

Phần III

CÁC CHỦ ĐỀ TRONG KINH TẾ LƯỢNG

Trong Phần I ta đã giới thiệu mô hình hồi quy tuyến tính cổ điển với tất cả các giả thiết của nó. Trong Phần II, ta xem xét chi tiết các hậu quả xảy ra khi một hay nhiều *giả thiết không* được thỏa mãn và làm thế nào để khắc phục. Trong Phần III, ta sẽ chuyển sang nghiên cứu một số kỹ thuật kinh tế lượng có chọn lựa nhưng thường gặp phải.

Trong Chương 15, ta xem xét vai trò của các biến giải thích định tính trong phân tích hồi quy. Các biến định tính, gọi là **biến giả** (dummy variables) là công cụ để đưa vào mô hình hồi quy những biến mà không thể lượng hóa ngay được, ví dụ như giới tính, tôn giáo, màu da, nhưng lại tác động tới hành vi của biến phụ thuộc. Bằng một số ví dụ, ta sẽ chỉ ra rằng các biến này có thể tăng cường phạm vi của mô hình hồi quy tuyến tính như thế nào.

Trong Chương 16, ta cho phép biến phụ thuộc trong một mô hình hồi quy là biến định tính về bản chất. Những mô hình như vậy được sử dụng trong các trường hợp mà biến phụ thuộc có phạm trù “có” hoặc “không”, như sở hữu nhà, xe hơi, và các vật dụng gia đình hay có một thuộc tính như thành viên của công đoàn hay một hiệp hội chuyên môn. Các mô hình trong đó bao gồm các biến phụ thuộc có dạng có - không được gọi là các **mô hình hồi quy có biến phụ thuộc phân đôi, hay biến phụ thuộc giả**. Ta xem xét ba phương pháp để ước lượng các mô hình dạng này: (1) mô hình xác suất tuyến tính (**LPM**), (2) **mô hình logit**, và (3) **mô hình probit** (đơn vị xác suất). Trong số các mô hình này, LPM, mặc dù dễ tính toán, lại không thỏa đáng nhất vì nó vi phạm một số giả thiết OLS. Vì vậy, logit và probit là các mô hình thường được sử dụng nhiều nhất khi biến phụ thuộc có dạng phân đôi. Ta minh họa các mô hình này với một số ví dụ bằng số và ví dụ thực tế.

Ta cũng xem xét **mô hình tobit**, một mô hình có quan hệ với probit. Trong mô hình probit, ví dụ, ta cố gắng tìm xác suất sở hữu một ngôi nhà. Trong mô hình tobit, ta muốn tìm lượng tiền mà một người tiêu dùng sử dụng để mua một ngôi nhà trong quan hệ với thu nhập, v.v... Nhưng tất nhiên, nếu một người tiêu dùng không mua nhà, ta không có số liệu về chi tiêu cho nhà ở của những người tiêu dùng đó; thông tin này chỉ có đối với những người tiêu dùng mua nhà thực sự. Như vậy, ta có một **mẫu kiểm duyệt** (censored sample), tức là, một mẫu mà trong đó thông tin về biến phụ thuộc không có cho một số quan sát, mặc dù thông tin về các biến làm hồi quy lại có. Mô hình tobit mô tả làm thế nào ta có thể ước lượng các mô hình hồi quy có các mẫu kiểm duyệt.

Trong Chương 17, ta xem xét các mô hình hồi quy với các biến giải thích có giá trị hiện tại, quá khứ, hay trễ cùng với các mô hình trong đó đưa các giá trị trễ của biến phụ thuộc thành một trong các biến giải thích. Các mô hình này được gọi là tương ứng là mô hình **trễ phân phối** và **tự tương quan**. Mặc dù các mô hình dạng này vô cùng hữu ích trong kinh lượng thực nghiệm, chúng tạo ra một số khó khăn đặc biệt trong ước lượng. Ta sẽ xem xét các vấn đề khó khăn đặc biệt này trong bối cảnh của mô hình Koyck, kỳ vọng *thích nghi* (adaptive expectations - AE), và mô hình điều chỉnh riêng phần. Ta cũng lưu ý tới các chỉ trích về mô hình AE của những người ủng hộ cái gọi là trường phái kỳ vọng hợp lý (rational expectations - RE).

Với Chương 17, ta kết thúc thảo luận về mô hình hồi quy đơn phương trình mà ta đã bắt đầu từ Chương 1. 17 chương này bao trùm rất nhiều cơ sở trong các mô hình kinh tế lượng đơn phương trình nhưng không hề đề cập hết tất cả các vấn đề. Đặc biệt là ta đã không thảo luận các kỹ thuật ước lượng các tham số phi tuyến và cũng không xem xét phương pháp Bayes trong các

mô hình hồi quy đơn phương trình, tuyến tính hay phi tuyến. Nhưng trong một cuốn sách mang tính giới thiệu như thế này, hoàn toàn không thể phân tích rõ các chủ đề này bởi vì chúng đòi hỏi các cơ sở về toán và thống kê vượt ra ngoài phạm vi dự định của cuốn sách.

Chương 15

HỒI QUY THEO CÁC BIẾN GIẢ

Mục đích của chương này là xem xét vai trò của các biến giả thích định tính trong phân tích hồi quy. Ta sẽ chỉ ra rằng việc đưa ra các biến định tính, thường được gọi là **biến giả**, làm cho mô hình hồi quy tuyến tính trở thành một công cụ vô cùng linh hoạt, có khả năng giải quyết các vấn đề thú vị thường gặp trong nghiên cứu thực nghiệm.

15.1 BẢN CHẤT CỦA CÁC BIẾN GIẢ

Trong phân tích hồi quy, biến phụ thuộc thường bị tác động không chỉ bởi các biến có thể lượng hóa được ngay theo tỷ lệ đã xác định (ví dụ như thu nhập, sản lượng, giá cả, chi phí, chiều cao và nhiệt độ), mà còn bởi các biến có bản chất định tính (như giới tính, chủng tộc, màu da, tôn giáo, quốc tịch, chiến tranh, động đất, đình công, bất ổn chính trị và thay đổi chính sách kinh tế của chính phủ). Ví dụ, giữ tất cả các nhân tố khác không đổi, người ta nhận thấy các giáo sư nữ dạy đại học có thu nhập ít hơn các giáo sư nam, và những người không phải da trắng có thu nhập thấp hơn những người da trắng. Hình thái này có thể nảy sinh từ sự phân biệt giới tính hay chủng tộc. Nhưng vì lý do gì đi nữa thì các biến định tính như giới tính và chủng tộc rõ ràng có tác động tới biến phụ thuộc và phải được đưa vào mô hình làm biến giả thích.

Do các biến định tính như vậy thường mô tả sự xuất hiện hay thiếu vắng một “tính chất” hay đặc điểm, như nam hay nữ, đen hay trắng, theo công giáo hay không theo công giáo, phương pháp “lượng hóa” các thuộc tính như vậy là thiết lập các biến nhân tạo với giá trị 1 biểu thị xuất hiện (hay có) thuộc tính đó. Ví dụ, 1 có thể biểu thị rằng một người là nam, và 0 có thể biểu thị một người là nữ; hay 1 có thể biểu thị một người đã tốt nghiệp đại học, và 0 biểu thị người đó chưa tốt nghiệp, và v.v... Các biến nhận các giá trị 0 và 1 được gọi là các **biến giả**.¹ Các tên gọi khác là biến chỉ định (indicator variables), biến nhị phân (binary variables), biến phân loại hay biến phạm trù (category variable), biến định tính (qualitative variables) và biến phân đôi (dichotomous variables).

Các biến giả có thể được sử dụng trong các mô hình hồi quy một cách dễ dàng như các biến định lượng. Trên thực tế, một mô hình hồi quy có thể gồm các biến giả thích hoàn toàn là biến giả, hay định tính, về bản chất. Các mô hình như thế được gọi là các **mô hình phân tích phương sai (ANOVA)**. Hãy lấy mô hình sau làm ví dụ xem xét:

$$Y_i = \alpha + \beta D_i + u_i \quad (15.1.1)$$

với Y = mức lương hàng năm của một giáo sư đại học

D_i = 1 nếu là giáo sư nam

= 0 nếu khác (nghĩa là giáo sư nữ).

Lưu ý rằng (15.1.1) giống các mô hình hồi quy hai biến gặp phải trước đây ngoại trừ thay cho biến định lượng X , ta có một biến giả D (sau đây ta ký hiệu tất cả các biến giả bằng ký tự D).

¹ Không hoàn toàn nhất thiết là các biến giả phải lấy các giá trị 0 và 1. Cặp (0, 1) có thể được biến đổi thành mọi cặp khác bằng một hàm tuyến tính như $Z = a + bD$ ($b \neq 0$), với a và b là các hằng số và với $D = 1$ hay 0. Khi $D = 1$, ta có $Z = a + b$; và khi $D = 0$, ta có $Z = a$. Vậy, cặp (0, 1) có thể trở thành $(a, a + b)$. Ví dụ, nếu $a = 1$ và $b = 2$, các biến giả sẽ là (1, 3). **Biểu thức này cho thấy các biến định tính không có một tỷ lệ đo tự nhiên.**

Mô hình (15.1.1) có thể cho phép ta tìm xem giới tính có tạo ra khác biệt trong mức lương của giáo sư đại học hay không, tất nhiên là với giả thiết rằng tất cả các biến khác như tuổi, học vị, và năm kinh nghiệm được giữ không đổi. Giả sử rằng các yếu tố nhiễu thỏa mãn giả thiết của mô hình hồi quy cổ điển, từ (15.1.1) ta có:

$$\text{Mức lương trung bình của giáo sư đại học nữ: } E(Y_i | D_i = 0) = \alpha \quad (15.1.2)$$

$$\text{Mức lương trung bình của giáo sư đại học nam: } E(Y_i | D_i = 1) = \alpha + \beta$$

tức là, tung độ gốc α cho ta mức lương trung bình của các giáo sư đại học nữ và hệ số góc β cho ta biết mức lương trung bình của một giáo sư đại học nam khác bao nhiêu so với mức lương trung bình của một giáo sư đại học nữ, $\alpha + \beta$ biểu thị mức lương trung bình của giáo sư đại học nam.

Một kiểm định *giả thiết không* cho rằng không có phân biệt giới tính ($H_0: \beta = 0$) có thể được dễ dàng thực hiện bằng cách chạy hồi quy (15.1.1) theo cách thông thường và tìm xem trên cơ sở của kiểm định t , giá trị ước lượng của β có ý nghĩa thống kê hay không.

Ví dụ 15.1 Lương giáo sư theo giới tính

Bảng 15.1 biểu thị số liệu giả thiết về các mức lương khởi điểm của 10 giáo sư đại học theo giới tính. Sau đây là các kết quả tương ứng với hồi quy (15.1.1):

$$\begin{aligned} Y_i &= 18,00 + 3,28D_i \\ &\quad (0,32) \quad (0,44) \\ t &= (57,74) \quad (7,439) \quad R^2 = 0,8737 \end{aligned} \quad (15.1.3)$$

BẢNG 15.1

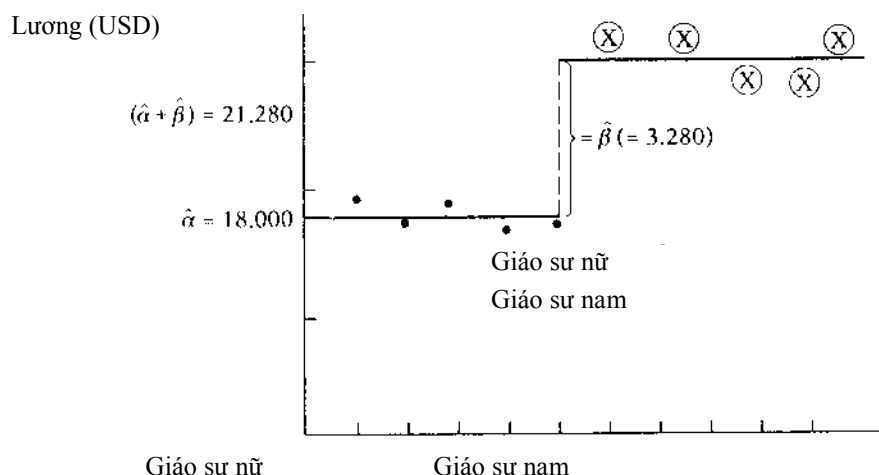
Số liệu giả thiết về mức lương khởi điểm của các giáo sư đại học theo giới tính

Lương khởi điểm, Y (nghìn USD)	Giới tính (1 = nam, 0 = nữ)
22,0	1
19,0	0
18,0	0
21,7	1
18,5	0
21,0	1
20,5	1
17,0	0
17,5	0
21,2	1

Như các kết quả biểu thị, mức lương trung bình ước lượng của các giáo sư đại học nữ là 18.000 USD ($= \alpha$) và của các giáo sư nam là 21.2800 USD ($\alpha + \beta$); từ số liệu trong Bảng 15.1 ta có thể tính ngay được các mức lương của giáo sư đại học nữ và nam, tương ứng là 18.000 và 21.800 USD, bằng chính xác với các giá trị ước lượng.

Do β có ý nghĩa thống kê, các kết quả chỉ ra rằng các mức lương trung bình của hai nhóm khác nhau; thực tế, lương trung bình của giáo sư nữ thấp hơn giáo sư nam. Nếu tất cả các biến được giữ không đổi (một chữ nếu không thực tế lắm), rất có thể là có phân biệt giới

tính trong các mức lương của hai giới. Tất nhiên, mô hình đang xem xét quá đơn giản nên không thể trả lời được câu hỏi này một cách xác đáng, đặc biệt là trên cơ sở của tính chất giả thiết của số liệu sử dụng trong phân tích.



HÌNH 15.1

Hàm số mức lương của giáo sư nữ và nam.

Nhân đây, hãy xem xét hồi quy (15.1.3) trên đồ thị Hình 15.1. Trong hình này, số liệu được xếp thứ tự để nhóm chúng thành hai nhóm, giáo sư nam và nữ. Như bạn có thể thấy từ hình vẽ, hàm hồi quy tính được là một *hàm bậc thang* – lương trung bình của giáo sư nữ là 18.000 USD và của giáo sư nam nhảy cách 3.280 USD ($= \hat{\beta}_2$) lên 21.280 USD; mức lương của các giáo sư riêng lẻ trong hai nhóm nằm rải rác xung quanh các giá trị mức lương trung bình tương ứng.

Các mô hình ANOVA theo kiểu (15.1.1), mặc dù thông dụng trong các lĩnh vực như xã hội học, tâm lý học, giáo dục, và nghiên cứu thị trường, lại không phổ biến trong kinh tế học. Thường thì trong phần lớn các nghiên cứu kinh tế, mô hình hồi quy chứa một số biến giải thích định lượng và một số biến định tính. Các mô hình hồi quy chứa hỗn hợp các biến định lượng và định tính được gọi là các **mô hình phân tích tích sai (Analysis of Covariance, ANCOVA)**, và trong chương này, ta sẽ chủ yếu phân tích các mô hình này.

15.2 HỒI QUY THEO MỘT BIẾN ĐỊNH LƯỢNG VÀ MỘT BIẾN ĐỊNH TÍNH CÓ HAI LOẠI HAY HAI PHẠM TRÙ

Để đưa ra ví dụ cho mô hình ANCOVA, hãy biến đổi mô hình (15.1.1) như sau:

$$Y_i = \alpha_1 + \alpha_2 D_i + \beta X_i + u_i \quad (15.2.1)$$

với Y_i = lương trung bình của một giáo sư đại học

X_i = số năm kinh nghiệm giảng dạy

D_i = 1 nếu là nam

= 0 nếu khác

Mô hình (15.2.1) chứa một biến định lượng (số năm kinh nghiệm giảng dạy) và một biến định tính (giới tính) có hai lớp (hay cấp, phân loại, hay phạm trù), cụ thể là nam và nữ.

