai phân tổng quát, ρ chứ không phải là ρ xuất hiện, nhưng khi thực hiện hồi qui OLS chúng ta sử dụng ρ. Bỏ qua việc thâm nhập vào các kỹ thuật phức tạp, ta có thể khẳng định *như một nguyên tắc tổng quát, bất cứ lúc nào chúng ta sử dụng một hàm ước lượng thay cho giá trị thực, các hệ số OLS ước lượng có các tính chất tối ưu thông thường chỉ là một cách tiếp cận, tức là, trong các mẫu lớn. Đồng thời, các qui trình kiểm định giả thiết thông dụng, nói một cách chặt chẽ, cũng chỉ hiệu lực một cách tiệm cận. Vì vậy, trong các mẫu nhỏ ta cần phải thận trọng khi giải thích các kết quả ước lượng.*

Quy trình lặp Cochrane-Orcutt để ước lượng ρ [33]. Một cách khác để ước lượng ρ từ trị thống kê Durbin-Watson d là phương pháp Cochrane-Orcutt được sử dụng thường xuyên, nó sử dụng các phần dư ước lượng u_t để thu thông tin về ρ chưa biết.

Để giải thích phương pháp này, hãy xem xét mô hình hai biến:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \tag{12.6.13}$$

và giả định rằng u_t được tạo bởi sơ đồ AR (1), cụ thể là,

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{12.2.1}$$

Sau đó Cochrane và Orcutt giới thiệu các bước sau đây để ước lượng ρ:

1. Ước lượng mô hình hai biến bằng trình tự OLS chuẩn và thu được các phần dư, u_t

2. Sử dụng các phần dư ước lượng, thực hiện hồi qui sau:

$$u_t = \rho\, u_{t-1} + \nu_t \tag{12.6.14}$$

nó là phần tương ứng theo kinh nghiệm của sơ đồ AR (1) đã cho trước đây.[34]

[32] Sự cải biến này được cho trong bài tập 12.6. Xem bài "Kiểm định sự phụ thuộc của các nhiễu hồi qui", Journal of the American Statistical Association, tập 56, 1961, trang 793-806.

[33] D. Cochrane và G.H. Arcutt, "Ứng dụng của các phép hồi qui bình phương tối thiểu cho các mối liên hệ bao gồm các số hạng sai số tự tương quan", Journal of the American Statistical Association, Tập 44, 1949, trang 32-61.

3. Sử dụng ρ đã thu được từ (12.6.14), thực hiện phương trình sai phân tổng quát hóa (12.6.5), cụ thể là

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

hoặc

$$Y^*_t = \beta^*_1 + \beta^*_2\ X^*_t + e^*_t \tag{12.6.15}$$

(Lưu ý: Chúng ta có thể thực hiện hồi qui này do đã biết ρ. Cũng lưu ý rằng $\beta^*_1 = \beta_1 (1-\rho)$.

4. Bởi vì một tiên nghiệm cho rằng ta không biết ρ thu được từ (12.6.14) là ước lượng "tốt nhất" của ρ, thay giá trị của $\beta^*_1 = \beta_1 (1-\rho)$ và β^*_2 thu được từ (12.6.15) vào hồi qui *ban đầu* (12.6.13) và thu được các phần dư mới, chẳng hạn u_t^{**}, là:

$$u_t^{**} = Y_t = \beta^*_1 = \beta^t_2 X_t \tag{12.6.16}$$

chúng có thể tính toán dễ dàng vì đã biết tất cả Y_t, X_t, β^*_1 và β^*_2.

5. Bây giờ ước lượng hồi qui này:

$$u_t^{**} = \rho\, u^{**}_{t-1} + w_t \tag{12.6.17}$$

nó tương tự với (12.6.14). Vì vậy, ρ là ước lượng vòng hai của ρ.

Bởi vì ta không biết liệu có phải ước lượng vòng hai này của ρ là ước lượng tốt nhất của ρ hay không, ta có thể ước lượng vòng 3, và tiếp theo. Như các bước trước đây đã đề đạt, phương pháp Cochrane – Orcutt có tính lặp. Nhưng chúng ta cần tiếp tục lặp bao nhiêu lần? Qui trình tổng quát là ngừng thực hiện các phép lặp khi các ước lượng liên tiếp của ρ khác nhau bởi một lượng rất nhỏ, chẳng hạn, nhỏ hơn 0,01 hoặc 0,005. Như một ví dụ minh họa sẽ được trình bày sau đây, trên thực tế, rất thường gặp là 3 hoặc 4 phép lặp là đủ.

Qui trình hai bước Cochrane–Orcutt. Đây là một dạng rút gọn của quá trình lặp. Trong bước một, chúng ta ước lượng ρ từ phép lặp thứ nhất, tức là từ hồi qui (12.6.14), và trong bước hai chúng ta sử dụng ước lượng này của ρ để thực hiện phương trình sai phân tổng quát hóa. Trên thực tế, đôi khi phương pháp hai bước này cho các kết quả tương tự như các kết quả thu được từ qui trình lặp tỉ mỉ hơn đã thảo luận ở trên.

[34] Ghi chú: $\rho = \Sigma^n_{t-2}\, ut\, u_{t-1} / \Sigma^n_{t-2}\, u^2_{t-1}$ (Vì sao?) (Xem chú thích 6). Khi kết thúc, ghi chú rằng mặc dù bị thiên lệch, đây là một hàm ước lượng nhất quán của ρ, tức là, khi cỡ của mẫu tăng một cách vô định, ρ đồng qui tại ρ thực.

Đối với ví dụ tiền công–năng suất của chúng ta, ρ được ước lượng từ (12.6.14) là 0,9404. Sử dụng ước lượng này và phương trình sai phân tổng quát hóa (12.6.15), chúng ta có:

$$
\begin{aligned}
Y^*_t &= 1{,}7152 + 0{,}7152\, X^*_t \\
\text{se} &= (1{,}1069) \quad (0{,}1569) \qquad R^2 = 0{,}4174 \\
& \qquad\qquad\qquad\qquad\quad d = 1{,}5886
\end{aligned}
\tag{12.6.18}
$$

trong đó $Y^*_t = (Y_t – 0{,}9404\, Y_{t-1})$, $X^*_t = (X_t – 0{,}9404\, X_{t-1})$, và $1{,}7152 = \beta_1 (1-0{,}9404)$, β_1 có thể được ước lượng là 28,7785. So sánh các kết quả này với hồi qui ban đầu cho trong Phụ lục 12A, Phần 12A.1.

Phương pháp hai bước của Durbin để ước lượng ρ. [35] Để minh họa phương pháp này, chúng ta hãy viết phương trình sai phân tổng quát hóa (12.6.5), một cách tương đương như sau:

$$Y_t = \beta_1 (1-\rho) + \beta_2 X_t - \rho\beta_2 X_{t-1} + \rho\, Y_{t-1} + \varepsilon_t \tag{12.6.19}$$

Durbin đề xuất qui trình hai bước như sau để ước lượng ρ:

1. Xử lý (12.6.19) như là một mô hình hồi qui đa biến, hồi qui Y_t theo X_t, X_{t-1} và Y_{t-1}; Và xử lý giá trị ước lượng của hệ số hồi qui cho Y_{t-1} (= ρ) như là một ước lượng của ρ. Mặc dù bị thiên lệch, nó cho ta một ước lượng nhất quán của ρ.

2. Sau khi thu được ρ biến đổi các biến như $Y^*_t = (Y_t - \rho^y{}_{t-1})$ và $X^*_t = (X_t - \rho\, X_{t-1})$ và thực hiện hồi qui OLS theo các biến đã biến đổi như trong (12.6.6).

Từ thảo luận trên đây, rõ ràng là bước thứ nhất trong qui trình hai bước của Durbin là để đạt được một ước lượng của ρ và bước thứ hai bao gồm việc thu các ước lượng của các thông số. Sau đây chúng ta sẽ nhận xét phương pháp này thông qua các phương pháp khác.

Đối với ví dụ tiền công–năng suất, chúng ta thu được ước lượng của (12.6.19) như sau:

$$
\begin{aligned}
Y_t &= 3{,}4879 + 0{,}7335\, X_t - 0{,}7122\, X_{t-1} + 0{,}9422\, Y_{t-1} \\
\text{se} &= (2{,}0889) \quad (0{,}1578) \qquad (0{,}1681) \qquad (0{,}0699) \qquad R^2 = 0{,}9922 \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad d = 1{,}7664
\end{aligned}
\tag{12.6.20}
$$

Từ hệ số của Y_{t-1} chúng ta thấy rằng ước lượng của ρ là 0,9422, nó không khác nhiều so với kết quả thu được từ giá trị của d trong hồi qui ban đầu hoặc kết quả thu được từ qui trình hai bước Cochrane-Orcutt.

[35] Durbin, "Ước lượng các thông số trong các mô hình hồi qui chuỗi thời gian", Journal of The Royal Statistical Society, loạt B, tập 22, 1960, trang 139-153.

Các phương pháp ước lượng ρ khác. Chúng ta vừa thảo luận một vài phương pháp thường được sử dụng của việc ước lượng ρ, nhưng danh sách này không phải là toàn bộ. Ví dụ, người ta có thể sử dụng phương pháp thích hợp tối đa để ước lượng các thông số, chẳng hạn, của phương trình (12.6.19) một cách trực tiếp mà không cần phải sử dụng một số các chu trình lặp đã thảo luận trước đây. Nhưng phương pháp ML bao gồm các qui trình ước lượng phi tuyến tính (trong các thông số) và nằm ngoài phạm vi của bài viết này.[36] Sau đó, còn có phép rà Hildreth-Lu hoặc qui trình khác (Xem bài tập 12.7). Nhưng phương pháp này là hoàn toàn tốn thời gian và đã được phát hiện là phương pháp không hiệu quả về tổng thể so với ước lượng ML và vì vậy ngày nay không được sử dụng nhiều.