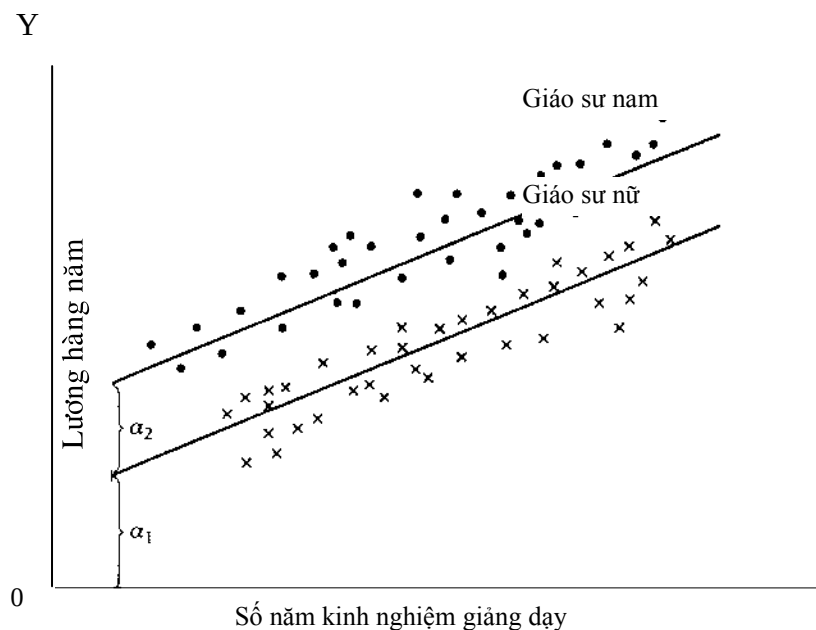
Ý nghĩa của (15.2.1) là gì? Theo thông lệ, giả sử rằng $E(u_i) = 0$, ta có thể thấy rằng
Mức lương trung bình một giáo sư đại học nữ:

$$E(Y_i | X_i, D_i = 0) = \alpha_1 + \beta X_i \quad (15.2.2)$$

Mức lương của một giáo sư nam:

$$E(Y_i | X_i, D_i = 1) = (\alpha_1 + \alpha_2) + \beta X_i \quad (15.2.3)$$

Về hình học, ta có trường hợp như trong Hình 15.2 (để minh họa, ta giả sử rằng $\alpha_1 > 0$). Diễn đạt bằng lời, mô hình (15.2.1) mặc định rằng các hàm số mức lương của các giáo sư đại học nam và nữ trong quan hệ với số năm kinh nghiệm dạy học có cùng độ dốc (β) nhưng tung độ gốc khác nhau. Nói một cách khác, ta giả sử rằng mức lương trung bình của giáo sư nam khác với giáo sư nữ (là α_2) nhưng tốc độ thay đổi mức lương trung bình hàng năm theo số năm kinh nghiệm giống nhau ở cả hai giới.



HÌNH 15.2

Đồ thị phân tán giả thiết giữa mức lương hàng năm và số năm kinh nghiệm giảng dạy của các giáo sư đại học.

Nếu giả thiết về độ dốc chung có hiệu lực,² một kiểm định giả thiết cho rằng hai hồi quy (15.2.2) và (15.2.3) có cùng tung độ gốc (nghĩa là không có phân biệt giới tính) có thể được thực hiện dễ dàng bằng cách chạy hồi quy (15.2.1) và kiểm định ý nghĩa thống kê của giá trị ước lượng của α_2 trên cơ sở của kiểm định truyền thống t . Nếu kiểm định t cho thấy α_2 có ý nghĩa thống kê, ta bác bỏ *giả thiết không* cho rằng các mức lương trung bình của giáo sư đại học nam và nữ là như nhau.

² Giá trị của giả thiết này có thể được kiểm định bằng các thủ tục tóm lược trong Mục 15.7.

Trước khi phân tích sâu hơn, hãy lưu ý các đặc điểm sau đây của mô hình hồi quy có biến giả xem xét ở trên:

- Để phân biệt giữa hai phạm trù, nam và nữ, ta chỉ đưa ra một biến giả D_i . Bởi vì nếu $D_i = 1$ luôn luôn biểu thị nam, khi $D_i = 0$ ta biết rằng đó là nữ do chỉ có hai kết quả có thể xảy ra. Vậy, một biến giả là đủ để phân biệt hai phạm trù. Hãy giả thiết rằng mô hình hồi quy có tung độ gốc; nếu ta phải viết mô hình (15.2.1) dưới dạng

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + u_i \quad (15.2.4)$$

với Y_i và X_i được định nghĩa như trước

$$\begin{aligned} D_{2i} &= 1 && \text{nếu là nam giáo sư} \\ &= 0 && \text{nếu khác} \\ D_{3i} &= 1 && \text{nếu là nữ giáo sư} \\ &= 0 && \text{nếu khác} \end{aligned}$$

thì mô hình (15.2.4) không thể ước lượng được bởi vì có đa cộng tuyến hoàn hảo giữa D_2 và D_3 . Để xem xét vấn đề này, giả sử ta có một mẫu ba giáo sư nam và hai giáo sư nữ. Ma trận số liệu sẽ có dạng như sau:

		D_2	D_3	X	
Nam	Y_1	1	1	0	X_1
Nam	Y_2	1	1	0	X_2
Nữ	Y_3	1	0	1	X_3
Nam	Y_4	1	1	0	X_4
Nữ	Y_5	1	0	1	X_5

Cột thứ nhất ở bên phải của ma trận số liệu trên đại diện cho tung độ gốc α_1 . Bây giờ, ta có thể thấy ngay rằng $D_2 = 1 - D_3$ hay $D_3 = 1 - D_2$; tức là, D_2 và D_3 có đa cộng tuyến hoàn hảo. Và như đã chỉ ra trong Chương 10, trong các trường hợp đa cộng tuyến hoàn hảo, ta không thể thực hiện ước lượng OLS thông thường. Có nhiều cách khác nhau để giải quyết vấn đề này, nhưng cách đơn giản nhất là đưa ra các biến giả như ta đã làm trong mô hình (15.2.1), cụ thể là chỉ sử dụng một biến giả nếu có hai cấp hay hai loại của biến định tính. Trong trường hợp này, ma trận số liệu ở trên sẽ không có cột D_3 , như vậy loại bỏ được vấn đề đa cộng tuyến hoàn hảo. Quy tắc tổng quát là: **Nếu một biến giả có m phạm trù thì chỉ đưa ra $m - 1$ biến giả.** Trong ví dụ của chúng ta, giới tính có hai phạm trù, và do vậy ta chỉ đưa ra một biến giả. Nếu quy tắc này không được tuân thủ, ta sẽ rơi vào cái gọi là **bẫy biến giả**, tức là, trường hợp đa cộng tuyến hoàn hảo. (Về thảo luận thêm, xem Mục 15.13).

- Việc gán các giá trị 1 và 0 cho hai phạm trù, như nam và nữ, là tùy ý trên khía cạnh là trong ví dụ hiện tại, ta có thể cho $D = 1$ biểu thị nữ và $D = 0$ biểu thị nam. Trong trường hợp này, hai hồi quy tính được ở (15.2.1) sẽ là

$$\text{Giáo sư nữ: } E(Y_i | X_i, D_i = 1) = (\alpha_1 + \alpha_2) + \beta X_i \quad (15.2.5)$$

$$\text{Giáo sư nam: } E(Y_i | X_i, D_i = 0) = \alpha_1 + \beta X_i \quad (15.2.6)$$

Tương phản với (15.2.2) và (15.2.3) trong các mô hình trước, α_2 cho biết sự khác biệt giữa mức lương của một giáo sư đại học nữ và mức lương của giáo sư đại học nam: trong trường hợp này, nếu có phân biệt giới tính, α_2 sẽ được dự kiến là âm ngược lại với trước đây nó được dự kiến là dương. **Do vậy, trong việc giải thích các kết quả của các mô hình sử dụng biến giả, điều then chốt là phải biết được các giá trị 1 và 0 được gán cho như thế nào.**

3. Nhóm, phạm trù hay phân loại được gán cho giá trị 0 thường được coi là phạm trù *cơ sở*, *mốc*, *kiểm soát*, *so sánh*, *tham chiếu* hay *loại bỏ*. Nó là cơ sở xét trên khía cạnh là ta thực hiện các so sánh với phạm trù đó. Vậy, trong mô hình (15.2.1), giáo sư nữ là phạm trù cơ sở. Lưu ý rằng tung độ gốc (chung) α_1 là tung độ gốc cho phạm trù cơ sở xét trên khía cạnh là nếu ta chạy hồi quy với $D = 0$, tức là, chỉ có giáo sư nữ, tung độ gốc sẽ là α_1 . Cũng cần lưu ý rằng việc phạm trù nào đóng vai trò phạm trù cơ sở là vấn đề lựa chọn, đôi khi được xác định bởi các nghiên cứu tiên nghiệm.
4. Hệ số α_2 gắn với biến giả D có thể được gọi là **hệ số tung độ gốc chênh lệch** do nó cho biết giá trị của tung độ gốc của phạm trù nhận giá trị 1 khác với hệ số tung độ gốc của phạm trù cơ sở là bao nhiêu.

Ví dụ 15.2 Hàng tồn kho có nhạy cảm với lãi suất không?

Dan M. Bechter và Stephen H. Pollock đã ước lượng mô hình sau để giải thích các biến động hàng tồn kho trong ngành thương nghiệp bán sỉ của nền kinh tế Hoa Kỳ trong giai đoạn 1967-IV đến 1979-IV (các tỷ số t ở trong ngoặc):³

$$\begin{aligned} I/S = & 1,269 - 0,3615C + 0,0215S^e - 0,0227S \\ & (19,6) \quad (-2,2) \quad (5,7) \quad (-2,4) \\ & -0,2552U + 0,0734DUM \\ & (-2,4) \quad (4,8) \end{aligned} \quad R^2 = 0,71 \quad d = 1,91$$

với I/S = hàng tồn kho tính theo USD cố định chia cho doanh thu tính theo USD cố định, C = mức lãi suất cơ bản của giấy nợ thương mại từ 4 đến 6 tháng trừ đi tỷ lệ thay đổi chỉ số giá sản xuất so với năm trước đối với hàng tiêu dùng cuối cùng, S^e = doanh thu kỳ vọng trong giai đoạn hiện hành, với doanh thu kỳ vọng bằng doanh thu xu hướng có hiệu chỉnh độ lệch khỏi xu hướng trong năm trước, tất cả đều tính theo USD cố định, U = tính không chắc chắn trong doanh thu tính bằng độ biến thiên của doanh thu xung quanh xu hướng, và DUM = biến giả, nhận giá trị 0 trong giai đoạn từ 1967-IV đến 1974-I và giá trị 1 trong giai đoạn từ 1974-II đến 1979-IV.

Mặc dù tất cả các hệ số đều có ý nghĩa thống kê và có dấu như mong đợi, trong thảo luận hiện tại, ta sẽ tập trung vào biến giả. Các kết quả cho thấy tỷ lệ hàng tồn kho/ doanh thu cao hơn ($= 1,269 + 0,0734$) trong giai đoạn sau suy thoái năm 1974 so với giai đoạn trước. Vậy, đường hồi quy, thực tế là mặt phẳng, trong giai đoạn sau song song nhưng nằm ở vị trí cao hơn so với giai đoạn trước (đối chiếu Hình 15.2). Các tác giả không thảo luận lý do tại sao nhưng hiện tượng này có thể phản ánh tính trầm trọng của suy thoái 1974.

15.3 HỒI QUY THEO MỘT BIẾN ĐỊNH LƯỢNG VÀ MỘT BIẾN ĐỊNH TÍNH CÓ NHIỀU PHẠM TRÙ

Giả sử rằng, trên cơ sở của số liệu chéo, ta muốn thực hiện hồi quy chi tiêu y tế hàng năm của một cá nhân theo thu nhập và trình độ học vấn của cá nhân đó. Do biến trình độ học vấn là biến định tính về bản chất, giả sử ta xem xét ba cấp loại trừ lẫn nhau của trình độ học vấn: dưới trung học, trung học và đại học. Bây giờ, không giống như trường hợp trước, ta có nhiều hơn hai phạm trù của biến định tính về giáo dục. Do vậy, **theo quy tắc số biến giả phải bằng số phạm**

³ “Are Inventories Sensitive to Interest Rates?”, *Economic Review* (Hàng tồn kho có nhạy cảm với lãi suất không? Tạp chí Kinh tế, Ngân hàng Dự trữ Liên bang Kansas, 4/1980, trang 24 (Bảng 2). Lưu ý: Các kết quả được hiệu chỉnh tự tương quan bậc 2; giá trị d ban đầu là 1,12.

trừ của biến trừ đi 1, ta phải đưa ra hai biến giả để giải quyết ba cấp của trình độ học vấn. Giả sử rằng ba nhóm trình độ học vấn có cùng độ dốc nhưng tung độ gốc khác nhau trong hồi quy chi tiêu y tế hàng năm theo thu nhập hàng năm, ta có thể sử dụng mô hình sau:

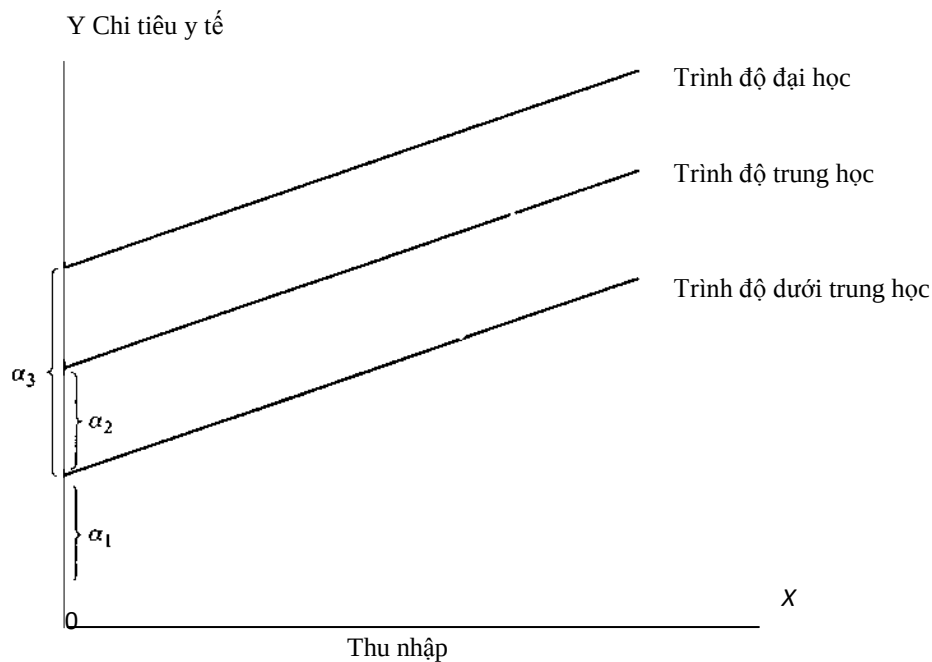
$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + u_i \quad (15.3.1)$$

với Y_i = chi tiêu y tế hàng năm

X_i = thu nhập hàng năm

$D_2 = 1$ nếu có trình độ trung học
 $= 0$ nếu có trình độ khác

$D_3 = 1$ nếu có trình độ đại học
 $= 0$ nếu có trình độ khác



HÌNH 15.3

Chi tiêu y tế trong quan hệ với thu nhập và ba cấp trình độ học vấn.

Lưu ý rằng trong việc gán giá trị của các biến giả ở trên, ta tùy ý coi phạm trù “dưới trung học” là phạm trù cơ sở. Do vậy, tung độ gốc α_1 sẽ biểu thị tung độ gốc của phạm trù này. Các tung độ gốc chênh lệch α_2 và α_3 cho biết các tung độ gốc của hai phạm trù kia khác với tung độ gốc của phạm trù cơ sở là bao nhiêu. Điều này có thể được kiểm tra ngay như sau: Giả sử $E(u_i) = 0$, từ (15.3.1) ta có

$$E(Y_i | D_2 = 0, D_3 = 0, X_i) = \alpha_1 + \beta X_i \quad (15.3.2)$$

$$E(Y_i | D_2 = 1, D_3 = 0, X_i) = (\alpha_1 + \alpha_2) + \beta X_i \quad (15.3.3)$$

$$E(Y_i | D_2 = 0, D_3 = 1, X_i) = (\alpha_1 + \alpha_3) + \beta X_i \quad (15.3.4)$$

Đây là các hàm số chỉ tiêu y tế trung bình tương ứng cho 3 cấp trình độ học vấn, cụ thể là dưới trung học, trung học và đại học. Trên đồ thị, trường hợp này được mô tả trong Hình 15.3 (để minh họa, giả sử rằng $\alpha_3 > \alpha_2$).

Sau khi chạy hồi quy (15.3.1), ta có thể dễ dàng tìm xem từng tung độ gốc chênh lệch α_2 và α_3 có ý nghĩa thống kê hay không, tức là, khác với nhóm cơ sở. Một kiểm định giả thiết rằng $\alpha_2 = \alpha_3$ và cùng đồng thời bằng 0 cũng có thể được thực hiện bằng kỹ thuật ANOVA và kiểm định F kèm theo, như được chỉ ra trong Chương 8 [xem Phương trình (8.7.9)].

Trước khi chuyển sang phần kế tiếp, lưu ý rằng việc giải thích hồi quy (15.3.1) sẽ thay đổi nếu ta áp dụng cách gán giá trị các biến giả theo kiểu khác. Vậy, nếu ta cho $D_2 = 1$ biểu thị “phạm trù dưới trung học” và $D_3 = 1$ biểu thị “phạm trù trung học”, phạm trù tham chiếu sẽ là “đại học” và tất cả các so sánh sẽ đặt trong quan hệ với phạm trù này.

15.4 HỒI QUY THEO MỘT BIẾN ĐỊNH LƯỢNG VÀ HAI BIẾN ĐỊNH TÍNH

Kỹ thuật biến giả có thể được dễ dàng mở rộng để giải quyết mô hình có nhiều biến định tính. Hãy quay lại với hồi quy lương giáo sư đại học (15.2.1), nhưng bây giờ giả thiết rằng ngoài số năm kinh nghiệm giảng dạy và giới tính, màu da của giáo viên cũng là một yếu tố quan trọng trong việc xác định mức lương. Để đơn giản, giả sử màu da có hai phạm trù: đen và trắng. Bây giờ, ta có thể viết (15.2.1) dưới dạng:

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + u_i \quad (15.4.1)$$

với Y_i = lương hàng năm

X_i = số năm kinh nghiệm giảng dạy

$D_2 = 1$ nếu là nam

= 0 nếu khác

$D_3 = 1$ nếu là da trắng

= 0 nếu khác

Lưu ý rằng mỗi biến định tính, giới tính và màu da, có hai phạm trù và do vậy chỉ cần một biến giả cho mỗi biến định tính. Cũng lưu ý rằng phạm trù loại bỏ hay cơ sở bây giờ là “giáo sư nữ da đen”.

Giả sử $E(u_i) = 0$, ta có thể tính hàm hồi quy sau đây từ (15.4.1):

Mức lương trung bình một giáo sư nữ da đen:

$$E(Y_i | D_2 = 0, D_3 = 0, X_i) = \alpha_1 + \beta X_i \quad (15.4.2)$$

Mức lương trung bình của một giáo sư nam da đen:

$$E(Y_i | D_2 = 1, D_3 = 0, X_i) = (\alpha_1 + \alpha_2) + \beta X_i \quad (15.4.3)$$

Mức lương trung bình một giáo sư nữ da trắng:

$$E(Y_i | D_2 = 0, D_3 = 1, X_i) = (\alpha_1 + \alpha_3) + \beta X_i \quad (15.4.4)$$

Mức lương trung bình của một giáo sư nam da trắng:

$$E(Y_i | D_2 = 1, D_3 = 1, X_i) = (\alpha_1 + \alpha_2 + \alpha_3) + \beta X_i \quad (15.4.5)$$

Một lần nữa, ta lại giả sử rằng các hồi quy ở trên chỉ khác nhau ở hệ số tung độ gốc chứ không phải ở hệ số góc β .

Ước lượng OLS của (15.4.1) sẽ cho phép ta kiểm định các giả thiết. Như vậy, nếu α_3 có ý nghĩa thống kê thì nó có nghĩa là màu da có tác động tới mức lương của một giáo sư. Tương tự, nếu α_2 có ý nghĩa thống kê thì nó có nghĩa là giới tính cũng có tác động tới mức lương của một giáo sư. Nếu cả hai tung độ góc chênh lệch đều có ý nghĩa thống kê thì điều này có nghĩa là cả giới tính và màu da đều là các yếu tố quan trọng trong việc xác định mức lương giáo sư.

Thảo luận ở trên suy ra rằng ta có thể mở rộng mô hình cho nhiều hơn một biến định lượng và nhiều hơn hai biến định tính. **Điều duy nhất phải cần trọng là số các biến giả của mỗi biến định tính phải bằng số các phạm trù của biến đó trừ đi một.** Phần sau đây là một ví dụ minh họa.

15.5 VÍ DỤ 15.3. KINH TẾ HỌC CỦA VIỆC “LÀM THÊM NGOÀI GIỜ”

Một người có hai hay nhiều việc làm, một công việc chính và một hay nhiều công việc phụ được gọi là người *làm thêm ngoài giờ*. Shisko và Rostker đã quan tâm tới việc tìm xem các yếu tố nào xác định các mức lương của những người làm thêm ngoài giờ.⁴ Dựa vào một mẫu 318 người làm thêm ngoài giờ, họ tính được hồi quy sau, với ký hiệu của tác giả (các sai số chuẩn trong ngoặc):

$$w_m = 37,07 + 0,403w_0 - 90,06 \text{ race} + 75,51 \text{ urban} + 47,33 \text{ hisch} + 113,64 \text{ reg} + 2,26 \text{ age}$$

(0,062) (24,47) (21,60) (23,42) (27,62) (0,94)

$R^2 = 0,34$ bậc tự do = 311

(15.5.1)

với	w_m	=	lương làm thêm (xu/giờ)
	w_0	=	lương chính (xu/giờ)
race (chủng tộc)	=	0	nếu là da trắng
	=	1	nếu không phải da trắng
urban (thành thị)	=	0	không phải thành thị
	=	1	thành thị
reg (vùng)	=	0	không phải miền tây
	=	1	miền tây
hisch (tr.độ trung học)	=	0	chưa tốt nghiệp
	=	1	tốt nghiệp trung học
	age	=	tuổi (năm)

Trong mô hình (15.5.1) có hai biến giải thích định lượng, w_0 và tuổi với bốn biến định tính. Lưu ý rằng các hệ số của tất cả các biến này đều có ý nghĩa thống kê ở mức 5%. Điều thú vị là tất cả các biến định tính đều có tác động đáng kể tới các mức lương làm thêm. Ví dụ, giữ tất cả các nhân tố khác không đổi, mức lương theo giờ được dự kiến là cao hơn khoảng 47 xu đối với người tốt nghiệp trung học so với người có trình độ dưới trung học.

Từ hồi quy (15.5.1), ta có thể tính một số hồi quy riêng rẽ, sau đây là hai trong số đó: Mức lương trung bình theo giờ của người làm thêm da trắng, không ở thành thị, không ở miền tây, chưa tốt nghiệp trung học (nghĩa là khi tất cả các biến giả có giá trị bằng 0) là

$$w_m = 37,07 + 0,403w_0 + 2,26 \text{ age} \quad (15.5.2)$$

⁴ Robert Shisko & Bernard Rostker, “The Economics of Multiple Job Holding”, *The American Economic Review*, (Kinh tế học về hiện tượng làm nhiều việc, *Tạp chí Kinh tế Hoa Kỳ*), tập 66, số 3 6/1976, trang 298-308.

Mức lương trung bình theo giờ của người làm thêm không phải da trắng, sống ở thành thị, ở miền tây, đã tốt nghiệp trung học (nghĩa là khi tất cả các biến giả có giá trị bằng 1) là

$$w_m = 183,49 + 0,403w_0 + 2,26 \text{ age} \quad (15.5.3)$$

15.6 KIỂM ĐỊNH TÍNH ỔN ĐỊNH CẤU TRÚC CỦA CÁC MÔ HÌNH HỒI QUY

Cho tới nay, trong các mô hình xem xét ở chương này ta giả sử rằng các biến định tính tác động tới tung độ gốc nhưng không tác động tới hệ số góc của các hồi quy nhóm khác nhau. Nhưng nếu hệ số góc cũng khác nhau thì sao? Nếu các độ dốc khác nhau trên thực tế, kiểm định sự khác nhau của tung độ gốc có thể có ít ý nghĩa thực tiễn. Do vậy, ta cần phải xây dựng một phương pháp tổng quát để tìm xem hai (hay nhiều) hồi quy có khác nhau không, với sự khác nhau có thể ở tung độ gốc hay độ dốc hay cả hai. Để tìm hiểu vấn đề này được giải quyết như thế nào, hãy xem xét số liệu tiết kiệm - thu nhập của Anh quốc trong Bảng 8.8. Để thuận tiện, số liệu được trình bày lại trong Bảng 15.2.

Ví dụ 15.4 Tiết kiệm và thu nhập, Anh Quốc, 1946-1963

Như trình bày trong bảng, số liệu được chia làm hai giai đoạn, 1946-1954 (thời kỳ ngay sau Chiến tranh Thế giới thứ II, gọi là thời kỳ tái thiết) và 1955-1963 (thời kỳ hậu tái thiết). Giả sử ta muốn tìm xem nếu quan hệ tiết kiệm - thu nhập có thay đổi giữa hai thời kỳ không. Cụ thể, đặt

$$\text{Thời kỳ tái thiết: } Y_i = \lambda_1 + \lambda_2 X_i + u_{1i} \quad (15.6.1) \\ i = 1, 2, \dots, n_1$$

$$\text{Thời kỳ hậu tái thiết: } Y_i = \gamma_1 + \gamma_2 X_i + u_{2i} \quad (15.6.2) \\ i = 1, 2, \dots, n_2$$

BẢNG 15.2

Số liệu tiết kiệm và thu nhập cá nhân, Anh Quốc, 1946-1963 (triệu pound)

Thời kỳ I	Tiết kiệm	Thu nhập	Thời kỳ II	Tiết kiệm	Thu nhập
1946	0,36	8,8	1955	0,59	15,5
1947	0,21	9,4	1956	0,90	16,7
1948	0,08	10,0	1957	0,95	17,7
1949	0,20	10,6	1958	0,82	18,6
1950	0,10	11,0	1959	1,04	19,7
1951	0,12	11,9	1960	1,53	21,1
1952	0,41	12,7	1961	1,94	22,8
1953	0,50	13,5	1962	1,75	23,9
1954	0,43	14,3	1963	1,99	25,2

Nguồn: Cục Thống kê Trung ương, Anh Quốc

với Y = tiết kiệm (triệu pound)

X = thu nhập (triệu pound)

u_{1i}, u_{2i} = các yếu tố nhiễu trong hai hồi quy

Lưu ý: Số các quan sát n_1 và n_2 trong hai nhóm (các giai đoạn) không cần phải bằng nhau.

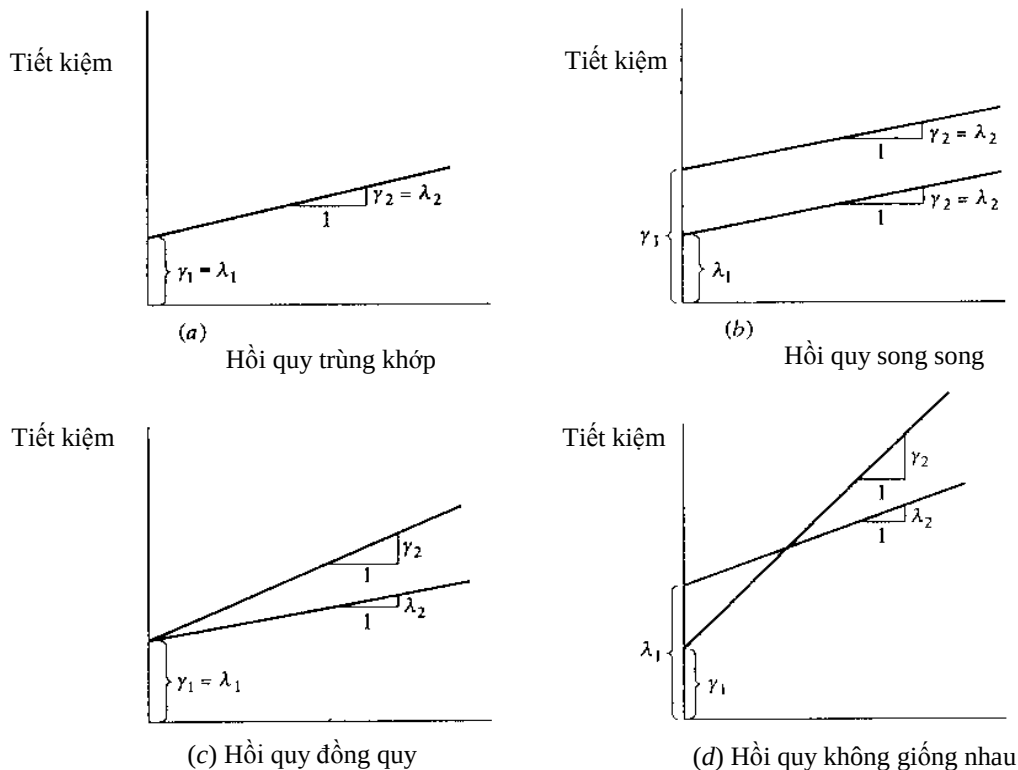
Bây giờ, các hồi quy (15.6.1) và (15.6.2) đại diện cho bốn khả năng sau:

1. $\lambda_1 = \gamma_1$ và $\lambda_2 = \gamma_2$; tức là, hai hồi quy đồng nhất nhau. (**Hồi quy trùng khớp**).
2. $\lambda_1 \neq \gamma_1$ và $\lambda_2 = \gamma_2$; tức là, hai hồi quy chỉ khác nhau ở vị trí của chúng (nghĩa là tung độ gốc). (**Hồi quy song song**).
3. $\lambda_1 = \gamma_1$ và $\lambda_2 \neq \gamma_2$; tức là, hai hồi quy có cùng tung độ gốc nhưng độ dốc khác nhau. (**Hồi quy đồng quy**).
4. $\lambda_1 \neq \gamma_1$ và $\lambda_2 \neq \gamma_2$; tức là, hai hồi quy hoàn toàn khác nhau. (**Hồi quy không giống nhau**).

Từ số liệu trong Bảng 15.2, ta có thể chạy hai hồi quy riêng (15.6.1) và (15.6.2) và sau đó sử dụng (các) kỹ thuật thống kê để kiểm định tất cả các khả năng ở trên, tức là, để tìm xem hàm tiết kiệm có bị **thay đổi cấu trúc** giữa hai thời đoạn hay không. Thay đổi cấu trúc có nghĩa là các tham số của hàm tiết kiệm thay đổi.

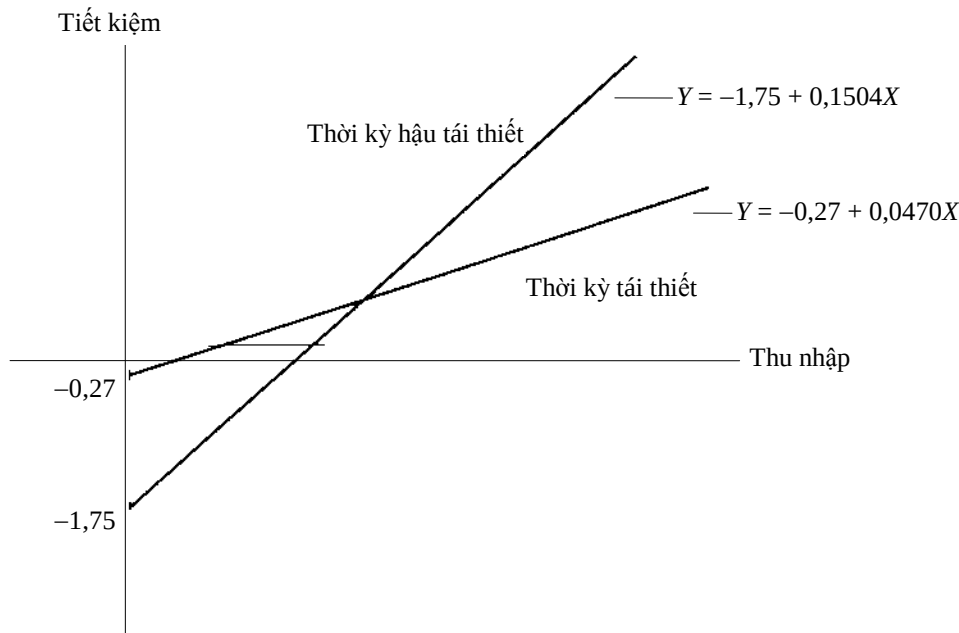
Một trong số các kỹ thuật đó được gọi là **kiểm định Chow**,⁵ mà ta đã thảo luận trong Mục 8.8. Kiểm định Chow chỉ ra rằng các tham số của hàm tiết kiệm giữa thời kỳ tái thiết và hậu tái thiết thật sự đã thay đổi.