Chúng ta kết luận phần này với các quan sát này. Các phương pháp khác nhau vừa được thảo luận về cơ bản là các phương pháp hai bước: Trong Bước một chúng ta thu được một ước lượng của ρ chưa biết và trong Bước hai chúng ta sử dụng ước lượng này để biến đổi các biến nhằm ước lượng phương trình sai phân tổng quát, về cơ bản đó là GLS. Nhưng do chúng ta sử dụng ρ thay cho ρ thực, tất cả các phương pháp ước lượng này được đề cập đến trong lý thuyết như là **các phương pháp bình phương tối thiểu tổng quát hóa đã ước lượng hoặc khả thi**.

12.7 VÍ DỤ MINH HỌA: MỐI QUAN HỆ GIỮA CHỈ SỐ MUỐN GIÚP ĐỠ VÀ TỈ LỆ THẤT NGHIỆP, HOA KỲ: SO SÁNH CÁC PHƯƠNG PHÁP.

Như một minh họa của các phương pháp khác nhau vừa được thảo luận, xét ví dụ sau (Xem Bảng 12.7)

Mô hình hồi qui đã chọn đối với khảo sát thực nghiệm là:

$$\ln HWI_t = \beta_1 + \beta_2 \, LnU_t + u_t$$

trong đó HWI là chỉ số muốn giúp đỡ và U là tỉ lệ thất nghiệp.[37] Trước đây, β_2 được kỳ vọng là âm (Vì sao?) Giả sử rằng tất cả các giả định OLS là đúng, chúng ta có thể viết hồi qui ước lượng như sau:

$$\begin{aligned} \ln HWI_t &= 7{,}3084 - 1{,}5375 \ln U_t \qquad (12.7.1)\\ &\quad (0{,}1110) \quad (0{,}0711) \qquad N = 24\\ t &= (65{,}825) \quad (-21{,}612) \qquad r^2 = 0{,}9550\\ &\qquad d = 0{,}9108 \end{aligned}$$

Từ hồi qui đã ước lượng chúng ta thấy rằng Durbin-Watson d cho thấy sự hiện hiện của tương quan chuỗi đồng biến: Đối với 24 quan sát và 1 biến giải thích, bảng Durin-Watson 5% cho ta $d_L = 1,27$ và $d_U = 1,45$, và d ước lượng là 0,9108 nằm dưới giới hạn tới hạn thấp.

[36] Xem J. Johnston, Op. Cit, trang 325-325.

[37] Hiện tại chúng ta hãy đừng lo lắng về vấn đề tính đồng thời, tức là, liệu U tạo ra HWI hay ngược lại.

BẢNG 12 .6
Mối liên hệ giữa chỉ số muốn giúp đỡ (HWI) và tỷ lệ thất nghiệp (U)

Năm và Quí	HWI, 1957-1959 = 100	U,%
1962–1	104.66	5.63
–2	103.53	5.46
–3	97.30	5.63
–4	95.96	5.60
1963–1	98.83	5.83
–2	97.23	5.76
–3	99.06	5.56
–4	l13.66	5.63
1964–1	l17.00	5.46
–2	l19.66	5.26
–3	124.33	5.06
–4	133.00	5.06
1965–1	143.33	4.83
–2	144.66	4.73
–3	152.33	4.46
–4	178.33	4.20
1966–1	192.00	3.83
–2	186.00	3.90
–3	188.00	3.86
–4	193.33	3.70
1967–1	187.66	3.66
–2	175.33	3.83
–3	178.00	3.93
–4	187.66	3.96

Nguồn: Damodar Gujarati, "Quan hệ giữa chỉ số muốn giúp đỡ và tỉ lệ thất nghiệp. Phân tích thống kê, 1962-1967, "The Quartely Review of Economies and Business, tập 8, 1968, trang 67-73.

Bởi vì hồi qui (12.7.1) gặp khó khăn do tương quan chuỗi, chúng ta không thể tin vào các sai số chuẩn đã ước lượng và các tỉ lệ t vì các lý do đã nêu ra. Vì vậy, các biện pháp sửa chữa là cần thiết. Tất nhiên, biện pháp sửa chữa phụ thuộc vào ρ, nó có thể được ước lượng bởi một hoặc nhiều hơn các phương pháp đã thảo luận trước đây. Đối với ví dụ minh họa của chúng ta, ρ ước lượng từ các phương pháp khác nhau là như sau:

Phương pháp sử dụng	**ρ**	**Ghi chú**
Durbin-Watson d	0.5446	Xem (12.6.12)
Theil-Nagar d	0.5554	Xem bài tập 12.6
Bước 1 của quy trình Cochrance-Orcutt	0.5457	
Qui trình lặp Cochrance-Orcutt		
Phép lặp I	0.54571	
Phép lặp II	0.57223	
Phép lặp III	0.57836	
Phép lặp IV	0.57999	
Phép lặp V	0.58040	
Durbin hai bước	0.79517	

Như là người đọc có thể thấy, trị thống kê Durbin-Watson d, Theil-Nagar cải biến d, Bước I của qui trình hai bước Cochrane-Orcutt, và qui trình lặp Cochrane-Orcutt, tất cả đều thu được các ước lượng của ρ, chúng cũng hoàn toàn tương tự, nhưng ước lượng từ qui trình hai bước Durbin là khác hẳn.[38]

Câu hỏi thực tiễn sau đó là: Phương pháp ước lượng ρ nào ta cần chọn trên thực tế? Chúng ta sẽ trả lời câu hỏi này ngay sau đây. Hiện tại, chúng ta sẽ tiếp tục với ví dụ của chúng ta và minh họa việc làm thế nào để tìm ước lượng cho phương trình sai phân tổng quát hóa (hoặc ước lượng GLS khả thi), bằng cách sử dụng một trong các giá trị ρ này.

Chúng ta sử dụng phép tìm xấp xỉ mẫu nhỏ Theil-Nager của d. Sử dụng công thức cho trong bài tập 12.6, chúng ta thu được ρ = 0,5554. Với ước lượng này, chúng ta biến đổi dữ liệu của chúng ta như sau:

$$\ln HWI^*_t = \ln HWI_t - 0{,}5554 \ln HWI_{t-1}$$
$$\ln U^*_t = \ln U_t - 0{,}5554 \ln U_{t-1}$$

Tức là, lấy các giá trị hiện tại của biến trừ 0,5554 lần giá trị trước đó của biến. Do quan sát đầu tiên không có giá trị trước, chúng ta có hai lựa chọn: (1) bỏ nó ra khỏi phân tích, hoặc (2) bao gồm nó qua biến đổi Prais-Winsten, nó trở thành $\left[\sqrt{1-(0{,}5554)^2}\cdot \ln HWI_1\right]$ và $\left[\sqrt{1-(0{,}5554)^2}\cdot \ln U_1\right]$ trong trường hợp này. Chúng ta trình bày các kết quả của chúng ta theo 2 cách:

[38] Chênh lệch này có thể do một nguyên nhân kỹ thuật. Nếu bạn xem xét (12.6.19) cẩn thận, bạn sẽ thấy có hai ước lượng của ρ, một thu được trực tiếp từ giá trị trễ của Y và một thu được từ phép chia hệ số của giá trị trễ của X cho hệ số của X. Không bảo đảm rằng hai ước lượng sẽ như nhau. Vấn đề thực sự ở đây là (12.6.19) là một mô hình phi tuyến tính (theo thông số) *một cách nội tại* và cần được ước lượng bằng các qui trình ước lượng hồi qui phi tuyến tính, chúng nằm ngoài phạm vi cuốn sách này.

Bỏ qua quan sát thứ nhất:

$$\widehat{\ln HWI^*_t} = 3{,}1284 - 1{,}4672 \ln U^*_t \qquad N = 23 \qquad (12.7.2)$$
$$se = (0{,}0886)\ (0{,}1328) \qquad r^2 = 0{,}9685$$
$$t = (38{,}326)\ (-11{,}045) \qquad d = 1{,}77$$

trong đó các biến được lưu ý là các biến đã cải biến như chỉ ra trước đâu. Lưu ý rằng 3,1284 = β1 (1-ρ) = β1 (1- 0,5554) từ đó ta thu được β1 = 7,0364, có thể so sánh được với β1 trong phép hồi qui ban đầu (12.7.1).

Bao gồm quan sát thứ nhất (biến đổi Prais-Winsten)[39]

$$\ln HWI^*_t = 3{,}1361 - 1{,}4800 \ln U^*_t \qquad N = 24 \qquad (12.7.3)$$
$$se = (0{,}0813)\ (0{,}1198) \qquad r^2 = 0{,}968$$
$$t = (38{,}583)\ (-12{,}351) \qquad d = 1{,}83$$

So sánh phép hồi qui ban đầu (bị khó khăn bởi tự tương quan) (12.7.1) với hồi qui biến đổi (12.7.2) và hồi qui Prais-Winsten (12.7.3), chúng ta thấy rằng nói chung các kết quả là tương hợp.[40] Câu hỏi thực tế là: Chúng ta đã giải quyết xong vấn đề tự tương quan chưa? Nếu chúng ta lấy các giá trị Durbin-Watson có trong (12.7.2) và (12.7.3) với các giá trị bề mặt của chúng, có vẻ như là không còn tự tương quan (bậc 1) nữa (Vì sao?) Tuy nhiên, như Kenneth White đã ghi nhận trong SHAZAM (trang 86), các bảng Durbin-Watson có thể không phù hợp với kiểm định tương quan chuỗi trong dữ liệu đã được điều chỉnh đối với tự tương quan. Vì vậy, chúng ta có thể sử dụng một trong các kiểm định phi thông số đã thảo luận trước đây. Đối với hồi qui (12.7.2), có thể thấy rằng trên cơ sở kiểm định các cuộc chạy, người ta không thể bác bỏ giả thiết cho rằng không có tương quan chuỗi trong các phần dư từ hồi qui này. (Xem bài tập 12.20). Đối với phép hồi qui Prais-Winsten (12.7.3) cũng thấy rằng các phần dư ước lượng từ hồi qui này không có vấn đề tương quan chuỗi. Hãy kiểm tra điều này trực tiếp. Để tham khảo, có 11 phần dư mang giá trị dương. 13 phần dư mang giá trị âm, và số cuộc chạy là 12).