Với vai trò một phương pháp thay thế cho kiểm định Chow, ta sẽ chỉ ra trong mục sau là làm thế nào mà kỹ thuật biến giả có thể giải quyết vấn đề thay đổi hay phá vỡ cấu trúc và đâu là các lợi thế của nó so với kiểm định Chow.



HÌNH 15.4
Các hồi quy tiết kiệm - thu nhập có thể xảy ra.

⁵ Về chi tiết của kiểm định Chow, xem Mục 8.8.



HÌNH 15.5
Các hồi quy tiết kiệm - thu nhập.

15.7 SO SÁNH HAI HỒI QUY: PHƯƠNG PHÁP BIẾN GIẢ

Thủ tục kiểm định Chow gồm nhiều bước thảo luận trong Mục 8.8 có thể được rút ngắn rất nhiều bằng cách sử dụng các biến giả. Mặc dù các kết luận toàn bộ rút ra từ các kiểm định Chow và biến giả trong mọi áp dụng đều như nhau, phương pháp biến giả có một số lợi thế mà ta sẽ giải thích sau khi trình bày phương pháp qua cùng ví dụ tiết kiệm - thu nhập.⁶

Hãy tập hợp tất cả các quan sát n_1 và n_2 lại và ước lượng hồi quy sau.⁷

$$Y_i = \alpha_1 + \alpha_2 D_i + \beta_1 X_i + \beta_2 (D_i X_i) + u_i \quad (15.7.1)$$

với Y_i và X_i là tiết kiệm và thu nhập như trước và với $D_i = 1$ cho các quan sát trong thời kỳ đầu hay tái thiết và 0 cho các quan sát trong thời kỳ hậu tái thiết.

Để xem các ý nghĩa của mô hình (15.7.1), và giả thiết rằng $E(u_i) = 0$, ta có

$$E(Y_i | D_i = 0, X_i) = \alpha_1 + \beta_1 X_i \quad (15.7.2)$$

$$E(Y_i | D_i = 1, X_i) = (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2) X_i \quad (15.7.3)$$

⁶ Các tài liệu trong phần này được lấy chủ yếu từ các bài viết của tác giả, “Use of Dummy Variables in Testing for Equality between Sets of Coefficients in Two Linear Regressions: A Note” (Sử dụng các biến giả trong kiểm định sự bằng nhau giữa hai tập hợp các hệ số trong hai hồi quy tuyến tính: một lưu ý) và “Use of Dummy Variables ...: A Generalization” (Sử dụng các biến giả ...: một sự tổng quát hóa”, cả hai đều được xuất bản trong *American Statistician* (Tập chí Nhà Thống kê Hoa Kỳ), tập 24, số 1 và 5, 1970, trang 50-52 và 18-21.

⁷ Như trong kiểm định Chow, kỹ thuật tập hợp đưa ra giả thiết rằng có phương sai thuần nhất, tức là, $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Nhưng từ Chương 11 bây giờ ta có một số phương pháp để kiểm định giả thiết này.

tương ứng là các hàm số tiết kiệm trung bình cho thời kỳ thứ hai (hậu tái thiết) và thứ nhất (tái thiết). Chúng giống như (15.6.2) và (15.6.1) với $\gamma_1 = \alpha_1$, $\gamma_2 = \beta_1$, $\lambda_1 = (\alpha_1 + \alpha_2)$ và $\lambda_2 = (\beta_1 + \beta_2)$. Do vậy, ước lượng (15.7.1) tương đương với ước lượng hai hàm tiết kiệm riêng (15.6.1) và (15.6.2).

Trong (15.7.1), α_2 là **tung độ gốc chênh lệch**, như trước đây, và β_2 là **hệ số góc chênh lệch** biểu thị hệ số góc của hàm tiết kiệm trong giai đoạn thứ nhất khác với hệ số góc của hàm tiết kiệm trong giai đoạn thứ hai là bao nhiêu. Lưu ý rằng việc biến giả D được đưa ra dưới dạng tích như thế nào (D nhân với X) để có thể cho phép ta phân biệt giữa hai hệ số góc của hai thời kỳ, cũng như là việc đưa ra biến giả dưới dạng tổng cho phép ta phân biệt giữa hai tung độ gốc trong hai thời kỳ.

Quay lại với số liệu tiết kiệm - thu nhập trong Bảng 15.2, ta tìm ra hàm thực nghiệm của (15.7.1) như sau:

$$\begin{aligned} Y_t &= -1,7502 + 1,4839D_t + 0,1504X_t - 0,1034D_tX_t \\ &\quad (0,3319) \quad (0,4704) \quad (0,0163) \quad (0,0322) \\ t &= (-5,2733) \quad (3,1545) \quad (9,2238) \quad (-3,1144) \\ \bar{R}^2 &= 0,9425 \end{aligned} \quad (15.7.4)$$

Như hồi quy này cho thấy, cả tung độ gốc chênh lệch và các hệ số góc chênh lệch đều có ý nghĩa thống kê. Điều này chỉ ra rõ ràng rằng các hồi quy trong hai thời kỳ khác nhau (đối chiếu Hình 15.4d). Sau đó, theo (15.7.2) và (15.7.3) ta có thể tính hai hồi quy như sau [(Lưu ý: $D = 1$ trong thời kỳ thứ nhất (xem Hình 15.5))]:

Thời kỳ tái thiết:

$$\begin{aligned} Y_t &= (-1,7502 + 1,4839) + (0,1504 - 0,1034)X_t \\ &= -0,2663 + 0,0470X_t \end{aligned} \quad (15.7.5)$$

Thời kỳ hậu tái thiết:

$$Y_t = -1,7502 + 0,1504X_t \quad (15.7.6)$$

Người đọc có thể thấy, hai hồi quy này giống những hồi quy tính từ quy tắc nhiều bước của Chow. Ta có thể nhận thấy điều này từ các hồi quy trong Mục 8.8.

Bây giờ thì các lợi thế của kỹ thuật biến giả [nghĩa là ước lượng (15.7.1)] so với kiểm định Chow [nghĩa là ba hồi quy riêng rẽ (8.8.1), (8.8.2) và hồi quy “tổng hợp”] có thể nhận thấy ngay là:

1. Ta chỉ cần chạy một hồi quy đơn bởi các hồi quy riêng có thể được suy ra từ nó một cách dễ dàng theo như các phương trình (15.7.2) và (15.7.3).
2. Hồi quy đơn phương trình có thể sử dụng để kiểm định các giả thiết khác nhau. Vậy, nếu hệ số tung độ gốc chênh lệch α_2 không có ý nghĩa thống kê, ta có thể chấp nhận giả thiết rằng hai hồi quy có cùng tung độ gốc, tức là, hai hồi quy đồng quy (xem Hình 15.4c). Tương tự, nếu hệ số góc chênh lệch β_2 không có ý nghĩa thống kê nhưng α_2 có ý nghĩa, ta có thể ít nhất là không bác bỏ giả thiết cho rằng hai hồi quy có cùng độ dốc, tức là, các đường hồi quy song song (đối chiếu Hình 15.4b). Kiểm định tính ổn định của toàn bộ hồi quy (nghĩa là $\alpha_2 = \beta_2$ và đồng thời = 0) có thể được thực hiện bằng kiểm định F về ý nghĩa toàn bộ của hồi quy ước lượng thảo luận trong Chương 8. Nếu giả thiết này đứng vững, các đường hồi quy sẽ trùng nhau như trong Hình 15.4a.

3. Kiểm định Chow không cho biết rõ ràng hệ số nào, tung độ gốc hay độ dốc, khác nhau, hay cả hai khác nhau (như trong ví dụ này) trong hai thời kỳ, tức là, ta có thể có một kiểm định Chow có ý nghĩa khi *chỉ có độ dốc* khác nhau hay khi *chỉ có tung độ gốc* khác nhau, hay cả hai đều khác nhau. Nói một cách khác, ta không thể nói, qua kiểm định Chow, khả năng nào trong số bốn khả năng minh họa trong Hình 15.4 tồn tại trong một ví dụ cụ thể. Về khía cạnh này, phương pháp biến giả có ưu thế rõ ràng, do nó không chỉ cho ta biết hai hồi quy có khác nhau không mà còn chỉ chính xác (các) nguồn gốc của sự khác nhau – đó là do tung độ gốc hay do độ dốc hay cả hai. Trên thực tế, việc biết rằng hai hồi quy khác nhau ở hệ số này hay hệ số kia cũng quan trọng như, nếu không muốn nói là quan trọng hơn, việc chỉ biết rằng chúng khác nhau.
4. Sau cùng, do việc tổng hợp làm tăng số bậc tự do, nó có thể cải thiện tính chính xác tương đối của các tham số ước lượng.⁸

15.8 SO SÁNH HAI HỒI QUY: MINH HỌA THÊM

Do tầm quan trọng thực tiễn của biến giả, ta xem xét một ví dụ nữa về cách sử dụng kỹ thuật này trong kiểm định tính tương đương của hai (hay nhiều) hồi quy.

Ví dụ 15.5 Hành vi của thất nghiệp và việc làm còn trống: Anh Quốc, 1958-1971⁹

Trong nghiên cứu quan hệ giữa thất nghiệp và tỷ lệ việc làm còn trống tại Anh Quốc trong giai đoạn 1958-IV đến 1971-II, tác giả đã có được đồ thị phân tán trong Hình 15.6. Như hình vẽ mô tả, bắt đầu quý IV của năm 1966, quan hệ thất nghiệp - chỗ làm việc còn trống có vẻ như đã thay đổi; đường cong biểu diễn quan hệ giữa hai biến dường như chuyển dịch lên trên vào đầu quý này. Sự dịch chuyển lên trên có nghĩa là đối với một tỷ lệ việc làm còn trống nhất định, tỷ lệ thất nghiệp vào quý IV năm 1966 cao hơn trước. Trong nghiên cứu của mình, tác giả đã tìm ra rằng nguyên nhân hợp lý tạo ra sự dịch chuyển lên trên là trong tháng 11 năm 1966 (tức là quý IV) Chính phủ Công đảng lên nắm quyền đã tự do hóa Luật Bảo hiểm Quốc gia bằng cách thay thế hệ thống tỷ lệ phúc lợi thất nghiệp cố định bằng một hệ thống hỗn hợp gồm phúc lợi thất nghiệp theo tỷ lệ cố định và theo thu nhập. Hệ thống này rõ ràng đã làm tăng mức phúc lợi thất nghiệp. Nếu phúc lợi thất nghiệp tăng, những người thất nghiệp có nhiều khả năng tìm kiếm việc làm trong thời gian lâu hơn, và do vậy tạo nên số thất nghiệp cao hơn đối với một tỷ lệ chỗ làm việc còn trống nhất định.

Để tìm xem sự thay đổi quan sát được trong mối quan hệ thất nghiệp - việc làm còn trống bắt đầu từ quý IV năm 1966 có ý nghĩa thống kê hay không, tác giả đã sử dụng mô hình sau:

$$UN_t = \alpha_1 + \alpha_2 D_t + \beta_1 V_t + \beta_2 (D_t V_t) + u_t \quad (15.8.1)$$

với UN = tỷ lệ thất nghiệp (%)
 V = tỷ lệ việc làm còn trống (%)
 D = 1 cho giai đoạn bắt đầu từ 1966-IV
 = 0 cho giai đoạn trước 1966-IV
 t = thời gian, tính theo quý

⁸ Nhưng lưu ý rằng việc thêm một biến giả sẽ sử dụng một bậc tự do.

⁹ Damodar Gujarati, "The Behaviour of Unemployment and Unfilled Vacancies: Great Britain, 1958-1971", *The Economic Journal* (Hành vi của thất nghiệp và chỗ làm việc còn trống: Anh Quốc, 1958-1971, *Tạp chí Kinh tế*), tập 82, 3/1972, trang 195-202.

Dựa vào 51 quan sát hàng quý trong giai đoạn 1958-IV đến 1971-II, các kết quả sau được tính toán (số liệu thực tế sử dụng được trình bày trong Phụ 15A), Mục 15A.1; người đọc có thể muốn xem những số liệu này vì chúng cho biết các biến giả được thiết lập như thế nào):

$$\begin{aligned}\widehat{UN}_t &= 2,7491 + 1,1507D_t - 1,5294V_t - 0,8511(D_tV_t) \\ &\quad (0,1022) \quad (0,3171) \quad (0,1218) \quad (0,4294) \\ t &= (26,896) \quad (3,6288) \quad (-12,5552) \quad (-1,9819) \quad R^2 = 0,9128\end{aligned}\quad (15.8.2)$$

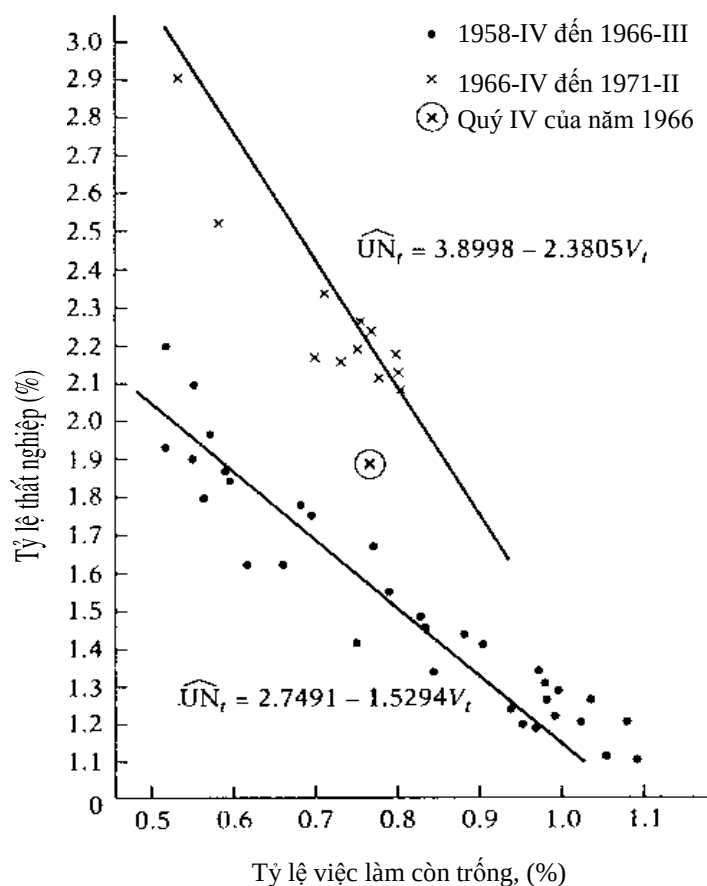
Đánh giá theo các tiêu chí thông thường, hồi quy ước lượng được cho ta một sự thích hợp tuyệt vời. Lưu ý rằng cả hệ số tung độ gốc và hệ số góc chênh lệch đều có ý nghĩa thống kê ở mức 5% (một phía). Vậy, ta có thể chấp nhận giả thiết rằng rõ ràng có sự chuyển dịch trong mối quan hệ UN-V bắt đầu từ quý IV của năm 1966.¹⁰

Từ hồi quy ở trên, ta có thể tính các hồi quy sau:

$$1958-IV \text{ đến } 1966-III: \widehat{UN}_t = 2,7491 - 1,5294V_t \quad (15.8.3)$$

$$\begin{aligned}1966-IV \text{ đến } 1971-II: \widehat{UN}_t &= (2,7491 + 1,15) - (1,5294 + 0,8511)V_t \\ &= 3,8998 - 2,3805V_t\end{aligned}\quad (15.8.4)$$

Các hồi quy trên được biểu diễn trong Hình 15.6. Các hồi quy chỉ ra rằng trong giai đoạn bắt đầu từ 1966-IV, đường cong UN-V dốc hơn và có tung độ gốc cao hơn nhiều so với giai đoạn bắt đầu từ 1958-IV.



HÌNH 15.6

Đồ thị phân tán giữa tỷ lệ thất nghiệp và tỷ lệ việc làm còn trồng, Anh Quốc, 1958-IV đến 1971-II.

¹⁰ Các kết quả được tính dựa trên giả thiết rằng các phương sai của sai số bằng nhau trong hai giai đoạn. Nhưng như đã lưu ý trong chú thích 7, giả thiết này phải được kiểm định một cách rõ ràng (xem bài tập 15.18).

15.9 CÁC TÁC ĐỘNG QUA LẠI

Xem xét mô hình sau:

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + u_i \quad (15.9.1)$$

với Y_i = chi tiêu may mặc hàng năm

X_i = thu nhập

$D_2 = 1$ nếu là nữ
 $= 0$ nếu là nam

$D_3 = 1$ nếu đã tốt nghiệp đại học
 $= 0$ nếu khác

Ấn trong mô hình là giả thiết rằng tác động chênh lệch của biến giả giới tính D_2 đối với hai cấp trình độ học vấn là hằng số và tác động chênh lệch của biến giả trình độ học vấn D_3 đối với hai giới tính cũng là hằng số. Tức là, nếu chi tiêu may mặc bình quân của nữ cao hơn nam thì nó là như vậy cho dù người ta đã tốt nghiệp đại học hay chưa. Cũng như vậy, nếu những người tốt nghiệp đại học tính trung bình chi tiêu cho may mặc nhiều hơn những người chưa tốt nghiệp đại học, thì nó là như vậy cho dù họ là nữ hay nam.

Trong nhiều áp dụng, một giả thiết như vậy có thể không đúng. Một người nữ đã tốt nghiệp đại học có thể chi nhiều hơn cho may mặc nhiều hơn là một người nam đã tốt nghiệp đại học. Nói một cách khác, có thể có tương tác giữa những biến định tính D_2 và D_3 và do vậy tác động của chúng đối với giá trị trung bình của Y và X có thể không đơn giản là tổng như trong (15.9.1) mà cũng có thể là tích như trong mô hình sau:

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 (D_{2i} D_{3i}) + \beta X_i + u_i \quad (15.9.2)$$

Từ (15.9.2) ta có

$$E(Y_i | D_2 = 1, D_3 = 1, X_i) = (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) + \beta X_i \quad (15.9.3)$$

là chi tiêu may mặc bình quân của nữ đã tốt nghiệp đại học. Lưu ý rằng

α_2 = tác động chênh lệch khi là nữ

α_3 = tác động chênh lệch khi đã tốt nghiệp đại học

α_4 = tác động chênh lệch khi là nữ đã tốt nghiệp đại học

Điều này cho thấy chi tiêu may mặc trung bình của nữ đã tốt nghiệp đại học khác với chi tiêu may mặc trung bình của nữ hay của người đã tốt nghiệp đại học (sự khác biệt là α_4). Nếu α_2 , α_3 và α_4 đều dương, chi tiêu may mặc trung bình của nữ cao hơn (so với phạm trù cơ sở, mà ở đây là nam chưa tốt nghiệp đại học), nhưng còn cao hơn nhiều nếu là nữ đã tốt nghiệp đại học. Tương tự, chi tiêu may mặc trung bình của một người đã tốt nghiệp đại học có xu hướng cao hơn phạm trù cơ sở, nhưng sẽ cao hơn nhiều nếu người đã tốt nghiệp đại học lại là nữ. Điều này chỉ ra rằng **biến giả tương tác** làm thay đổi tác động của hai thuộc tính xem xét một cách riêng rẽ như thế nào.

Việc hệ số của biến giả tương tác có ý nghĩa thống kê hay không có thể được kiểm định bằng kiểm định t thông thường. Nếu nó có ý nghĩa, sự có mặt đồng thời của hai thuộc tính sẽ làm giảm bớt hay tăng cường các tác động riêng lẻ của các thuộc tính này. Không cần phải nói, bỏ qua một số hạng tương tác quan trọng một cách không đúng đắn sẽ dẫn tới một thiên lệch đặc trưng.

15.10 CÁCH SỬ DỤNG CÁC BIẾN GIẢ TRONG PHÂN TÍCH MÙA

Nhiều chuỗi số liệu kinh tế theo thời gian dựa vào số liệu hàng tháng hay hàng quý có hình thái theo mùa (dao động đều đặn). Các ví dụ gồm có doanh thu của các cửa hàng tổng hợp vào dịp Giáng sinh, mức cầu tiền tệ (cân đối tiền mặt) của gia đình vào kỳ nghỉ, nhu cầu kem và nước giải khát trong mùa hè và giá nông sản ngay sau vụ thu hoạch. Thường thì cần phải loại bỏ các yếu tố hay *thành phần* mùa khỏi chuỗi thời gian để ta có thể tập trung vào các thành phần khác, như xu hướng.¹¹ Quá trình loại bỏ thành phần mùa khỏi một chuỗi thời gian được gọi là **loại bỏ yếu tố mùa** hay **hiệu chỉnh yếu tố mùa**, và chuỗi thời gian sau khi tính được gọi là chuỗi thời gian đã loại bỏ yếu tố mùa hay đã hiệu chỉnh yếu tố mùa, ví dụ như chỉ số giá tiêu dùng, chỉ số giá bán buôn, chỉ số sản xuất công nghiệp thường được xuất bản ở dạng đã hiệu chỉnh yếu tố mùa.

Có một số phương pháp để loại bỏ yếu tố mùa trong một chuỗi thời gian, nhưng ta sẽ chỉ xem xét một trong những phương pháp này, gọi là, *phương pháp biến giả*.¹² Để minh họa các biến giả được sử dụng như thế nào để loại bỏ yếu tố mùa trong chuỗi thời gian kinh tế, giả sử rằng ta muốn thực hiện hồi quy lợi nhuận của các công ty công nghiệp chế tạo Hoa Kỳ theo doanh thu trong các giai đoạn theo quý, 1965-1970. Các số liệu thích hợp chưa hiệu chỉnh yếu tố mùa được trình bày trong Phụ lục 15A, Mục 15A.2. Phần phụ lục cũng chỉ ra làm thế nào ta có thể chuẩn bị *ma trận số liệu* để lập các biến giả. Xem xét các số liệu này cho thấy một hình thái thú vị. Cả lợi nhuận và doanh thu đều cao hơn trong quý II so với quý I và quý III của từng năm. Có lẽ quý II có tác động mùa. Để điều tra hiện tượng này, ta tiến hành như sau:

Ví dụ 15.6 Hành vi lợi nhuận - doanh thu trong ngành công nghiệp chế tạo Hoa Kỳ

$$\text{Lợi nhuận}_t = \alpha_1 + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{4t} + \beta(\text{doanh thu})_t + u_t \quad (15.10.1)$$

$$\begin{aligned} \text{với } D_2 &= 1 && \text{đối với quý II} \\ &= 0 && \text{các quý khác} \\ D_3 &= 1 && \text{đối với quý III} \\ &= 0 && \text{các quý khác} \\ D_4 &= 1 && \text{đối với quý IV} \\ &= 0 && \text{các quý khác} \end{aligned}$$

Lưu ý rằng ta đang giả thiết là biến “mùa” có bốn loại, bốn quý trong một năm, do vậy cần phải sử dụng ba biến. Vậy, nếu xuất hiện hình thái mùa trong các quý khác nhau, các tung độ gốc chênh lệch ước lượng α_2 , α_3 và α_4 , nếu có ý nghĩa thống kê, sẽ phản ánh hình thái mùa. Có thể là chỉ có một số tung độ gốc có ý nghĩa thống kê cho nên chỉ có một số quý có thể phản ánh hình thái mùa. Nhưng mô hình (15.10.1) không đủ tổng quát để tính cho tất cả các trường hợp này. (Lưu ý, ta coi quý I của năm là quý cơ sở).

Sử dụng số liệu trong Phụ lục 15A, Mục 15A.2, ta có các kết quả sau (số liệu lợi nhuận và doanh thu được tính theo triệu USD):

¹¹ Một chuỗi thời gian có bốn thành phần: mùa, chu kỳ, xu hướng và một thành phần hoàn toàn ngẫu nhiên.

¹² Một số phương pháp khác là phương pháp tỷ lệ - trung bình dịch chuyển, phương pháp liên kết tương đối và phương pháp tỷ lệ phần trăm của trung bình năm. Về thảo luận không mang tính kỹ thuật của những phương pháp này, xem Morris Hamburg, *Statistical Analysis for Decision Making* (Phân tích thống kê cho việc ra quyết định), Harcourt, Brace & World, New York, 1970, trang 563-575.

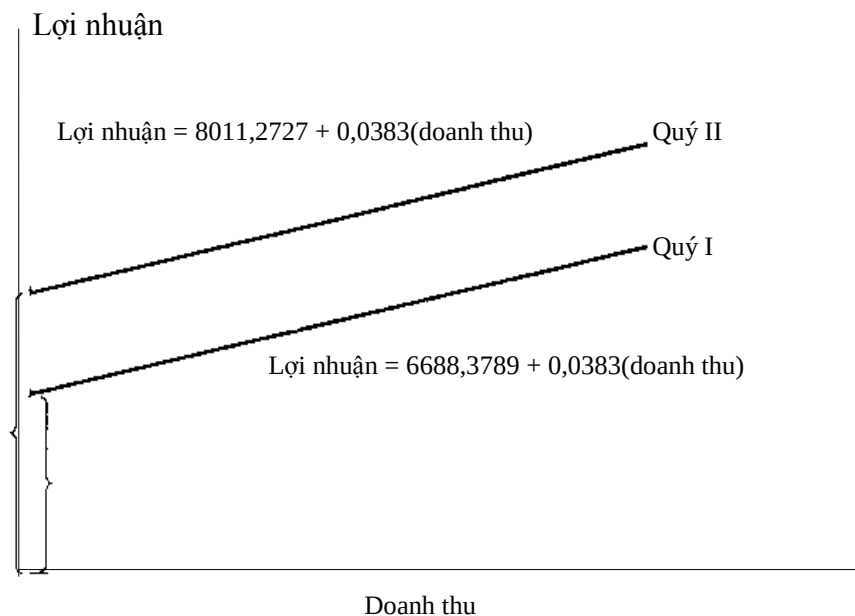
$$\begin{aligned}
 \text{Lợi nhuận} = & 6688,3789 + 1322,8938D_{2t} - 217,8037D_{3t} + 183,8597D_{4t} + 0,0383(\text{doanh thu})_t \\
 & (1711,3707) \quad (638,4753) \quad (632,2561) \quad (654,2937) \quad (0,0115) \\
 t = & (3,9082) \quad (2,0720) \quad (-0,3445) \quad (0,2810) \quad (3,3313) \\
 & R^2 = 0,5255 \quad (15.10.2)
 \end{aligned}$$

Các kết quả cho thấy chỉ các hệ số doanh thu và tung độ gốc chênh lệch gần với quý II là có ý nghĩa thống kê ở mức 5%. Vậy, ta có thể kết luận rằng có yếu tố mùa trong quý II của từng năm: Hệ số doanh thu 0,0383 cho ta biết rằng, sau khi tính đến tác động của mùa, nếu doanh thu tăng lên, ví dụ, 1 USD, lợi nhuận bình quân theo dự kiến sẽ tăng lên khoảng 4 xu. Mức lợi nhuận trung bình trong quý cơ sở hay quý I là 6688 USD và trong quý II, nó có mức cao hơn khoảng 1323 USD tức là vào khoảng 8011 USD. (Xem Hình 15.7).¹³

Do quý II dường như khác với các quý khác, nếu muốn ta có thể chạy lại (15.10.2) chỉ sử dụng một biến giả để phân biệt quý II với các quý khác như sau:

$$\begin{aligned}
 Y_t = & 6516,6 + 1311,4D_2 + 0,0393(\text{doanh thu}) \\
 & (1623,1) \quad (493,02) \quad (0,0106) \\
 t = & (4,0143) \quad (2,7004) \quad (3,7173) \\
 & R^2 = 0,5155 \quad (15.10.3)
 \end{aligned}$$

với $D_2 = 1$ đối với quan sát trong quý II và 0 đối với các quý khác.



HÌNH 15.7

Mối quan hệ giữa lợi nhuận và doanh thu trong các công ty công nghiệp chế tạo tại Hoa Kỳ, 1965-I đến 1970-II.

¹³ Lưu ý: Về số, các tung độ gốc trong quý II và III khác với quý I nhưng về thống kê thì chúng như nhau. (Tại sao?)

Người đọc sẽ nhận thấy rằng (15.10.3) là dạng giới hạn của (15.10.2). Sự giới hạn là tung độ gốc của quý I, III và IV bằng nhau. Để đánh giá các kết quả (15.10.2) ta sẽ dự kiến rằng các hạn chế này sẽ có giá trị nhưng từ Chương 8 ta đã biết làm thế nào để kiểm định chúng một cách rõ ràng. Trong bài tập 15.21, bạn được yêu cầu kiểm chứng rằng các hạn chế này có hiệu lực thực sự. Do vậy, kết luận vẫn như trước – chỉ có hình thái mùa trong quý II.

Trong công thức của mô hình (15.10.1), ta đã giả thiết rằng chỉ có tung độ gốc là khác nhau giữa các quý, hệ số góc của biến doanh thu ở mỗi quý là như nhau. Nhưng giả thiết này có thể được kiểm định bằng kỹ thuật biến giả dạng tích số đã thảo luận ở trên. (xem bài tập 15.22).

15.11 HỒI QUY TUYẾN TÍNH TỪNG KHÚC

Để minh họa một ứng dụng nữa của các biến giả, hãy xem Hình 15.8, trong đó biểu diễn một công ty trên giả thiết thưởng cho các đại lý bán hàng của mình như thế nào. Công ty trả hoa hồng dựa vào doanh thu theo cách thức mà cho tới một mức nhất định, gọi là mức *mục tiêu* hay *ngưỡng*, X^* , thì có một cơ chế hoa hồng (ngẫu nhiên) và cao hơn mức đó thì một cơ chế khác. (Lưu ý: Bên cạnh doanh thu, các yếu tố khác cũng tác động tới hoa hồng doanh thu). Giả sử rằng các nhân tố khác được đại diện bởi yếu tố nhiễu ngẫu nhiên). Cụ thể hơn, ta giả thiết rằng hoa hồng doanh thu tăng tuyến tính theo doanh thu cho tới ngưỡng X^* , sau đó cũng tăng lên tuyến tính theo doanh thu nhưng với tốc độ nhanh hơn nhiều. Như vậy, ta có **hồi quy tuyến tính từng khúc** gồm hai phần hay hai đoạn tuyến tính, ký hiệu là I và II trong Hình 15.8, và hàm hoa hồng thay đổi độ dốc của nó tại giá trị ngưỡng. Với số liệu về hoa hồng, doanh thu và giá trị ngưỡng X^* , kỹ thuật biến giả có thể được sử dụng để ước lượng các độ dốc (khác nhau) của hai đoạn của hồi quy tuyến tính từng khúc biểu diễn trong Hình 15.8. Ta tiến hành như sau:

$$Y_i = \alpha_1 + \beta_1 X_i + \beta_2 (X_i - X^*) D_i + u_i \quad (15.11.1)$$

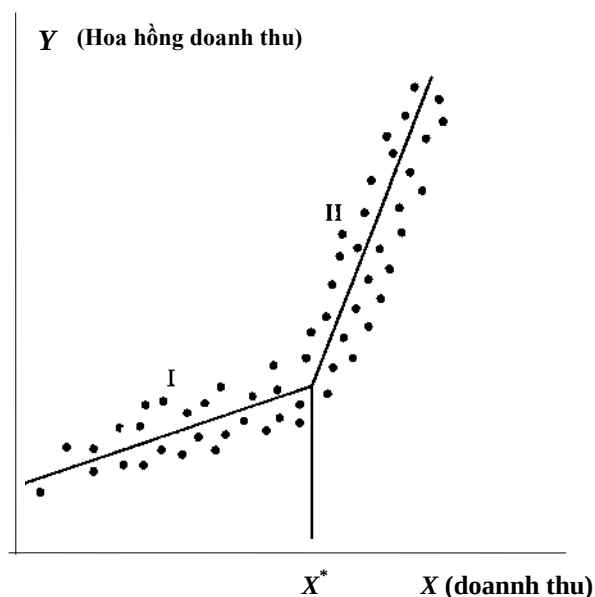
với Y_i = hoa hồng doanh thu

X_i = doanh thu phát sinh từ đại lý

X^* = giá trị ngưỡng của doanh thu, cũng được gọi là **điểm gãy khúc** (đã biết trước)¹⁴

$D = 1$ nếu $X_i > X^*$
 $= 0$ nếu $X_i < X^*$

¹⁴ Tuy nhiên, giá trị ngưỡng có thể không luôn luôn biết trước. Phương pháp cụ thể là vẽ biến phụ thuộc theo (các) biến giải thích và quan sát nếu có một thay đổi sắc nét trong quan hệ sau một giá trị nhất định của X (nghĩa là X^*). Một phương pháp phân tích xác định điểm gãy có thể được tìm thấy trong các **mô hình hồi quy chuyển đổi**. Nhưng đây là nội dung ở trình độ cao. Thảo luận mang tính giáo khoa về nội dung này có thể tìm thấy trong Thomas Fomby, R. Carter Hill và Stanley Johnson, *Advanced Econometric Methods* (Các phương pháp kinh tế lượng cao cấp), Springer-Verlag, New York, 1984, Chương 14.