Nếu chúng ta muốn kiểm định giả thiết về các thông số, bây giờ chúng ta có thể tiếp tục theo kiểu thông thường. Nhưng lưu ý rằng, do chúng ta đang ước lượng ρ, các kiểm định thông thường về mức ý nghĩa sẽ có hiệu lực chặt chẽ chỉ với các mẫu lớn. Trong các mẫu nhỏ, các kết quả kiểm định sẽ chỉ là xấp xỉ. Ví dụ, từ (12.7.3) chúng ta có thể kết luận rằng hệ số góc thực là khác không về mặt thống kê. Nhưng chúng ta cần cảnh giác một chút ở đây vì mẫu 23 quan sát của chúng ta không phải là quá lớn.

---

[39] Một điểm kỹ thuật. Số hạng tung độ gốc trong phép hồi qui Prais-Winsten hơi phức tạp. Kết quả là, ta cần thực hiện phép hồi qui này qua gốc tọa độ. Tung độ gốc cho trong (12.7.3) đã thay đổi. Về chi tiết, xin xem Kenneth J. White và Linda T.M. Bùi, sổ tay máy vi tính sử dụng SHAZAM, McGraw-Hill, New York, 1985, trang 86. Đối với các chi tiết lý thuyết, xin xem Jan Kmenta; Các cơ sở Kinh tế lượng, in lần thứ 2, Macmillan, New York, 1986, trang 303-305.

[40] Nhưng hãy nhớ rằng trong các mẫu nhỏ, các kết quả có thể nhạy cảm với việc có bao gồm hay không bao gồm quan sát thứ nhất.

So sánh các phương pháp. Chúng ta quay lại vấn đề đã nêu ra trước đây: Phương pháp ước lượng ρ nào ta cần sử dụng trên thực tế để thực hiện hồi qui sai phân tổng quát hóa, hoặc hồi qui GLS khả thi? Nếu chúng ta đang xử lý các mẫu lớn (chẳng hạn, vượt quá 60-70 quan sát), thì không có khác biệt nhiều khi chọn phương pháp khác nhau, vì tất cả chúng đều thu được các kết quả gần như nhau. Nhưng điều đó không đúng trong các mẫu hữu hạn hoặc nhỏ, vì các kết quả có thể phụ thuộc vào phương pháp đã chọn. Trong các mẫu nhỏ, lúc đó, phương pháp nào được ưa thích? Không may, không có câu trả lời xác định cho câu hỏi này vì các nghiên cứu mẫu nhỏ đã được thực hiện theo các phương pháp khác nhau, qua các mô phỏng Monte Carlo, không có lợi cho hẳn một phương pháp nào.[41]Tuy nhiên, trên thực tế phương pháp thường được sử dụng là phương pháp lặp Cochrane-Oreutt mà hiện nay được kết vào nhiều chương trình máy vi tính, như là ET, SHAZAM, TSP, và SAS. Khi phần mềm máy vi tính trở nên phức tạp hơn, chúng ta có thể sử dụng các phương pháp ước lượng ρ phục vụ đặc biệt cho việc xử lý các mẫu nhỏ. Các phần mềm như SAS đã có ML và một số qui trình phi tuyến tính để ước lượng ρ (Xem trình tự AUTOREG của ASA).

12.8 MÔ HÌNH PHƯƠNG SAI THAY ĐỔI CÓ ĐIỀU KIỆN TỰ HỒI QUI (Autoregressive Conditional Heteroscesdaticity-ARCH)

Kinh nghiệm thông thường cho thấy là vấn đề tự hồi qui là một nét đặc trưng của dữ liệu chuỗi thời gian và phương sai thay đổi là một nét đặc trưng của dữ liệu chéo. Phương sai thay đổi có thể xuất hiện trong dữ liệu chuỗi thời gian không? Và tại sao?

Các nhà nghiên cứu có tham gia vào việc dự báo chuỗi thời gian tài chính, như là giá cổ phiếu, tỉ lệ lạm phát, tỉ giá hối đoái, v.v…. đã quan sát thấy rằng khả năng dự báo các biến số này của họ khác nhau đáng kể từ thời đoạn này sang thời đoạn khác.[42] Đối với một số giai đoạn, các sai số dự báo là tương đối nhỏ, đối với một thời đoạn chúng lại tương đối lớn, và sau đó chúng lại nhỏ lại đối với các thời đoạn khác. Sự biến thiên này có thể do sự biến động trên thị trường tài chính, nhạy cảm như khi chúng phản hồi các sự đồn đại, các biến động chính trị, các thay đổi trong chính sách tiền tệ và tài chính của chính phủ, và những gì tương tự. Điều này có thể đề xuất rằng phương sai của các sai số dự là hằng số và thay đổi từ thời đoạn này sang thời đoạn khác, tức là có một kiểu tự tương quan trong phương sai của các sai số dự báo.

Do hành vi của các sai số có thể được giả định là phụ thuộc vào hành vi của các nhiễu (hồi qui) u_t, ta có thể giải quyết đối với tự tương quan trong phương sai của u_t. Để nắm được tương quan này, Engle đã phát triển **mô hình phương sai thay đổi có điều kiện tự hồi qui (ARCH).**[43] Ý nghĩa then chốt của ARCH là phương sai của u tại thời gian t ($= \sigma_t^2$) phụ thuộc vào kích cỡ của số hạng bình phương sai số tại thời gian (t-1), tức là, vào u^2_{t-1}.

---

[41] Để ôn lại các nghiên cứu này, Xem J. Johnson, op.cit, trang 326-327. Một xử lý hơi cao cấp hơn có thể thấy trong A.C, Harvey, Phân tích Kinh tế lượng Chuỗi thời gian, John Wiley & Sons, New York, 1981, trang 196-199.
[42] Ví dụ, xem M.Mandelbrot, "Biến thiên của giá đầu cơ cụ thể", Journal of Business, tập 36, 1963, trang 394-419.
[43] R. Engle, "Phương sai thay đổi có điều kiện tự hồi qui với các ước lượng phương sai của lạm phát tại Vương Quốc Anh", Econometrica, tập 50, số 1, 1982, trang 987-1007. Đồng thời xem A. Bera và M. Higgins, "Tóm tắt các mô hình ARCH: Sự thúc đẩy, lý thuyết và thực hành" Journal of Economic Surveys, Forthcoming.

Để cụ thể hơn, chúng ta hãy quay lại mô hình hồi qui k biến:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + u_t \tag{12.8.1}$$

và giả định là có điều kiện về thông tin đang có sẵn tại thời gian (t-1), số hạng nhiễu tuân theo phân phối như sau:

$$u_t \sim N\,[0, (\alpha_o + \alpha_1 u^2_{\ t-1})] \tag{12.8.2}$$

tức là u_t có phân phối chuẩn với giá trị trung bình bằng 0 và phương sai là $(\alpha_o + \alpha_1 u^2_{t-1})$.

Tính chuẩn của u_t không phải là mới đối với chúng ta (Vì sao?). Điểm mới là phương sai của u tại thời gian t phụ thuộc vào bình phương nhiều tại thời gian (t-1), do đó cho thấy tương quan chuỗi.

Do trong (12.8.2) phương sai của u_t phụ thuộc vào số hạng bình phương nhiễu trong giai đoạn trước, nó được gọi là quá trình **ARCH (1)**. Nhưng chúng ta có thể tổng quát hóa dễ dàng. Do đó, một quá trình **ARCH (p)** có thể được viết là:

$$\text{var}\,(u_t) = \sigma^2_t \quad = \alpha_0\, \alpha_1 u^{\ 2}_{t-1} + \alpha_2 u^2_{t-2} + \ldots + \alpha_p u^2_{t-p} \tag{12.8.3}$$

Nếu không có tự tương quan trong phương sai sai số, chúng ta có H_o: $\alpha_1 = \alpha_2 = \ldots = \alpha_p = 0$, trong trường hợp này Var (u_t) α_o và chúng ta có trường hợp phương sai sai số không đổi.

Như Engle đã trình bày, một kiểm định của *giả thiết không* trên đây có thể dễ dàng thực hiện bằng cách thực hiện phép hồi qui sau:

$$u^2_t = \alpha_0 + \alpha_1 u^2_{t-1} + \alpha_2 u^2_{t-2} + \ldots + \alpha_p u^2_{t-p} \tag{12.8.4}$$

trong đó u, như thường lệ, ký hiệu cho các phần dư OLS đã ước lượng từ mô hình hồi qui ban đầu, (12.8.1)

Ta có thể kiểm định *giả thiết không* H_o bằng kiểm định F thông thường trong Chương 8 hoặc, khác đi, là bằng cách tính nR^2, trong đó R^2 là hệ số xác định từ hồi qui phụ trợ (12.8.4). Ta có thể thấy rằng:

$$nR^2 \sim \chi^2_p \tag{12.8.5}$$

tức là, nR^2 tuân theo phân phối Chi-bình phương với bậc tự do df bằng số hạng tự hồi qui trong hồi qui phụ.