**HÌNH 15.8**

Mối quan hệ giả thiết giữa hoa hồng bán hàng và doanh thu. (Lưu ý: Tung độ gốc trên trục Y biểu diễn mức hoa hồng đảm bảo tối thiểu).

Giả sử $E(u_i) = 0$, ta nhận thấy ngay rằng

$$E(Y_i | D_i = 0, X_i, X^*) = \alpha_1 + \beta_1 X_i \quad (15.11.2)$$

(15.11.2) cho biết mức hoa hồng doanh thu trung bình cho tới mức mục tiêu X^* .

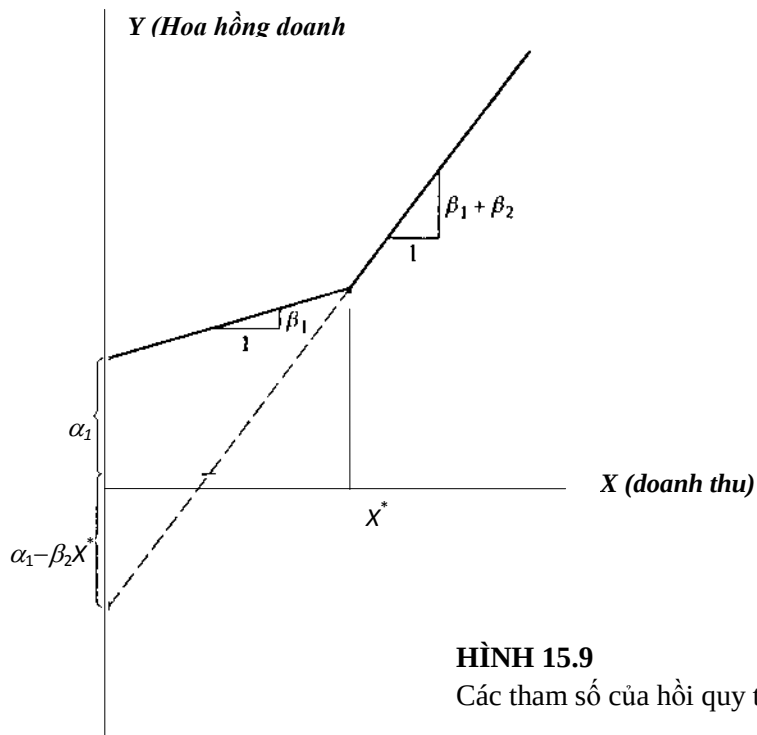
$$E(Y_i | D_i = 1, X_i, X^*) = \alpha_1 - \beta_2 X^* + (\beta_1 + \beta_2) X_i \quad (15.11.3)$$

(15.11.3) cho biết mức hoa hồng doanh thu trung bình khi doanh thu lớn hơn mức X^* .

Vậy, β_1 cho biết độ dốc của đường hồi quy trong đoạn I, và $\beta_1 + \beta_2$ cho biết độ dốc của đường hồi quy trong đoạn II của hồi quy từng khúc biểu diễn trong Hình 15.8. Một kiểm định về giả thiết cho rằng không có gãy khúc trong hồi quy tại giá trị ngưỡng X^* có thể được thực hiện dễ dàng bằng cách xác định mức ý nghĩa thống kê của hệ số góc chênh lệch ước lượng $\hat{\beta}_2$ (xem Hình 15.9).

Nhân đây, hồi quy tuyến tính gãy khúc ta vừa thảo luận là một ví dụ của một lớp hàm số tổng quát hơn gọi là hàm đa thức gãy khúc (spline).¹⁵

¹⁵ Đối với thảo luận về splines (nghĩa là đa thức bậc k gãy khúc), xem Douglas C. Montgomery & Elizabeth A. Peck, *Introduction to Linear Regression Analysis* (Giới thiệu phân tích hồi quy tuyến tính), John Wiley & Sons, xuất bản lần thứ 2, New York, 1992, trang 210-218.



HÌNH 15.9

Các tham số của hồi quy tuyến tính từng khúc.

Ví dụ 15.7 Tổng chi phí trong quan hệ với sản lượng

Để ví dụ về ứng dụng của hồi quy tuyến tính từng khúc, hãy xem xét số liệu tổng chi phí - tổng sản lượng giả thiết trong Bảng 15.3. Ta được cho biết rằng tổng chi phí có thể thay đổi độ dốc của nó tại mức sản lượng 5500 đơn vị.

Gọi Y trong (15.11.1) đại diện cho tổng chi phí và X đại diện cho tổng sản lượng, ta có các kết quả sau:

$$\begin{aligned} Y_i &= -145,72 + 0,2791X_i + 0,0945(X_i - X_i^*)D_i \\ t &= (-0,8245) \quad (6,0669) \quad (1,1447) \\ R^2 &= 0,9737 \quad X^* = 5500 \end{aligned} \quad (15.11.4)$$

BẢNG 15.3
Số liệu giả thiết về sản lượng và tổng chi phí

Tổng chi phí (USD)	Sản lượng (đơn vị)
256	1.000
414	2.000
634	3.000
778	4.000
1.003	5.000
1.839	6.000
2.081	7.000
2.423	8.000
2.734	9.000
2.914	10.000

Như các kết quả này biểu thị, chi phí sản xuất biên tế vào khoảng 28 xu một đơn vị sản lượng và mặc dù nó vào khoảng 37 xu ($28 + 9$) khi sản lượng vượt 5500 đơn vị, sự khác nhau giữa hai phần không có ý nghĩa thống kê do biến giả không có ý nghĩa ở mức 5%. Do vậy, đối với tất cả các mục đích thực tiễn, ta có thể thực hiện hồi quy tổng chi phí theo tổng sản lượng và loại bỏ biến giả.

15.12 SỬ DỤNG CÁC BIẾN GIẢ TRONG VIỆC KẾT HỢP SỐ LIỆU CHUỖI THỜI GIAN VÀ SỐ LIỆU CHÉO

Để minh họa tính đa dụng của biến giả, ta xem xét trong mục này một ứng dụng nữa của nó.

Hồi quy tổng hợp: Tổng hợp số liệu chuỗi thời gian và số liệu chéo

Xem xét số liệu trong Bảng 15.4 được lấy từ nghiên cứu nổi tiếng về lý thuyết đầu tư do Grunfeld đề xuất.¹⁶ Grunfeld đã quan tâm tới việc tìm xem đầu tư gộp (Y) phụ thuộc như thế nào vào giá trị của công ty (X_2) và lượng vốn (X_3). Trong bảng này, các số liệu cho từng biến trong mỗi năm được xem xét cho công ty G.M. và Westinghouse (Tạm thời bỏ qua số liệu của công ty G.E., nhưng sẽ xem xét nó trong bài tập 15.31). Những số liệu này là một ví dụ của **số liệu chéo**. Đồng thời, với từng công ty ta có số liệu về các biến trong 20 năm. Đây là ví dụ về **số liệu chuỗi thời gian**. Bây giờ, để nghiên cứu phản ứng của Y theo X_2 và X_3 , ta có thể tiến hành theo một trong ba cách.

Thứ nhất, ta có thể chạy **hồi quy chuỗi thời gian** sau cho từng công ty riêng biệt:

$$\text{G.M.:} \quad Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + u_t \quad (15.12.1)$$

$$\text{Westinghouse:} \quad Y_t = \alpha'_1 + \alpha'_2 X_{2t} + \alpha'_3 X_{3t} + u_t \quad (15.12.2)$$

Sử dụng kỹ thuật biến giả hay kiểm định Chow, ta có thể tìm xem các tham số của hai hàm đầu tư có như nhau hay không.

¹⁶ Các số liệu này được tái lập trong một số sách. Chúng tôi lấy chúng từ H. D. Vinod & Aman Ullah, *Recent Advances in Regression Methods* (Những tiến bộ gần đây trong các phương pháp hồi quy), Marcel Dekker, New York, 1981, trang 259-261. Nguồn ban đầu là Y. Grunfeld, *The Determinants of Corporate Investment* (Các yếu tố xác định của đầu tư công ty), luận văn tiến sĩ không xuất bản, Khoa Kinh tế học, Đại học Chicago, 1958.

Thứ hai, với từng năm ta có thể ước lượng **hồi quy chéo**. Không may là trong trường hợp này, ta không thể thực hiện được bởi vì chỉ có hai quan sát chéo (hai công ty) mà lại có ba tham số phải ước lượng. Ví dụ, nếu có ít nhất bốn công ty thì ta có thể ước lượng hồi quy chéo như thế cho 20 năm, tạo ra tổng cộng 20 hồi quy chéo.

BẢNG 15.4**Số liệu đầu tư của các công ty G.M., Westinghouse và G.E.**

	G.M.			Westinghouse			G.E.		
	<i>I</i>	<i>F</i> ₋₁	<i>C</i> ₋₁	<i>I</i>	<i>F</i> ₋₁	<i>C</i> ₋₁	<i>I</i>	<i>F</i> ₋₁	<i>C</i> ₋₁
1935	317,6	3.078, 5	2,8	12,9	191,5	1,8	33,1	1.170, 6	97,8
36	391,8	4.661, 7	52,6	25,9	516,0	0,8	45,0	2.015, 8	104,4
37	410,6	5.387, 1	156,9	35,1	729,0	7,4	77,2	2.803, 3	118,0
38	257,7	2.792, 2	209,2	22,9	560,4	18,1	44,6	2.039, 7	156,2
39	330,8	4.313, 2	203,4	18,8	519,9	23,5	48,1	2.256, 2	172,6
40	461,2	4.643, 9	207,2	28,6	628,5	26,5	74,4	2.132, 2	186,6
41	512,0	4.551, 2	255,2	48,5	537,1	36,2	113,0	1.834, 1	220,9
42	448,0	3.244, 1	303,7	43,3	561,2	60,8	91,9	1.588, 0	287,8
43	499,6	4.053, 7	264,1	37,0	617,2	84,4	61,3	1.749, 4	319,9
44	547,5	4.379, 3	201,6	37,8	626,7	91,2	56,8	1.687, 2	321,3
45	561,2	4.840, 9	265,0	39,3	737,2	92,4	93,6	2.007, 7	319,6
46	688,1	4.900, 9	402,2	53,5	760,5	86,0	159,9	2.208, 3	346,0
47	568,9	3.526, 5	761,5	55,6	581,4	111,1	147,2	1.656, 7	456,4
48	529,2	3.251, 7	922,4	49,6	662,3	130,6	146,3	1.604, 4	543,4
49	555,1	3.700, 2	1.020, 1	32,0	583,8	141,8	98,3	1.431, 8	618,3
50	642,9	3.755, 6	1.099, 0	32,2	635,2	136,7	93,5	1.610, 5	647,4
51	755,9	4.833, 0	1.207, 7	54,4	723,8	129,7	135,2	1.819, 4	671,3
52	891,2	4.924, 9	1.430, 5	71,8	864,1	145,5	157,3	2.079, 7	726,1
53	1.304,	6.241,	1.777,	90,1	1.193,	174,8	179,5	2.371,	800,3

	4	7	3		5		6	
54	1.486,	5.593,	2.226,	68,6	1.188,	213,5	189,6	2.759, 888,9
	7	6	3		9		9	

Lưu ý: $Y = I$ = đầu tư gộp = gia tăng máy móc, thiết bị cộng bảo trì và sửa chữa, tính theo triệu USD, hiệu chỉnh theo tỷ số giảm phát P_1 .

$X_2 = F$ = giá trị công ty = giá cổ phiếu thường và cổ phiếu ưu đãi vào ngày 31/12 (hay giá trung bình của ngày 31/12 và 31/1 năm sau) nhân với số cổ phiếu thường và ưu đãi do bên ngoài nắm giữ cộng với giá trị ghi sổ của số nợ vào ngày 31/12, tính theo triệu USD, hiệu chỉnh theo tỷ số giảm phát P_2 .

$X_3 = C$ = giá trị máy móc và thiết bị = tổng gia tăng ròng máy móc, thiết bị lũy tích hiệu chỉnh theo tỷ số giảm phát P_1 trừ cho khấu hao có hiệu chỉnh theo tỷ số giảm phát P_3 theo các định nghĩa sau.

P_1 = hệ số giảm phát ngầm của thiết bị lâu bền của nhà sản xuất (1947 = 100)

P_2 = hệ số giảm phát ngầm của GNP (1947 = 100)

P_3 = hệ số lạm phát chi phí khấu hao = trung bình dịch chuyển 10 năm của chỉ số giá bán buôn của kim loại và sản phẩm kim loại (1947 = 100)

Nguồn: Lấy từ H. D. Vinod & Aman Ullah, *Recent Advances in Regression Methods* (Những tiến bộ gần đây trong các phương pháp hồi quy), Marcel Dekker, New York, 1981, trang 259-261.

Thứ ba, tại sao lại không kết hợp tất cả 40 quan sát (20 quan sát chuỗi cho 2 công ty) và ước lượng mô hình sau?

$$Y_{it} = \beta_1 + \beta_2 X_{2it} + \beta_3 X_{3it} + u_{it} \quad (15.12.3)$$

với i đại diện cho công ty thứ i và t đại diện cho thời đoạn thứ t . Trong ví dụ của ta, $i = 2$ và $t = 20$, như vậy cho ta tổng cộng 40 quan sát. Phương trình (15.12.3) là một ví dụ của **hồi quy tổng hợp** với các quan sát chuỗi thời gian và quan sát chéo được kết hợp hay tổng hợp lại với nhau. Những hồi quy như vậy thường được ước lượng trong các tình thế mà ta có quá ít quan sát chéo (như trong ví dụ này) và số lượng các quan sát chuỗi thời gian khá tốt. Như Vinod và Ullah viết:

Khi làm việc với số liệu chéo và chuỗi thời gian, khi từng mẫu số liệu chéo riêng lẻ quá nhỏ nên các suy luận sâu về các hệ số không thể thực hiện được, thông lệ chung trong nghiên cứu ứng dụng là kết hợp tất cả số liệu lại với nhau và ước lượng một hồi quy chung. Động lực thúc đẩy cơ bản để tổng hợp số liệu chéo và chuỗi thời gian là nếu mô hình được xác định đúng đắn, kết hợp số liệu cho ta sự ước lượng, suy luận và dự báo hiệu quả hơn.¹⁷

Bây giờ hãy xem xét các vấn đề khó khăn trong ước lượng (15.12.3). Giả sử ta ước lượng nó bằng thủ tục OLS thông thường: đơn giản là ta nhập các quan sát của G.M. và Westinghouse, 20 quan sát đầu là của G.M. và 20 quan sát cuối là của Westinghouse. Cách làm này có sai hay không?

Cách làm như vậy giả định ngầm là các tham số của hồi quy không thay đổi theo thời gian (ổn định theo thời gian) và chúng không khác nhau giữa các đơn vị số liệu chéo (ổn định chéo). Cũng ngầm định trong một thủ tục như thế là giả thiết rằng phương sai của sai số của hàm đầu tư G.M. và Westinghouse thuần nhất và sai số trong hàm đầu tư G.M. tại thời gian t không tương quan với sai số trong hàm đầu tư Westinghouse tại thời gian t . Rõ ràng đây là các

¹⁷ Ibid., trang 248.

giả thiết quá mức. Có nhiều cách khác nhau để tháo bỏ các giả thiết này và đưa chúng vào trong thủ tục ước lượng. Không may là các hạn chế về thời gian, không gian và toán học không cho phép ta theo đuổi phương pháp này.¹⁸ Ta sẽ chỉ trình bày một trường hợp mà ta giả thiết rằng các giá trị tung độ gốc trong hàm đầu tư G.M. và Westinghouse khác nhau (câu hỏi về ổn định chéo) nhưng các hệ số góc như nhau. Ta cũng giả thiết rằng sai số trong hồi quy tổng hợp có các tính chất OLS thông thường đối với tất cả các quan sát chéo và chuỗi thời gian.