Ví dụ minh họa: Chúng ta hãy tiếp tục với ví dụ tiền công–năng suất đã được sử dụng quá nhiều của chúng ta. Bằng cách sử dụng các phần dư thu được từ phép hồi qui này, chúng ta ước lượng các mô hình ARCH (1), ARCH (2), ARCH (3), ARCH (4) và ARCH (5). Nhưng chỉ có mô hình ARCH (1) được chứng minh là có ý nghĩa. Các kết quả của mô hình này như sau:

$$\begin{aligned} u^2_t &= 2{,}0746 + 0{,}6946\ u^2_{t-1} \\ t &= (1{,}0583)\ \ (5{,}0364) \qquad R^2 = 0{,}4665 \\ & \qquad\qquad\qquad\qquad\quad d = 1{,}67 \end{aligned} \tag{12.8.6}$$

Áp dụng (12.8.5), chúng ta thấy rằng $nR^2 = (31)\ (0{,}4665) = 14{,}46$, nó xấp xỉ χ^2 với một bậc tự do. Từ bảng Chi-bình phương rõ ràng xác suất của việc thu được giá trị Chi-bình phương như vậy là nhỏ hơn 0,005 (giá trị p là khoảng 0,000143). Điều này đề xuất rằng trong ví dụ của chúng ta, phương sai sai số có tương quan chuỗi.

Phải làm gì nếu tồn tại ARCH?

Hãy nhớ lại rằng chúng ta đã thảo luận nhiều phương pháp hiệu chỉnh phương sai thay đổi, nó về cơ bản bao gồm việc áp dụng OLS vào dữ liệu đã cải biến. Hãy nhớ rằng OLS áp dụng vào dữ liệu đã cải biến là bình phương tối thiểu tổng quát hóa (GLS). Nếu đã tìm ra hiệu ứng ARCH, chúng ta sẽ cần sử dụng GLS. Để cho gọn, các chi tiết của lý thuyết và các cơ chế của điều này được bố trí vào phần tham khảo.[44]

Nhân đây, sự tổng quát hóa của mô hình ARCH được gọi là **GARCH**, trong đó phương sai có điều kiện của u tại thời gian t phụ thuộc không chỉ vào bình phương nhiễu trong quá khứ, mà còn vào các phương sai có điều kiện trong quá khứ. Chi tiết có thể thấy trong phần tham khảo.[45]

Một lời về trị thống kê Durbin-Watson d và hiệu ứng ARCH

Hãy nhớ lại khi chúng ta hồi qui tiền công theo năng suất, chúng ta đã thu được giá trị của d là 0,1380, cho thấy rõ là có mối tương quan chuỗi bậc 1 đồng biến trong số hạng sai số. Nhưng kết luận này bây giờ có vẻ chưa chín muồi do hiệu ứng ARCH. Nói cách khác, tương quan chuỗi quan sát được trong u_t có thể là do hiệu ứng ARCH và không phải là do tương quan chuỗi trên mỗi sai số chuẩn, se. Vì vậy, trong các phân tích chuỗi thời gian, đặc biệt các phân tích có bao gồm dữ liệu tài chính, ta cần kiểm định hiệu ứng ARCH trước khi chấp nhận trị thống kê d được in thường xuyên với giá trị bề mặt của nó.

[44] Xem Davidson v2 Mac Kinnon , op.cit., Phần 164. Cũng xem William H, Greene, op.cit., Phần 15.9, và ET của ông ta: Công cụ Kinh tế học, ấn bản 3,6, Phần mềm Kinh tế lượng, Inc., Bellport, New York, 1992,Chương 29 MICRO TSP 7,0 và SHAZAM 7,0 có các qui trình kiểm định ARCH.

[45] T. Bollerslev, "Phương sai thay đổi có điều kiện tự hồi qui tổng quát hóa", Journal of Econometrica, tập 31, trang 307-326.

12.9 TÓM TẮT VÀ CÁC KẾT LUẬN

1. Nếu giả định của mô hình hồi qui tuyến tính cổ điển cho rằng các sai số hoặc các nhiễu u_t được đưa vào mô hình hồi qui tổng thể là ngẫu nhiên hoặc không có tương quan bị vi phạm, thì vấn đề tự tương quan hoặc tương quan chuỗi xuất hiện.

2. Tự tương quan có thể xuất hiện do nhiều nguyên do, như là sự trì trệ hoặc chậm chạp của các chuỗi thời gian kinh tế, các thiên lệch của đặc trưng do không đưa các biến quan trọng vào mô hình hoặc do sử dụng dạng hàm không chính xác, hiện tượng Cobweb, nhào nặn dữ liệu, v.v...

3. Mặc dù các hàm ước lượng OLS vẫn không thiên lệch và nhất quán khi có tự tương quan, chúng không còn hiệu quả nữa. Kết quả là, các kiểm định thông thường t và F về mức ý nghĩa không thể được áp dụng một cách hợp lệ. Do đó, ta cần phải có các số các biện pháp sửa chữa.

4. Cách sửa chữa phụ thuộc vào bản chất của mối liên phụ thuộc giữa các nhiễu u_t. Nhưng do u_t là không thể quan sát được, nên trên thực tế thường là giả định rằng chúng được tạo ra bởi một cơ chế nào đó.

5. Cơ chế thường được giả định là sơ đồ tự hồi qui bậc 1 Markov, nó giả định rằng nhiễu trong thời đoạn hiện tại có quan hệ tuyến tính với số hạng nhiễu trong thời đoạn trước đó, hệ số tự tương quan cho ta phạm vi của mối liên phụ thuộc. Cơ chế này được biết như là sơ đồ AR(1).

6. Nếu sơ đồ AR(1) có hiệu lực và đã biết hệ số tự tương quan, thì vấn đề tương quan chuỗi có thể bị tấn công một cách dễ dàng bằng cách biến đổi dữ liệu theo qui trình sai phân tổng quát. Sơ đồ AR(1) có thể được tổng quát hóa dễ dàng thành sơ đồ AR(p). Người ta cũng có thể giả định một cơ chế trung bình trượt (MA) hay một sự phối hợp giữa sơ đồ AR và MA gọi là ARMA.

7. Mặc dù ta sử dụng sơ đồ AR(1), hệ số tự tương quan ρ là không biết trước. Chúng ta đã xem xét nhiều phương pháp ước lượng ρ, như là Durbin-Warson d, Theil-Nagar đã cải biến d, qui trình hai bước Cochrane-Orcutt (C-O), qui trình lặp C-O, và phương pháp hai bước Durbin. Trong các mẫu lớn, các phương pháp này nói chung đều thu được các ước lượng tương tự, mặc dù trong các mẫu nhỏ chúng cho kết quả khác nhau. Trên thực tế, phương pháp lặp C-O đã trở nên hoàn toàn thông dụng.

8. Tất nhiên, trước khi các biện pháp sửa chữa trở thành sự khám phá quan hệ tự tương quan. Có nhiều phương pháp khám phá, trong đó phương pháp nổi tiếng nhất là trị thống kê Durbin-Watson d. Mặc dù được sử dụng rộng rãi và được in ra thường lệ bởi đa số các phần mềm máy vi tính, trị thống kê d có nhiều hạn chế. Rất thường gặp, trị thống kê d là dấu hiệu

không phải của tự tương quan thuần túy mà là của các thiên lệch trong đặc trưng mô hình hoặc hiệu ứng ARCH.

9. Mô hình đặc biệt mà ta đã thảo luận trong chương này là mô hình ARCH trong đó phương sai có điều kiện của số hạng sai số có tương quan chuỗi với các giá trị bình phương của số hạng sai số trong quá khứ. Mô hình này đã được chứng minh là hữu ích trong việc lập mô hình và dự báo nhiều biến tài chính, như là tỉ giá hối đoái, tỉ lệ lạm phát, v.v...

BÀI TẬP

Các câu hỏi:

12.1 Hãy khẳng định có phải các phát biểu sau đây là đúng hay sai. Hãy chứng minh ngắn gọn câu trả lời của bạn.

(a) Khi có tự tương quan, các hàm ước lượng OLS là thiên lệch cũng như không hiệu quả.

(b) Kiểm dịnh Durbin-Watson d giả định rằng phương sai của số hạng sai số u_t là không đổi.

(c) Sự biến đổi sai phân bậc 1 để hủy bỏ tự tương quan giả định rằng hệ số tự tương quan ρ là -1.

(d) Các giá trị R^2 của hai mô hình, một bao gồm hồi qui dưới dạng sai phân bậc 1 và mô hình kia dưới dạng mức độ, không thể so sánh một cách trực tiếp.

(e) Trị thống kê Durbin-Watson có ý nghĩa thì không nhất thiết là có tự tương quan bậc 1.

(f) Khi có tự tương quan, các phương sai và các sai số chuẩn của các giá trị dự báo được tính toán theo thông lệ là không hiệu quả.

(g) Việc bỏ (các) biến quan trọng ra ngoài mô hình hồi qui có thể cho một giá trị d đáng kể.

(h) Trong sơ đồ AR(1), kiểm định giả thiết cho rằng $\rho = 0$ có thể được lập ra bởi trị thống kê Berenblutt-Webb g cũng như trị thống kê Durbin-Watson d.

(i) Trong phép hồi qui sai phân bậc 1 của Y theo sai phân bậc 1 của X, nếu có một số hạng không đổi và một số hạng xu hướng tuyến tính, nó có nghĩa là trong mô hình ban đầu có một số hạng tuyến tính cũng như một số hạng xu hướng bậc 2.

12.2 Cho trước một mẫu gồm 50 quan sát và 4 biến giải thích, bạn có thể nói gì về tự tương quan nếu (a) d = 0,05? (b) d = 1,40? (c) d = 2,50? (d) d = 3,97?

12.3 Trong nghiên cứu biến động của tỉ phần công nhân sản xuất dưới dạng giá trị gia tăng (tức là tỉ phần nhân công), các mô hình sau đã được Gujarati xem xét:*

Mô hình A: $Y_t = \beta_0 + \beta_1 t + u_t$
Mô hình B: $Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + u_t$

trong đó Y = tỉ phần nhân công và t = thời gian. Dựa vào dữ liệu hàng năm từ 1949 đến 1964, thu được các kết quả sau đây đối với ngành công nghiệp luyện kim:

Mô hình A : $Y_t = 0{,}4529 - 0{,}0041t$ $\quad R^2 = 0{,}5284 \quad d = 0{,}8252$
(-3,9608)

Mô hình B : $Y_t = 0{,}4786 - 0{,}0127t + 0{,}0005t^2$
(-3,2724) (2,7777)
$R^2 = 0{,}6629 \quad d = 1{,}82$

trong đó các số trong ngoặc là các tỉ lệ t.

(a) Có tương quan chuỗi trong mô hình A? Trong mô hình B?
(b) Có giải thích nào cho tương quan chuỗi?
(c) Bạn phân biệt thế nào giữa tự tương quan "thuần túy" và các thiên lệch trong đặc trưng mô hình?

12.4 *Phát hiện tự tương quan*: **Kiểm định tỉ lệ von Neumann**[*]**.** Khi giả định rằng các phần dự u_t là các kết quả rút ra một cách ngẫu nhiên từ phân phối xác suất chuẩn, von Neumann đã chỉ ra rằng đối với *n lớn*, tỉ lệ

$$\frac{\delta^2}{s^2} = \frac{\sum(\hat{u}_i - \hat{u}_{i-1})^2/(n-1)}{\sum(\hat{u}_i - \bar{\hat{u}})^2/n} \qquad \text{lưu ý rằng } \bar{\hat{u}} = 0 \text{ trong OLS}$$

được gọi là *tỉ lệ von Neumann,* có phân bố xác suất xấp xỉ chuẩn với giá trị trung bình:

$$E\frac{\delta^2}{s^2} = \frac{2n}{n-1}$$

và phương sai:

$$\operatorname{var}\frac{\delta^2}{s^2} = 4n^2\frac{n-2}{(n+1)(n-1)^3}$$

(a) Nếu n là đủ lớn, bạn sử dụng tỉ lệ von Neumann như thế nào để kiểm định tự tương quan?

(b) Mối liên hệ giữa trị thống kê Durbin-Watson d và tỉ lệ này là gì?

[*] J. von Nemann, "Phân bố xác suất của tỉ lệ của trung bình bình phương sai phân liên tiếp đối với phương sai", Annals of Mathematical Statistics, tập 12, 1941, trang 367-395.

(c) Trị thống kê d nằm giữa 0 và 4. Các giới hạn tương ứng đối với tỉ lệ von Neumann là gì?

(d) Do tỉ lệ này phụ thuộc vào giả định rằng các u là các kết quả ngẫu nhiên từ phân bố xác suất chuẩn, giả định này hiệu lực như thế nào đối với các phần dư OLS?

(e) Giả sử trong một ứng dụng, tỉ lệ này được tìm ra là 2,88 với 100 quan sát. Hãy kiểm định giả thiết cho rằng không có tương quan chuỗi trong dữ liệu.

Lưu ý: B.I. Hart đã lập bảng các giá trị tới hạn của tỉ số von Neumann đối với các cỡ mẫu dưới 60 quan sát. [+]

12.5 Trong một chuỗi gồm 17 phần dư, 11 dương và 6 âm, số lần chạy là 3. Có bằng chứng tự tương quan không? Câu trả lời có thay đổi không nếu số lượng chạy là 14?

12.6 **Ước lượng Theil-Nagas về r dựa trên trị thống kê d**. Theil và Nagar đã đề xuất rằng trong các mẫu nhỏ, thay cho việc ước lượng ρ như là (1 - d/2), có thể ước lượng như là:

$$\hat{\rho} = \frac{n^2(1 - d/2) + k^2}{n^2 - k^2}$$

trong đó n = tổng số các quan sát, d = Durbin-Watson d, và k = số các hệ số (bao gồm tung độ gốc) cần phải ước lượng.

Hãy chứng tỏ rằng đối với n lớn, ước lượng ρ này là bằng với ước lượng thu được bởi công thức đơn giản hơn (1 - d/2).

12.7 *Ước lượng ρ* : **Qui trình rà soát hoặc nghiên cứu Hildreth - Lu.**[**] Do trong sơ đồ tự hồi qui bậc 1.

$$u_t = \rho u_{t-1} + \varepsilon_t$$

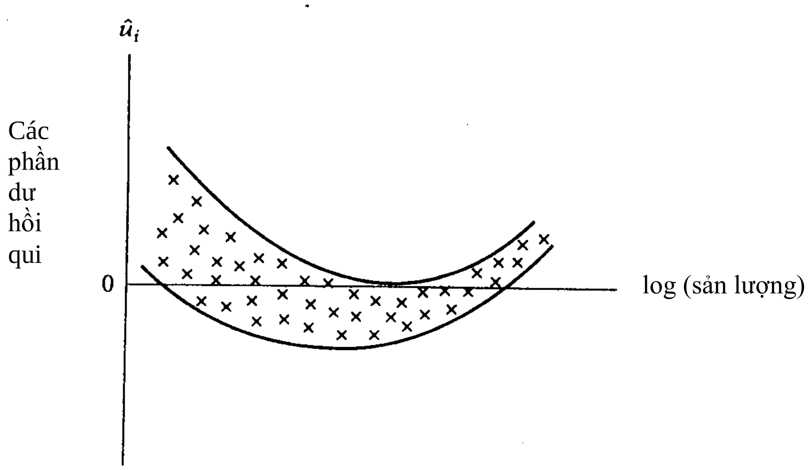
ρ được kỳ vọng nằm giữa -1 và +1, Hildreth và Lu đề xuất một qui trình "rà soát" có hệ thống hoặc nghiên cứu để tìm vị trí của nó. Họ kiến nghị chọn ρ giữa -1 và +1, bằng cách sử dụng, chẳng hạn các khoảng đơn vị là 0,1 và biến đổi dữ liệu nhờ phương trình sai phân tổng quát hóa (12.6.5). Vì vậy, ta có thể chọn ρ từ -0,9; -0,8;...; 0,8; 0,9. Đối với mỗi ρ đã chọn, chúng ta chạy phương trình sai phân tổng quát hóa và thu được RSS kèm theo: Σu^2_t. Hildreth và Lu đề xuất chọn ρ sao cho cực tiểu hóa RSS (vì vậy tối đa hóa R^2). Nếu việc chỉnh đốn tiếp theo là cần thiết, thì họ kiến nghị sử dụng các khoảng đơn vị nhỏ hơn, chẳng hạn 0,01 như là 0,99; 0,98; ...; 0,90; 0,91. v.v...

[+] Bảng này có thể thấy trong Johnston, op.cit, in lần thứ 3, trang 559.
[**] G. Hildreth và J.Y.Lu, "Các quan hệ cầu với các nhiễu tự tương quan", Trường ĐHTH Michigan, Trạm thí nghiệm nông nghiệp, Tech. Bull. 276, T.11/1960.

(a) Các lợi điểm của qui trình Hildreth - Lu là gì?

(b) Làm thế nào người ta biết được rằng giá trị ρ cuối cùng được chọn để biến đổi dữ liệu trên thực tế sẽ bảo đảm $\Sigma\ u^2_t$ cực tiểu?

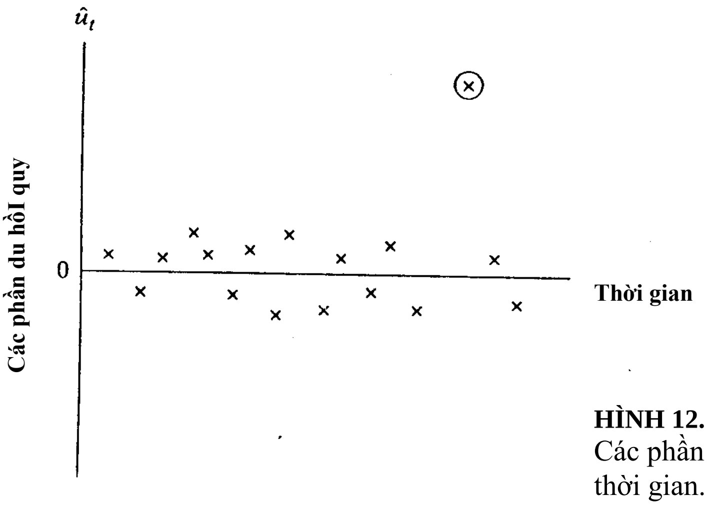
12.8 Khi đo các lợi ích theo qui mô trong việc cung cấp điện, Nerlove đã sử dụng dữ liệu chéo của 145 công ty tư nhân tại Hoa Kỳ trong thời đoạn 1955 và đã được thực hiện hồi qui lôgarit của tổng chi phí theo các lôgarit của sản lượng, mức tiền công, giá vốn và giá nhiên liệu. Ông ta đã tìm ra rằng các phần dư được ước lượng từ phép hồi qui này tỏ ra có tương quan "chuỗi", như đã chứng thực bởi Durbin-Watson d. Để tìm một biện pháp sửa chữa, ông ta đã thể hiện các phần dư đã ước lượng theo lôgarit của sản lượng trên đồ thị và thu được Hình 12.10.



HÌNH 12.10 Các phần dư hồi qui từ nghiên cứu của Nerlove (Từ Marc Nerlove, "Lợi ích theo qui mô trong việc cung cấp điện", trong Carl F. Christ et al., Đo lường trong kinh tế học, Stanford University Press, Stanford, Calif., 1963.