Với những giả thiết này, ta viết (15.12.3) dưới dạng

$$Y_{it} = \beta_1 + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_4 D_{it} + u_{it} \quad (15.12.4)$$

với $D_{it} = 1$ đối với các quan sát G.M. và 0 đối với các quan sát khác. Vậy, nếu β_4 trong (15.2.4) có ý nghĩa thống kê thì có nghĩa là giá trị của tung độ gốc của hàm đầu tư G.M. khác với Westinghouse. Nói cách khác, β_4 là giá trị *tung độ gốc chênh lệch*. Trong bài tập 15.32, người đọc được yêu cầu đưa ra các hệ số góc chênh lệch.

Ví dụ 15.8 Hàm đầu tư cho công ty General Motors và Westinghouse

Sử dụng số liệu trong Bảng 15.4, ta có các ước lượng sau đây của (15.12.4).¹⁹

$$\begin{aligned} Y_{it} &= -58,09 + 0,0986X_{2it} + 0,3763X_{3it} - 2,6929D_{it} & (15.12.5) \\ t &= (-3,291) (7,409) & (14,01) & (-0,0579) \\ &R^2 = 0,9673 & d = 0,7765 \end{aligned}$$

Như kết quả cho thấy, do biến giả tung độ gốc chênh lệch không có ý nghĩa thống kê, ta có thể kết luận rằng hàm đầu tư của G.M. và Westinghouse có các tung độ gốc như nhau về mặt thống kê. Tất nhiên là phải cẩn thận với kết luận này, do ta chỉ cho phép các tung độ gốc khác nhau mà không không phải là độ dốc. Việc thống kê Durbin-Watson có giá trị thấp cho thấy có thể có sai số cụ thể hóa trong (15.12.4). Phải thừa nhận rằng hồi quy (15.12.4) được lựa chọn chỉ để minh họa cách sử dụng các biến giả trong số liệu tổng hợp.

* 15.13 MỘT SỐ KHÍA CẠNH KỸ THUẬT CỦA PHƯƠNG PHÁP BIẾN GIẢ

Trong phần này, ta thảo luận một số điểm sâu hơn về sử dụng các biến giả trong phân tích hồi quy.

Giải thích các biến giả trong các hồi quy bán lôgarít

Nhớ lại thảo luận của chúng ta về các mô hình hồi quy log-lin trong đó biến được hồi quy có dạng lôgarít và biến làm hồi quy có dạng tuyến tính. Cụ thể, hãy xem xét mô hình sau:

$$\ln Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i \quad (15.13.1)$$

với Y = lương khởi điểm của giáo sư đại học, X = số năm kinh nghiệm giảng dạy, và $D = 1$ nếu là nam và 0 nếu khác.

¹⁸ Nguồn tốt nhất về các kỹ thuật ước lượng khác nhau là Terry E. Dielman, *Pooled Cross-Sectional and Time Series Data Analysis* (Phân tích số liệu chéo và chuỗi thời gian tổng hợp), Marcel Dekker, New York, 1989.

¹⁹ Để ước lượng hồi quy, nhập các quan sát của hai công ty. Tức là, 20 quan sát đầu thuộc G.M. và 20 quan sát cuối thuộc Westinghouse. Tạo biến giả và cho nó giá trị 1 đối với tất cả các quan sát G.M. và giá trị 0 đối với tất cả các quan sát Westinghouse.

* Không bắt buộc.

Từ Chương 6, ta giải thích hệ số β_2 là thay đổi tương đối (hay thay đổi phần trăm nếu thay đổi tương đối được nhân với 100) trong giá trị trung bình của Y khi X thay đổi đi một đơn vị. Như vậy, trong ví dụ này, nếu kinh nghiệm giảng dạy tăng lên một năm, thay đổi tương đối trong mức lương khởi điểm trung bình sẽ bằng β_2 . Một sự giải thích như vậy có thể áp dụng cho một thay đổi trong mọi giá trị của biến làm hồi quy, với điều kiện là biến được hồi quy là biến liên tục và không phân đôi như trong trường hợp biến giả. Nhưng ta có thể tính thay đổi tương đối của Y trung bình ngay cả đối với biến giả bằng công cụ do Halvorsen và Palmquist đề xuất:²⁰ Lấy đối lôgarít (antilog, cơ số e) của hệ số biến giả ước lượng và trừ cho 1.

Ví dụ 15.9 Hồi quy bán lôgarít với biến giả

Để minh họa, hãy xem xét số liệu trong Bảng 15.5 mô tả quan hệ giữa lương khởi điểm (Y) với số năm kinh nghiệm giảng dạy (X_2) và giới tính ($D = 1$ đối với giáo sư nam). Giả sử mô hình có dạng (15.13.1), ta tính được các kết quả sau:

$$\begin{aligned} \ln Y_i &= 2,9298 + 0,0546X_i + 0,134D_i \\ t &= (481,524) (48,3356) (27,2250) \\ R^2 &= 0,9958 \quad d = 2,51 \end{aligned} \quad (15.13.2)$$

BẢNG 15.5

Số liệu giả thiết về mức lương giáo sư đại học trong quan hệ với số năm kinh nghiệm giảng dạy

Lương khởi điểm, Y (nghìn USD)	Số năm kinh nghiệm giảng dạy, X_2	Giới tính (1 = nam) (0 = nữ)
23,0	1	1
19,5	1	0
24,0	2	1
21,0	2	0
25,0	3	1
22,0	3	0
26,5	4	1
23,1	4	0
25,0	5	0
28,0	5	1
29,5	6	1
26,0	6	0
27,5	7	0
31,5	7	1
29,0	8	0

Như các kết quả này cho thấy, giữ nguyên các yếu tố khác (ở đây là giới tính của giáo sư), lương trung bình tăng lên 5,46% hàng năm. Nhưng ta không thể nói rằng, giữ kinh nghiệm

²⁰ Robert Halvorsen & Raymond Palmquist, “The Interpretation of Dummy Variables in Semilogarithmic Equations”, *American Economic Review*, (Giải thích các biến giả trong các phương trình bán lôgarít, Tạp chí Kinh tế Hoa Kỳ), tập 70, số 3, 1980, trang 474-475.

giảng dạy không đổi, mức lương trung bình cao hơn 13,41% đối với giáo sư nam. Theo Halvorsen và Palmquist, ta tính antilog $0,1341 = 1,1435$. Trừ đi 1, ta có 0,1435 hay 14,35%; vậy lương trung bình của giáo sư nam cao hơn (so với giáo sư nữ) là 14,35%. Trong bài tập 15.33, bạn được yêu cầu so sánh các kết quả hồi quy trong (15.12.2) với các kết quả tính được từ mô hình tuyến tính.

Một phương pháp nữa để loại bỏ bias biến giả

Có một cách nữa để loại bỏ bias biến giả. Để tìm hiểu, ta hãy tiếp tục với mô hình (15.2.4) nhưng viết nó dưới dạng

$$Y_i = \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + u_i \quad (15.13.3)$$

với các biến giả được định nghĩa theo Phương trình (15.2.4). Lưu ý rằng trong (15.13.3) ta đã loại bỏ tung độ gốc α_1 . Bây giờ ta sẽ không rơi vào bias biến giả bởi vì ta không còn bị tác động bởi đa cộng tuyến hoàn hảo nữa. Điều này có thể thấy từ ma trận số liệu trình bày sau Phương trình (15.2.4) với cột 1 bị loại bỏ.

Lưu ý rằng từ kết quả của thay đổi này, ta cần phải giải thích α_2 và α_3 theo một cách khác. Chúng không còn là các hệ số tung độ gốc chênh lệch; bây giờ chúng là các ước lượng trực tiếp của các tung độ gốc trong các phạm trù khác nhau. Vậy, trong trường đang xem xét, với α_1 bị loại bỏ, α_2 sẽ cho biết giá trị tung độ gốc của hồi quy mức lương giáo sư nam và α_3 sẽ cho biết giá trị tung độ gốc của hồi quy mức lương giáo sư nữ. *Nhưng lưu ý rằng để ước lượng (15.13.3), ta sẽ phải sử dụng thủ tục ước lượng qua gốc tọa độ*, như đã thảo luận trong Chương 6. Tất nhiên, phần lớn các phần mềm thống kê đều có thể thực hiện việc này một cách tự động.

Quay trở lại với hồi quy (15.1.3), ta đã có thể ước lượng hồi quy

$$Y_i = \alpha_2 D_{2i} + \alpha_3 D_{3i} + u_i \quad (15.13.4)$$

với $D_{2i} = 1$ đối với giáo sư nam và 0 nếu khác, và $D_{3i} = 1$ đối với giáo sư nữ và 0 nếu khác. (Lưu ý: không có tung độ gốc trong hồi quy này).

Nếu tuân theo chiến lược này, bạn sẽ tính được các kết quả hồi quy sau:

$$\begin{aligned} Y_i &= 21,28D_{2i} + 18,00D_{3i} \\ se &= (0,3118) \quad (0,3118) \\ t &= (69,2556) \quad (57,7350) \quad R^2 = 0,8737 \end{aligned} \quad (15.13.5)$$

Các kết quả này giống như (15.1.3), nhưng dưới một hình thức khác.

Thông lệ chung là quy giá trị các biến giả theo cách mà nếu một biến có m phạm trù, ta chỉ đưa ra $(m - 1)$ biến giả. Lợi thế của cách làm này là thường thì ta muốn so sánh các kết quả của chúng ta theo một phạm trù tham chiếu. Ngoài ra, bằng cách giữ một tung độ gốc chung, ta có giá trị R^2 như thường lệ, trái lại với mô hình tung độ gốc bằng 0, R^2 quy ước không phải lúc nào cũng có ý nghĩa. Do vậy, ta sẽ tuân theo thông lệ chung.

Các biến giả và phương sai thay đổi

Hãy quay lại với ví dụ tiết kiệm - thu nhập của Anh Quốc mà ta đã thảo luận trong Mục 15.6. Bằng cách sử dụng kỹ thuật biến giả để kết hợp hai hồi quy (15.6.1) và (15.6.2) như trong (15.7.1), ta giả thiết ngầm rằng $\text{var}(u_{1i}) = \text{var}(u_{2i}) = \sigma^2$, tức là, phương sai thuần nhất. Nếu giả thiết này không thỏa mãn, tức là, nếu hai phương sai của sai số khác nhau, chúng ta sẽ có nhiều khả năng nhận thấy rằng hai tung độ gốc và hai hệ số góc không khác nhau về mặt thống kê mà

vẫn tìm ra rằng hệ số của biến giả trong hồi quy (15.7.1) có ý nghĩa thống kê.²¹ Do vậy, khi áp dụng kỹ thuật biến giả (hay kiểm định Chow cho vấn đề này), ta phải kiểm chứng rằng, trong một trường hợp cụ thể, ta không gặp phải vấn đề phương sai thay đổi. Nhưng tới giờ thì ta đã biết làm thế nào để giải quyết vấn đề này.

Các biến giả và tự tương quan

Xem xét mô hình sau có số liệu chuỗi thời gian:

$$Y_t = \beta_1 + \beta_2 D_t + \beta_3 X_t + \beta_4 (D_t X_t) + u_t \quad (15.13.6)$$

với $D_t = 0$ cho các quan sát trong giai đoạn đầu và 1 cho các quan sát trong giai đoạn thứ hai. Giả sử có n_1 quan sát trong giai đoạn đầu và n_2 quan sát trong giai đoạn thứ hai. Quan sát rằng (15.13.6), trong đó có tính tới các biến giả có tung độ gốc và độ dốc chênh lệch, chính là mô hình (15.7.1) mà ta đã từng nghiên cứu mối quan hệ tiết kiệm - thu nhập của Anh Quốc.

Giả sử tiếp theo cho rằng sai số u_t trong (15.13.6) được tạo bởi sơ đồ tự tương quan bậc 1 Markov, sơ đồ AR (1). Cụ thể,

$$u_t = \rho u_{t-1} + \varepsilon_t \quad (15.3.7)$$

với ε thỏa mãn các giả thiết chuẩn.

Từ Chương 12, ta đã biết cách biến đổi một mô hình hồi quy để loại bỏ tự tương quan (bậc nhất) (nhớ lại phương pháp sai phân tổng quát hóa): Giả sử ρ đã biết hay được ước lượng, ta sử dụng $(Y_t - \rho Y_{t-1})$ làm biến được hồi quy và $(X_t - \rho X_{t-1})$ làm biến làm hồi quy. Nhưng việc xuất hiện biến hồi quy thuộc loại biến giả D sẽ gây ra vấn đề đặc biệt: Lưu ý rằng biến giả chỉ đơn giản phân loại một quan sát thuộc giai đoạn thứ nhất hay thứ hai. Như vậy thì ta biến đổi nó như thế nào? Maddala đưa ra quy tắc sau:²²

1. Trong (15.13.6), các giá trị của D bằng 0 đối với tất cả các quan sát trong giai đoạn thứ nhất; trong giai đoạn thứ 2 giá trị của D đối với quan sát *đầu tiên* là $1/(1 - \rho)$ thay cho 1, và 1 đối với tất cả các quan sát khác.
2. Biến X_t được biến đổi là $(X_t - \rho X_{t-1})$. Lưu ý rằng ta mất một quan sát trong phép biến đổi này, trừ khi ta dùng **phép biến đổi Prais-Winsten**.
3. Giá trị của $D_t X_t$ bằng không đối với tất cả các quan sát trong giai đoạn thứ nhất (*Lưu ý: D_t bằng 0 trong giai đoạn thứ nhất*); trong giai đoạn thứ hai, *quan sát thứ nhất* lấy giá trị của $D_t X_t = X_t$ và các quan sát còn lại trong giai đoạn thứ hai được định là $(D_t X_t - D_t X_{t-1}) = (X_t - \rho X_{t-1})$. (*Lưu ý: Giá trị của D_t trong giai đoạn thứ hai là 1*).

Như phần thảo luận ở trên đã chỉ ra, *quan sát then chốt* là quan sát thứ nhất trong giai đoạn II. Nếu nó được thực hiện theo cách trình bày ở trên thì sẽ không có vấn đề gì trong việc ước lượng các hồi quy như (15.3.6) tùy theo sự tương quan như xác định trong (15.13.7).

²¹ Về vấn đề này, xem G. S. Maddala, *Introduction to Econometrics* (Giới thiệu kinh tế lượng), Macmillan, xuất bản lần thứ 2, New York, 1992, trang 320-322.

²² Maddala, *ibid.*, trang 321-322.

15.14 CÁC CHỦ ĐỀ CHO NGHIÊN CỨU TIẾP

Một số chủ đề liên quan tới các biến giả thảo luận trong sách vở có mức độ khá cao, bao gồm (1) **mô hình tham số ngẫu nhiên hoặc thay đổi**, (2) **mô hình hồi quy chuyển đổi** và (3) **mô hình bất cân bằng**.