(a) Hình 12.10 chỉ ra điều gì?
(b) Làm thế nào bạn loại bỏ tương quan "chuỗi" trong tình thế trên đây?

12.9 Các phần dư từ một phép hồi qui khi đã được vẽ lên đồ thị theo thời gian cho một đồ thị phân tán trong Hình 12.11. Phần dư "cực" được khoanh tròn được gọi là *điểm nằm ngoài*. Điểm nằm ngoài là một quan sát mà giá trị của nó vượt quá lớn so với các giá trị của các quan sát khác trong mẫu, có thể là cách xa giá trị trung bình của mọi quan sát ba hoặc bốn độ lệch chuẩn.



HÌNH 12.11
Các phần dư hồi qui giả thiết theo thời gian.

(a) Các nguyên nhân tồn tại (các) điểm nằm ngoài là gì?

(b) Nếu có (các) điểm nằm ngoài thì có nên loại bỏ (các) quan sát này và thực hiện hồi qui với các quan sát còn lại hay không.

(c) Durbin-Watson d có thể áp dụng được trong trường hợp có (các) điểm nằm ngoài hay không?

12.10 Chứng minh Phương trình (12.6.9)

*12.11 Giả định sơ đồ hồi qui bậc nhất $u_t = \rho u_{t-1} + \varepsilon_t$ trong đó ε_t thỏa mãn các giả định của mô hình hồi qui tuyến tính cổ điển.

(a) Hãy chứng tỏ rằng var $(u_t) = \sigma^2/(1 + \rho^2)$, trong đó σ^2 = var (ε_t)

(b) Đồng phương sai giữa u_t và u_{t-1} là gì? Giữa u_t và u_{t-2}? Hãy tổng quát hóa các kết quả của bạn.

(c) Hãy viết ma trận phương sai – đồng phương sai của các u

(d) Nếu $\rho = 1$, điều gì xảy ra với phương sai của u_t? Nó hàm chứa các ý nghĩa gì đối với việc biến đổi sai phân thứ nhất?

* Tùy ý (Không bắt buộc)

12.12 Liên quan đến Phương trình (12.4.1). Giả sử r = 0 nhưng $\rho \neq 0$. Tác động lên E (σ^2) là gì nếu (a) $0 < \rho < 1$ và (b) $-1 < \rho < 0$? Khi nào các thiên lệch trong $\hat{\sigma}^2$ sẽ có giá trị nhỏ một cách hợp lý?

12.13 Dựa trên trị thống kê Durbin-Watson d, làm thế nào bạn phân biệt được tự tương quan "thuần túy" và các thiên lệch trong đặc trưng mô hình?

12.14 Giả sử trong mô hình

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

các u trên thực tế là độc lập về chuỗi. Điều gì xảy ra trong tình huống này nếu, giả sử rằng $u_t = \rho u_{t-1} + \varepsilon_t$, chúng ta sử dụng phép hồi qui sai phân tổng quát hóa

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

Hãy thảo luận cụ thể từng tính chất của số hạng nhiễu ε_t.

12.15 Trong một nghiên cứu về việc định giá của sản lượng cuối cùng theo chi phí nhân tố sản xuất tại Vương quốc Anh, các kết quả sau đây đã thu được trên cơ sở dữ liệu hàng năm trong thời đoạn 1951-1969:

$$\begin{aligned}
\widehat{PF}_t = \ & 2{,}033 + 0{,}273W_t - 0{,}521X_t + 0{,}256M_t + 0{,}028M_t\text{-}1 + 0{,}121PF_{t-1} \\
& (0{,}992) \quad (0{,}127) \quad (0{,}099) \quad (0{,}024) \quad (0{,}039) \quad (0{,}119) \\
& R2 = 0{,}984 \qquad d = 2{,}54
\end{aligned}$$

trong đó PF = các giá của sản lượng cuối cùng theo chi phí nhân tố sản xuất, W = tiền công và tiền lương cho mỗi công nhân, X = tổng sản phẩm quốc nội theo đầu người lao động, M = giá nhập khẩu, M_{t-1} = giá nhập khẩu trễ 1 năm, và PF_{t-1} = giá của sản lượng cuối cùng với chi phí nhân tố sản xuất trong năm trước đó.[+]

"Do đối với 18 quan sát và 5 biến giải thích, các giá trị tới hạn dưới và trên của d ứng với 5% là 0,71 và 2,06, giá trị ước lượng của d là 2,54 cho thấy không có tự tương quan đồng biến". Hãy bình luận.

12.16 Hãy nêu các hoàn cảnh trong đó từng phương pháp ước lượng hệ số tự tương quan bậc nhất ρ có thể là phù hợp:

(a) Hồi qui sai phân thứ nhất
(b) Hồi qui trung bình trượt
(c) Biến đổi Theil-Nagar

[+] Nguồn: Các giá và lợi nhuận trong 1951 - 1969: Đánh giá kinh tế lượng, Phòng giới thiệu việc làm, Văn phòng Her Majesty's Stationety, 1971, Bảng C, trang 37, Phương trình 63.

(d) Qui trình lặp Cochrane và Oreutt
(e) Qui trình rà soát Hildreth - Lu
(f) Qui trình hai bước Durbin

12.17 Liệu β_1 = 7,0364 thu được từ hồi qui (12.7.2) có cho ta một ước lượng không thiên lệch của β_1 thực? Vì sao có hoặc vì sao không?

12.18 Bao gồm nhân tố hiệu chỉnh C, phương trình của β_2^{GLS} cho trong (12.3.1) là

$$\hat{\beta}_2^{GLS} = \frac{(1-\rho^2)x_1 y_1 + \sum_{t=2}^{n}(x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{(1-\rho^2)x_1^2 \sum_{t=2}^{n}(x_t - \rho x_{t-1})^2}$$

Cho trước công thức này và (12.3.1), hãy tìm biểu thức đối với nhân tố hiệu chỉnh C.

12.19 Hãy chứng tỏ rằng việc ước lượng (12.6.6) là tương đương với ước lượng GLS đã thảo luận trong Phần 12.3, không bao gồm quan sát đầu của Y và X.

12.20 Đối với hồi qui (12.7.2), các phần dư ước lượng có các dấu như sau:

$$- - - + - - + + - - + + + - + - + + - - - + +$$

Trên cơ sở của kiểm định các cuộc chạy, hãy chứng tỏ rằng người ta có thể chấp nhận giả thiết cho rằng không có tự tương quan trong các phần dư này.

*12.21 **Kiểm định đối với tương quan chuỗi bậc cao hơn**. Giả sử chúng ta có dữ liệu chuỗi thời gian trên cơ sở quí. Trong các mô hình hồi qui liên quan tới dữ liệu theo quí, thay vì sử dụng sơ đồ AR(1) cho trong (12.2.1), có thể phù hợp hơn nếu giả định sơ đồ AR(4) như sau:

$$u_t = \rho_4 u_{t-4} + \varepsilon_t$$

tức là, để giả định rằng số hạng nhiễu hiện tại có tương quan với chính nó trong cùng quí trong năm trước hơn là với chính nó của quí trước.

Để kiểm định giả thiết cho rằng $\rho_4 = 0$, Wallis + đề xuất kiểm định Durbin Watson cải biến d như sau:

---

* Tùy ý (không bắt buộc)
+ Kenneth Wallis, "Kiểm định tự tương quan bậc 4 trong các phương trình hồi qui theo quí", Econometrica, tập 40, 1972, trang 617-636. Các bảng d4 cũng có thể tìm thấy trong J. Johnston, op.cit, in lần thứ 3, trang 558.

$$d_4 = \frac{\sum_{t=5}^{n}(u_t - \hat{u}_{t-4})^2}{\sum_{t=1}^{n}\hat{u}_1^2}$$

Qui trình kiểm định tuân theo qui trình kiểm định d thông thường đã thảo luận trong bài này.

Wallis đã chuẩn bị các bảng d_4, có thể thấy trong bài báo gốc của ông ta.

Bây giờ giả sử chúng ta có dữ liệu hàng tháng. Kiểm định Durbin-Watson có thể được tổng quát hóa để xét tới dữ liệu như vậy hay không? Nếu như vậy, hãy viết ra công thức d_{12} phù hợp.

12.22 Giả sử bạn ước lượng phép hồi qui sau đây:

$$\Delta \ln \text{sản lượng}_t = \beta_1 + \beta_2 \Delta \ln L_t + \beta_3 \Delta \ln K_t + u_t$$

trong đó Y là sản lượng, L là nhập lượng nhân công, K là nhập lượng vốn, và Δ là toán tử sai phân thứ nhất. Bạn giải thích β_1 như thế nào trong mô hình này? Có thể xem nó như là một ước lượng thay đổi công nghệ hay không. Hãy chứng tỏ câu trả lời của bạn.

Các nhân tố xác định giá đồng nội địa Mỹ, 1951-1980

Năm	C	G	I	L	H	A
1951	21.89	330.2	45.1	220.4	1,491.0	19.00
52	22.29	347.2	50.9	259.5	1,504.0	19.41
53	19.63	366.1	53.3	256.3	1,438.0	20.93
54	22.85	366.3	53.6	249.3	1,551.0	21.78
55	33.77	399.3	54.6	352.3	1,646.0	23.68
56	39.18	420.7	61.1	329.1	1,349.0	26.01
57	30.58	442.0	61.9	219.6	1,224.0	27.52
58	26.30	447.0	57.9	234.8	1,182.0	26.89
59	30.70	483.0	64.8	237.4	1,553.7	26.85
60	32.10	506.0	66.2	245.8	1,296.I	27.23
61	30.00	523.3	66.7	229.2	1,365.0	25.46
62	30.80	563.8	72.2	233.9	1,492.5	23.88
63	30.80	594.7	76.5	234.2	1,634.9	22.62
64	32.60	635.7	81.7	347.0	1,561.0	23.72
65	35.40	688.1	89.8	468.1	1,509.7	24.50
66	36.60	753.0	97.8	555.0	1,195.8	24.50
67	38.60	796.3	100.0	418.0	1,321.9	24.98
68	42.20	868.5	106.3	525.2	1,545.4	25.58
69	47.90	935.5	111-1	620.7	1,499.5	27.18
70	58.20	982.4	107.8	588.6	1,469.0	28.72
71	52.00	1,063.4	109.6	444.4	2,084.5	29.00

72	51.20	1,171.I	l19.7	427.8	2,378.5	26.67
73	59.50	1,306.6	129.8	727.1	2,057.5	25.33
74	77.30	1,412.9	129.3	877.6	1,352.5	34.06
75	64.20	1,528.8	l17.8	556.6	1,171.4	39.79
76	69.60	1,700.I	129.8	780.6	1,547.6	44.49
77	66.80	1,887.2	137.1	750.7	1,989.8	51.23
78	66.50	2,127.6	145.2	709.8	2,023.3	54.42
79	98.30	2,628.8	152.5	935.7	1,749.2	61.01
80	101.40	2,633.1	147.1	940,9	1,298.5	70.87

Ghi chú: Dữ liệu này được thu thập bởi Gary R. Smith từ các nguồn như American Metal Market, Metals Week, và các công bố của Bộ Thương Mại Mỹ.

C = Giá đồng nội địa Mỹ trung bình 12 tháng (xu Mỹ / 1 pound)
G = Tổng sản phẩm quốc gia hàng năm (Tỉ đô la)
I = Chỉ số sản xuất công nghiệp trung bình 12 tháng
L = Giá trung bình 12 tháng theo London Metal Exchange (Bảng Anh)
H = Số housing starts hàng năm (ngàn đơn vị)
A = Giá nhôm trung bình 12 tháng (xu Mỹ / 1 pound)

12.23 Maddala đề xuất rằng nếu Durbin-Watson d nhỏ hơn R^2, ta cần thực hiện hồi qui ở dạng sai phân thứ nhất. Lý luận đằng sau đề xuất này là gì?

Bài tập:

12.24 Xét dữ liệu về công nghiệp đồng trong bảng đi kèm.

(a) Dựa trên dữ liệu này, hãy ước lượng mô hình hồi qui sau:

$$\ln C_t = \beta_1 + \beta_2 \ln I_t + \beta_3 \ln L_t + \beta_4 \ln H_t + \beta_5 \ln A_t + u_t$$

Hãy giải thích các kết quả

(b) Hãy thu thập các phần dư và phần dư chuẩn hóa từ phép hồi qui trên đây và vẽ đồ thị của chúng. Bạn có thể phỏng đoán gì về sự tồn tại của tự tương quan trong các phần dư này?

(c) Hãy ước lượng trị thống kê Durbin-Watson d và bình luận về bản chất của tự tương quan có trong dữ liệu.

(d) Hãy thực hiện kiểm định cuộc chạy và xem nếu như câu trả lời của bạn khác với những gì cho biết trong (c).

(e) Hãy kiểm định các phần dư đối với tác động ARCH

(f) Bạn sẽ tìm ra như thế nào nếu quá trình AR(p) mô tả tự tương quan tốt hơn là quá trình AR(1)?

Ghi nhớ: Hãy bảo toàn dữ liệu cho phân tích tiếp theo (Xem bài tập 12.26)

12.25 Bạn được biết dữ liệu đi kèm sau đây:

Y, chi tiêu tiêu dùng cá nhân (tỉ đô la theo giá trị năm 1958)	**X, thời gian**	**Y, Y ước lượng ***	**u, các phần dư**
281.4	1 (= 1956)	261.4208	19.9791
288.1	2	276.6026	11.4973
290.0	3	291.7844	-1.7844
307.3	4	306.9661	0.3338
316.1	5	322.1479	-6.0479
322.5	6	337.3297	-14.8297
338.4	7	352.5115	-14.1115
353.3	8	367.6933	-14.3933
373.7	9	382.8751	-9.1751
397.7	10	398.0569	-0.3569
418.1	11	413.2386	4.8613
430.1	12	428.4206	1.6795
452.7	13	443.6022	9.0977
469.1	14	458.7840	10.3159
476.9	15 (= 1970)	473.9658	2.9341

* Thu được từ phép hồi qui $Y_t + \beta_0\beta_1 X_t + u_t$

(a) Hãy chứng minh rằng Durbin-Watson d = 0,4147

(b) Có mối tương quan chuỗi đồng biến trong các nhiễu không?

(c) Nếu có, hãy ước tính ρ bằng

(i) Phương pháp Theil-Nagar
(ii) Qui trình hai bước Durbin
(iii) Phương pháp Cochrane-Orcutt

(d) Hãy sử dụng phương pháp Theil - Nagar để biến đổi dữ liệu và thực hiện hồi qui trên dữ liệu đã biến đổi.

(e) Có phải phép hồi qui đã ước lượng trong (d) có tự tương quan? Nếu như vậy, làm thế nào bạn loại bỏ nó?

12.26 Xét bài tập 12.24 và dữ liệu cho trong bảng với 12.24. Nếu các kết quả của bài tập này cho thấy tương quan chuỗi,

(a) Hãy sử dung qui trình hai giai đoạn Cochrane - Orcutt và tìm các ước lượng của GLS khả thi hoặc hồi qui sai phân tổng quát hóa và so sánh các kết quả của bạn.

(b) Nếu ρ ước tính từ phương pháp Cochrane - Orcutt trong (a) khác đáng kể so với giá trị được ước tính từ trị thống kê d, bạn sẽ chọn phương pháp ước tính ρ nào và vì sao?

12.27 Xét ví dụ 7.4. Bằng cách bỏ qua các biến X^2 và X^3, hãy thực hiện phép hồi qui và xem xét các phần dư để tìm tương quan "chuỗi". Nếu tương quan chuỗi được tìm ra, bạn sẽ xử lý nó thế nào? Bạn sẽ đề xuất các biện pháp sửa chữa nào?

12.28 Xét ví dụ 7.25. Một tự tương quan tiên nghiệm được kỳ vọng trong dữ liệu như vậy. Vì vậy, có kiến nghị rằng bạn hồi qui lôgarit của cung tiền thực theo các lôgarit của thu nhập quốc gia thực và lãi suất dài hạn dưới dạng sai phân thứ nhất. Hãy thực hiện phép hồi qui này, và sau đó thực hiện lại phép hồi qui dưới dạng ban đầu. Giả định của phép biến đổi dưới dạng sai phân thứ nhất có được thỏa mãn không? Nếu không, các loại thiên lệch nào có khả năng tác động từ phép biến đổi như vậy? Hãy minh họa bằng dữ liệu hiện có.

12.29 **Sử dụng Durbin-Watson d để kiểm định tính phi tuyến tính.** Tiếp tục với bài tập 12.27. Hãy sắp xếp các phần dư thu được trong phép hồi qui đó theo giá trị tăng dần của X. Sử dụng công thức cho trong (12.5.4), hãy ước tính d từ các phần dư đã sắp xếp lại. Nếu giá trị d đã tính cho thấy tự tương quan, điều đó có ý nghĩa rằng mô hình tuyến tính đã không chính xác và rằng mô hình đầy đủ cần phải có các số hạng X^2_i và X^3_i. Bạn có thể đưa ra một giải thích trực giác đối với một qui trình như vậy không? Hãy xem nếu như câu trả lời của bạn phù hợp với điều đưa ra bởi Henri Theil.*

12.30 Xét bài tập 11.20. Hãy thu thập các phần dư và tìm ra nếu có tự tương quan trong các phần dư. Bạn sẽ chuyển đổi dữ liệu thế nào trong trường hợp tương quan chuỗi được tìm ra? Ý nghĩa của tương quan chuỗi trong khoảng thời gian hiện tại là gì?

12.31 **Thử nghiệm Monte Carlo**. Xét các Bảng 12.1 và 12.2. Sử dụng dữ liệu ε_t và X_t đã cho, hãy tạo ra một mẫu gồm 10 giá trị Y từ mô hình

$$Y_t = 3,0 + 0,5\, X_t + u_t$$

trong đó $X_t = 0,9\, u_{t-1} + \varepsilon_t$. Giả định $u_0 = 10$

* Henri Theil, Nhập môn Kinh tế lượng, Prentice-Hall, Englewood Cliffs, N.J., 1978, trang 307-308.

(a) Hãy nhận xét các kết quả của bạn.

(b) Hãy lặp lại bài tập này 10 lần và nhận xét các kết quả.

(c) Giữ nguyên cơ cấu trên, ngoại trừ bây giờ ta cho ρ=0,3 thay cho ρ=0,9 và so sánh các kết quả của bạn với những gì cho ở (b).

12.32 Sử dụng dữ liệu trong bảng đi kèm, hãy ước lượng mô hình

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

trong đó Y= tồn kho và X= doanh thu bán, cả hai được tính bằng tỷ đô la.

(a) Hãy ước lượng phép hồi qui trên đây.

(b) Từ phần dư đã ước lượng, hãy tìm nếu có tự tương quan đồng biến, sử dụng (i) kiểm định Durbin-Watson và (ii) kiểm định tính chuẩn của mẫu lớn cho trong (12.5.11).

(c) Nếu ρ là dương, hãy áp dụng kiểm định Berenblutt-Webb để kiểm định giả thiết cho rằng ρ=1.

(d) Nếu bạn nghi ngờ rằng cấu trúc sai số tự hồi qui có bậc là p, hãy sử dụng kiểm định Bereusch-Godfrey để chứng minh điều này. Bạn sẽ chọn bậc của p như thế nào?

Tồn kho và Doanh thu trong ngành chế tạo Mỹ, 1950–1991
(triệu đô la)

Năm	Doanh thu *	Tồn kho +	Năm	Doanh thu *	Tồn kho +
1950	38,596	59,822	1970	108,352	178,594
1951	43,356	70,242	1971	117,023	188,991
1952	44,840	72.377	1972	131,227	203,227
1953	47,987	76,122	1973	153,881	234,406
1954	46,443	73,175	1974	178,201	287,144
1955	51,694	79,516	1975	182,412	288,992
1956	54,063	87,304	1976	204,386	318,345
1957	55,879	89,052	1977	229,786	350,706
1958	54,201	87,055	1978	260,755	400,929
1959	59,729	92,097	1979	298,328	452,636
1960	60,827	94,719	1980	328,l12	510,124
1961	61,159	95,580	1981	356,909	547,169
1962	65,662	101,049	1982	348,771	575,486
1963	68,995	105,463	1983	370,501	591,858
1964	73,682	111,504	1984	411,427	651,527

1965	80,283	120,929	1985	423,940	66.5,837
1966	87,187	136,824	1986	431,786	664,654
1967	90,913	145,681	1987	459,107	711,745
1968	98,794	156,611	1988	496,334	767,387
1969	105,812	170,400	1989	522,344	813,018
			1990	540,788	835,985
			1991	533,838	828,184

Nguồn: Báo cáo Kinh tế của Tổng thống, 1993, Bảng B-53, trang 408.

* Dữ liệu hàng năm là các trung bình con số hàng tháng không hiệu chỉnh theo mùa.
+ Số liệu đã được hiệu chỉnh theo mùa, vào cuối giai đoạn bắt đầu từ 1982 không tương thích với thời đoạn sớm hơn.

(e) Dựa trên các kết quả của kiểm định này, bạn sẽ biến đổi dữ liệu thế nào để loại bỏ tự tương quan? Hãy trình bày mọi tính toán của bạn.

(f) Hãy kiểm định mô hình của bạn đối với tác động ARCH. Nếu tác động ARCH được quan sát thấy, liệu bạn có cải biến các kết luận của bạn về tự tương quan đã đạt trước đó?

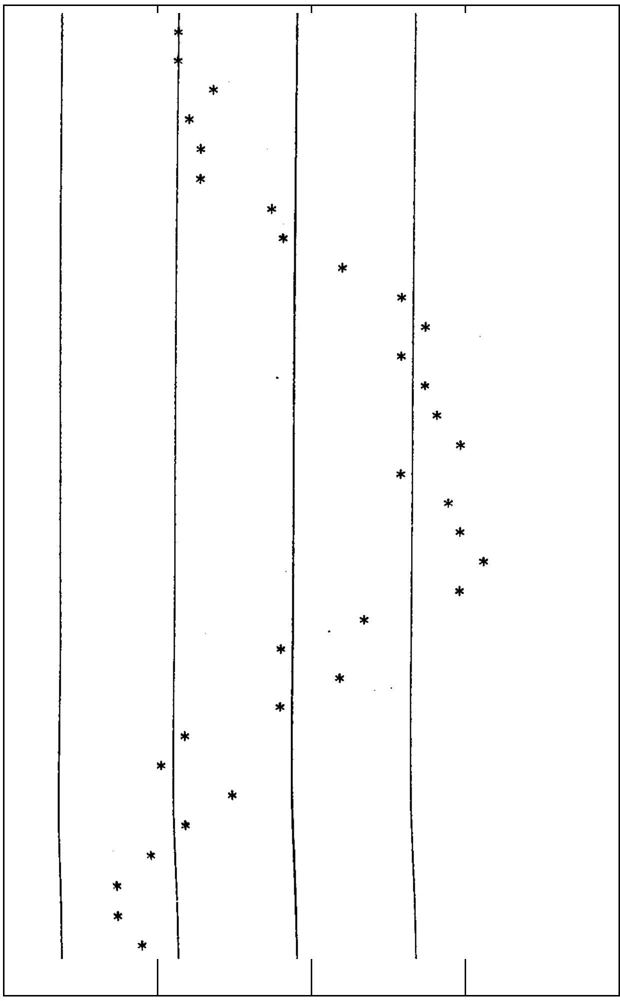
(g) Hãy nhắc lại các bước trước đây, sử dụng mô hình sau:

$$\ln Y_t = \beta_1 + \beta_2 \ln X_t + u_t$$

(g) Bạn sẽ quyết định thế nào giữa các đặc trưng mô hình tuyến tính và phi tuyến tính? Hãy trình bày rõ ràng (các) kiểm định mà bạn sử dụng.

12A.1 SẢN LƯỢNG TSP CỦA TIỀN CÔNG (Y)- NĂNG SUẤT (X) TẠI HOA KỲ, 1960–1961*

Đồ thị phần dư



Quan sát	PHẦN DƯ	THỰC	THÍCH HỢP
1960	-2.40999	68.700	71.1100
1961	-2.43360	70.700	73,1336
1962	-1.87626	73.200	75.0763
1963	-2.34270	75.000	77.3427
1964	-2.03292	77.900	79.9329
1965	-2.03275	79.600	81.6327
1966	-0.51352	82.900	83.4135
1967	-0,13240	84.900	85.0324
1968	1.06304	88.200	87.1370
1969	2.23926	89.700	87.4607
1970	2.76793	91.200	88.4321
1971	2.22055	93.000	90.7795
1972	2.75411	95.800	93.0459
1973	3.01145	98.000	94.9886
1974	3.46845	97.000	93.5316
1975	2.38767	97.700	95.3123
1976	3.22124	100.800	97.5788
1977	3.42612	102.300	98.8739
1978	4.04046	103.400	99.3595
1979	3.53084	102.000	98.4692
1980	1.59745	99.500	97.9025
1981	-0.25483	98.700	98.9548
1982	0.96423	100.000	99.0358
1983	-0,15465	100.500	100.6550
1984	-2.35920	100.400	102.7590
1985	-2.67336	101.300	103.9730
1986	-1.35414	104.400	105.7540
1987	-2.34453	104.300	106.6450
1988	-3.05397	104.400	107.4540
1989	-3.72547	103.000	106.7250
1990	-3.68736	103.200	106.8870
1991	-3.31114	103.900	107.2110

□ Dữ liệu thực được trình bày trong bảng đi kèm.

Chỉ số đền bù thực mỗi giờ (Y)

OLSQ // Dependent Variable is Y
SMPL range: 1960–1991
Number of observations: 32

VARIABLE	COEFFICIENT	STD. ERROR	t-STATISTIC	2-TAIL SIG.
C	18.091487	3.3106307	5.4646648	0.0000
X	0.8094428	0.0351369	23.036851	0.0000

R-squared	0.946495	Mean of dependent var	93.61250
Adjusted R-squared	0.944712	S.D. of dependent var	11.10898
S.E. of regression	2.612109	Sum of squared resid	204.69340
Log likelihood	−75.098470	F-statistic	530.69650
Durbin-Watson stat	0.138039	Prob(F-statistic)	0.00000

SMPL range: 1960–1991
Number of observations: 32

VARIABLE	MEAN	STANDARD DEVIATION	MAXIMUM	MINIMUM
RESID	−1.537E−08	2.5696328	4.0404560	−3.7254730

INTERVAL	COUNT	HISTOGRAM
−4.0 ≥ RESID < −3.5	2	***********
−3.5 ≥ RESID < −3.0	2	***********
−3.0 ≥ RESID < −2.5	1	******
−2.5 ≥ RESID < −2.0	7	**************************************
−2.0 ≥ RESID < −1.5	1	******
−1.5 ≥ RESID < −1.0	1	******
−1.0 ≥ RESID < −0.5	1	******
−0.5 ≥ RESID < 0.0	3	*****************
0.0 ≥ RESID < 0.5	0	
0.5 ≥ RESID < 1.0	1	******
1.0 ≥ RESID < 1.5	1	******
1.5 ≥ RESID < 2.0	1	******
2.0 ≥ RESID < 2.5	3	*****************
2.5 ≥ RESID < 3.0	2	***********
3.0 ≥ RESID < 3.5	4	**********************
3.5 ≥ RESID < 4.0	1	******
4.0 ≥ RESID < 4.5	1	******

Skewness 0.109606 Kurtosis 1.440579
Jarque-Bera normality test statistic 3.306464 Probability 0.191430

và chỉ số sản lượng mỗi giờ (X), khu vực doanh nghiệp, Hoa Kỳ, 1960-1961,1982=100

Quan sát	Y	X
1960	68.7000	65.5000
1961	70.7000	68.0000
1962	73.2000	70.4000
1963	75.0000	73.2000
1964	77.9000	76.4000

1965	79.6000	78.5000
1966	82.9000	80.7000
1967	84.9000	82.7000
1968	88.2000	8.5.3000
1969	89.7000	85.7000
1970	91.2000	86.9000
1971	93.0000	89.8000
1972	95.8000	92.6000
1973	98.0000	95.0000
1974	97.0000	93.2000
1975	97.7000	95.4000
1976	100.8000	98.2000
1977	102.3000	99.8000
1978	103.4000	100.4000
1979	102.0000	99.3000
1980	99.5000	98.6000
1981	98.7000	99.9000
1982	100.0000	100.0000
1983	100.5000	102.0000
1984	100.4000	104.6000
1985	101.3000	106.1000
1986	104.4000	108.3000
1987	104.3000	109.4000
1988	104.4000	1 10.4000
1989	103.0000	109.5000
1990	103.200(1	109.7000
1991	103.9000	1 10.1000

Nguồn: Báo cáo Kinh tế của Tổng thống, 1993, Bảng B–44, trang 398.