Trong các mô hình hồi quy xem xét trong cuốn sách này, ta giả sử rằng các tham số, β , là những đối tượng chưa biết nhưng cố định. Các mô hình hệ số ngẫu nhiên – và có một số dạng của chúng – giả thiết các tham số β cũng có thể ngẫu nhiên. Một công trình lớn trong lĩnh vực này mà ta có thể tham khảo là của Swamy.²³

Trong mô hình biến giả sử dụng cả các tung độ độ chênh lệch và hệ số góc chênh lệch, ta giả sử ngầm rằng điểm gãy khúc đã biết. Vậy, trong hồi quy tiết kiệm - thu nhập của Anh Quốc, ta xác định thời gian 1946-1954 là thời kỳ tái thiết và 1955- là thời kỳ hậu tái thiết. Nhưng điều gì xảy ra nếu ta không biết điểm gãy khúc xảy ra vào năm 1955 hay 1954 hay 1956? Kỹ thuật **mô hình hồi quy chuyển đổi** giải quyết câu hỏi này bằng cách cho phép bản thân điểm gãy khúc là ngẫu nhiên. Công trình có ảnh hưởng trong lĩnh vực này là của Goldfelt và Quandt.²⁴

Ta cần phải có các kỹ thuật ước lượng đặc biệt để giải quyết cái gọi là **tình thế bất cân bằng**, tức là, các trường hợp mà thị trường không cân bằng (nghĩa là cầu không bằng cung). Ví dụ kinh điển là cầu và cung hàng hóa. Cầu hàng hóa là một hàm của giá và các biến khác. Cung hàng hóa là một hàm của giá và các biến khác, một số khác với các biến trong hàm cầu. Bây giờ, lượng hàng mua và bán trên thực tế có thể không nhất thiết bằng với mức tính được khi cho cung bằng cầu, và do vậy dẫn tới bất cân bằng. Người đọc có thể tham khảo Quandt về thảo luận toàn diện các **mô hình bất cân bằng**.²⁵

15.5 TÓM TẮT VÀ KẾT LUẬN

1. Các biến giả nhận các giá trị 1 và 0 (hay các biến đổi tuyến tính của chúng) là phương tiện để đưa các biến giải thích định tính vào phân tích hồi quy.
2. Các biến giả là một công cụ phân loại số liệu trong đó chúng chia một mẫu thành nhiều tiểu nhóm dựa vào các tính chất hay thuộc tính (giới tính, tình trạng hôn nhân, chủng tộc, tôn giáo, v.v...) và *ngầm* cho phép ta chạy các hồi quy riêng lẻ cho từng tiểu nhóm. Nếu có những khác biệt trong sự phản ứng của biến được hồi quy đối với biến thiên trong các biến định lượng trong các tiểu nhóm khác nhau, chúng sẽ được phản ánh trong những khác biệt về các tung độ gốc hay hệ số góc, hay cả hai, của các hồi quy tiểu nhóm khác nhau.
3. Mặc dù là một công cụ đa dụng, kỹ thuật biến giả cần phải được sử dụng một cách cẩn thận. Thứ nhất, nếu hồi quy có số hạng không đổi, số các biến giả cần phải nhỏ hơn số các phân loại trong từng biến định tính. Thứ hai, hệ số gắn với các biến giả phải *luôn luôn* được giải thích trong quan hệ với nhóm cơ sở hay nhóm tham chiếu, tức là nhóm nhận giá trị 0. Sau cùng, nếu một mô hình có vài biến định tính với một số phân loại, việc đưa ra các biến giả

²³ P.A. V. B. Swamy, *Statistical Inference in Random Coefficient Regression Models* (Suy luận thống kê trong các mô hình hồi quy hệ số ngẫu nhiên), Springer-Verlag, Berlin, 1971.

²⁴ S. Goldfelt & R. Quandt, *Nonlinear Methods in Econometrics* (Các phương pháp phi tuyến trong kinh tế lượng), North Holland, Amsterdam, 1972.

²⁵ Richard E. Quandt, *The Econometrics of Disequilibrium* (Kinh tế học của sự bất cân bằng), Basil Blackwell, New York, 1988.

có thể sử dụng một số lượng lớn các bậc tự do. Do vậy, ta phải luôn luôn so sánh số các biến giả đưa vào mô hình với tổng số các quan sát có sẵn trong phân tích.

4. Trong số các ứng dụng khác nhau của biến giả, chương này chỉ xem xét một vài nội dung, bao gồm (1) so sánh hai (hay nhiều) hồi quy, (2) loại bỏ yếu tố mùa trong số liệu chuỗi thời gian, (3) kết hợp số liệu chéo và chuỗi thời gian, (4) các mô hình hồi quy từng khúc.
5. Do các biến giả là phi ngẫu nhiên, chúng không gây ra vấn đề gì đặc biệt trong việc áp dụng OLS. Tuy nhiên, ta phải cẩn thận khi thực hiện biến đổi số liệu có chứa các biến giả. Đặc biệt, các vấn đề tự tương quan và phương sai thay đổi cần phải được xử lý cẩn thận.

BÀI TẬP

Câu hỏi

- 15.1. Nếu có số liệu hàng tháng trong một số năm, bạn phải đưa ra bao nhiêu biến giả để kiểm định các giả thiết sau:
 - (a) Tất cả 12 tháng trong năm đều có hình thái mùa;
 - (b) Chỉ có tháng 2, 4, 6, 8, 10 và 12 có hình thái mùa.
- 15.2. Tham chiếu hồi quy (15.5.1) trong đó giải thích yếu tố xác định mức lương theo giờ của những người làm thêm. Từ phương trình này tính các phương trình mức lương theo giờ cho các loại người làm thêm sau:
 - (a) Da trắng, không ở thành thị, ở miền Tây và đã tốt nghiệp trung học
 - (b) Không phải da trắng, ở thành thị, không ở miền Tây và chưa tốt nghiệp trung học
 - (c) Da trắng, không ở thành thị, không ở miền Tây và đã tốt nghiệp trung học.
- 15.3. Trong nghiên cứu tác động của một số đặc điểm định tính tới giá vé xem phim tại một đô thị lớn trong giai đoạn 1961-1964, R. D. Lampson đã tính được hồi quy sau cho năm 1961:^{*}

$$\begin{aligned}
 Y = & 4,13 + 5,77D_1 + 8,21D_2 - 7,68D_3 - 1,13D_4 \\
 & (2,04) \quad (2,67) \quad (2,51) \quad (1,78) \\
 & + 27,09D_5 + 31,46\log X_1 + 0,81X_2 + 3 \text{ biến giả khác} \\
 & (3,58) \quad (13,78) \quad (0,17) \\
 & R^2 = 0,961
 \end{aligned}$$

- với D_1 = vị trí rạp chiếu phim: 1 nếu là ngoại thành, 0 nếu là trung tâm thành phố
 D_2 = tuổi của rạp chiếu phim: 1 nếu dưới 10 năm từ khi xây dựng hay cải tạo lớn, 0 nếu khác
 D_3 = loại rạp: 1 nếu ngoài trời, 0 nếu trong nhà
 D_4 = chỗ đỗ xe: 1 nếu có, 0 nếu khác
 D_5 = chính sách chiếu phim: 1 nếu chiếu lần đầu, 0 nếu khác
 X_1 = tỷ lệ % trung bình số ghế không sử dụng trong một buổi chiếu
 X_2 = giá thuê phim trung bình, tính theo xu/vé do phía phân phối phim áp đặt
 Y = giá vé buổi tối cho người lớn, tính theo xu
 và các số trong ngoặc là sai số chuẩn.

^{*} R. D. Lampson, "Measured Productivity and Price Change: Some Empirical Evidence on Service Industry Bias, Motion Picture Theaters", *Journal of Political Economy* (Năng suất tính toán và thay đổi giá cả: một số bằng chứng thực nghiệm về thiên lệch ngành dịch vụ, Rạp chiếu phim, *Tạp chí Kinh tế Chính trị*), tập 78, 3-4/1970, trang 291-305.

- (a) Bình luận các kết quả.
 (b) Làm sao bạn hợp lý hóa việc đưa biến X_1 vào mô hình?
 (c) Làm sao bạn giải thích giá trị âm của hệ số D_4 ?
- 15.4. Dựa vào số liệu giai đoạn 1972-1979, William Nordhaus đã ước lượng mô hình sau để giải thích hành vi giá dầu lửa OPEC (sai số chuẩn trong ngoặc):*

$$y_t = 0,3x_{1t} + 5,22x_{2t} \\ (0,03) \quad (0,50)$$

với y_t = khác biệt giữa giá hiện tại và giá năm ngoái (USD/thùng)
 x_{1t} = khác biệt giữa giá hiện hành tại thị trường và giá của OPEC trong năm trước
 x_{2t} = 1 cho năm 1974 và 0 cho các năm khác.

(Lưu ý: 1973-1974 là năm cấm vận dầu lửa). Giải thích kết quả này và biểu diễn kết quả trên đồ thị cho các giai đoạn trước và sau cấm vận dầu lửa.

- 15.5. Giả sử rằng ta sửa đổi hồi quy lương giáo sư đại học (15.4.1) như sau:

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 (D_{2i} D_{3i}) + \beta X_i + u_i$$

với Y_i = mức lương hàng năm của một giáo sư đại học
 X_i = số năm kinh nghiệm giảng dạy
 $D_2 = 1$ nếu là nam và 0 nếu khác
 $D_3 = 1$ nếu là da trắng và 0 nếu khác

- (a) Số hạng $D_{2i} D_{3i}$ đại diện cho *tác động qua lại*. Biểu thức này có ý nghĩa gì?
 (b) Đây là ý nghĩa của hệ số α_4 ?
 (c) Tìm $E(Y_i | D_2 = 1, D_3 = 1, X_i)$ và giải thích nó.

- 15.6. **Các biến giả so với mã phân bổ** (allocated codes). Tham chiếu hồi quy (15.2.1). Thay cho việc áp dụng biến giả, giả sử ta sử dụng *mã phân bổ*:

$$D_i = 1 \text{ nếu là nữ} \\ = 2 \text{ nếu là nam}$$

- (a) Giải thích hồi quy sử dụng các mã phân bổ.
 (b) Lợi thế, nếu có, của việc sử dụng mã phân bổ đã định so với việc sử dụng biến giả 0-1 là gì?

- 15.7. Tiếp tục bài tập 15.6 nhưng bây giờ xem xét cách phân bổ sau:

$$D_i = 1 \text{ nếu là nữ} \\ = -1 \text{ nếu là nam}$$

Giải thích hồi quy sử dụng cách phân bổ này và so sánh các kết quả với phương pháp biến giả 0-1 thông thường.

- 15.8. Tham chiếu hồi quy (15.10.1). Bạn kiểm định giả thiết sau như thế nào

- (a) $\alpha_2 = \alpha_3$
 (b) $\alpha_2 = \alpha_4$

* “Oil and Economic Performance in Industrial Countries”, *Brookings Papers on Economic Activity* (Dầu lửa và kết quả kinh tế tại các nước công nghiệp, *Bài viết Brookings về hoạt động kinh tế*), 1980, trang 341-388.

- (c) Nếu $\alpha_2 \neq \alpha_1$ và $\alpha_3 \neq \alpha_1$ về mặt thống kê thì có phải điều nó nghĩa là $\alpha_2 \neq \alpha_3$? Gợi ý: $\text{var}(A + B) = \text{var}(A) + \text{var}(B) + 2 \text{cov}(A, B)$, và $\text{var}(A - B) = \text{var}(A) + \text{var}(B) - 2 \text{cov}(A, B)$.
- 15.9** (a) Làm thế nào bạn tính được các sai số của các hệ số hồi quy trong mô hình (15.8.3) và (15.8.4) ước lượng từ hồi quy “tổng hợp” (15.8.2)?
- (b) Để đưa ra các câu trả lời bằng số, cần phải thêm thông tin gì, nếu có?
- 15.10.** Như trình bày trong chương, các ước lượng của các hệ số hồi quy tính từ (15.7.1) sẽ đồng nhất với những hệ số tính từ ước lượng riêng lẻ của hai hồi quy (15.6.1) và (15.6.2). Điều đó có đúng cho $\hat{\sigma}^2$, ước lượng của giá trị đúng của phương sai σ^2 hay không, tức là $\hat{\sigma}^2$ tính được từ (15.7.1) có giống như giá trị tính được từ (15.6.1) hay (15.6.2) không? Tại sao hay tại sao không?
- 15.11. Kết hợp số liệu chéo và chuỗi thời gian.** Giả sử bạn có số liệu về sản lượng, nhập lượng lao động và vốn của N công ty trong một ngành kinh tế trong các thời đoạn T và giả sử bạn muốn thực hiện hàm hồi quy từ số liệu với dạng sau:

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + u_{it} \quad i = 1, 2, 3, \dots, N; \quad t = 1, 2, 3, \dots, T$$

với Y = sản lượng

X_1 = nhập lượng vốn

X_2 = nhập lượng lao động

Giả sử có các số liệu thích hợp, bạn được yêu cầu xây dựng các mô hình như sau

- (a) Các công ty khác nhau về *hiệu quả quản lý*, những khác biệt chỉ tác động đến tung độ gốc α ; tác động này có thể gọi là *tác động công ty*.
- (b) Tất cả các công ty có hiệu quả quản lý như sau, nhưng tung độ gốc α thay đổi năm này qua năm khác; tác động này có thể được gọi là *tác động năm*.
- (c) Tung độ gốc của hàm sản xuất trên bị ảnh hưởng bởi tác động công ty lẫn tác động năm. Thêm vào đó,
- (d) Bạn phải đưa ra giả thiết nào về yếu tố nhiễu u_{it} ?
- 15.12.** Trong nghiên cứu của mình về thời gian FDIC (Công ty Bảo hiểm Tiền gửi Liên bang) sử dụng lao động trong việc kiểm tra 91 ngân hàng, R. J. Miller đã ước lượng hàm số sau:

$$\begin{aligned} \widehat{\ln Y} = & 2,41 + 0,3674 \ln X_1 + 0,2217 \ln X_2 + 0,0803 \ln X_3 \\ & (0,0477) \quad (0,0628) \quad (0,0287) \\ & -0,1755 D_1 + 0,2799 D_2 + 0,5634 D_3 - 0,2572 D_4 \\ & (0,2905) \quad (0,1044) \quad (0,1657) \quad (0,0787) \\ & R^2 = 0,766 \end{aligned}$$

với Y = số giờ lao động của kiểm tra viên FDIC

X_1 = tổng tài sản của ngân hàng

X_2 = tổng số văn phòng của ngân hàng

X_3 = tỷ lệ các khoản vay đã phân loại so với tổng số các khoản vay của ngân hàng

* “Examination of Man-Hour Cost for Independent, Joint and Divided Examination Programs”, *Journal of Bank Research* (Kiểm tra chi phí phí lao động/giờ đối với các chương trình kiểm tra độ lập, liên kết và phân chia, *Tạp chí Nghiên cứu Ngân hàng*), tập 11, 1980, trang 28-35. Lưu ý: Các ký hiệu đã được thay đổi để phù hợp với các ký hiệu của chúng ta.

- $D_1 = 1$ nếu mức đánh giá quản lý là “tốt”
 $D_2 = 1$ nếu mức đánh giá quản lý là “trung bình”
 $D_3 = 1$ nếu mức đánh giá quản lý là “thỏa đáng”
 $D_4 = 1$ nếu kiểm tra được thực hiện cùng với bang.

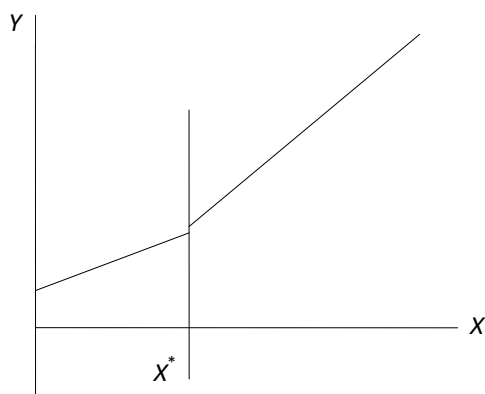
Các số trong ngoặc là sai số chuẩn ước lượng

- (a) Giải thích các kết quả.
 (b) Có xảy ra vấn đề gì hay không trong việc giải thích các biến giả trong mô hình này khi Y có dạng lôgarít?
 (c) Làm thế nào bạn giải thích các hệ số của biến giả?
- 15.13.** Để đánh giá tác động của chính sách của Hệ thống Dữ trữ Liên bang (Fed) trong việc nói lỏng kiểm soát lãi suất bắt đầu từ tháng 7 năm 1979, Sidney Langer, một sinh viên của tôi, đã ước lượng mô hình sau trong giai đoạn theo quý từ 1975-III đến 1983-II.*

$$\begin{aligned}
 Y_t = & 8,5871 - 0,1328P_t - 0,7102Un_t - 0,2389M_t \\
 & \text{se}(1,9563) \quad (0,0992) \quad (0,1909) \quad (0,0727) \\
 & + 0,6592Y_{t-1} + 2,5831Dum_t \quad R^2 = 0,9156 \\
 & (0,1036) \quad (0,7549)
 \end{aligned}$$

- với Y = mức lãi suất tín phiếu kho bạc kỳ hạn 3 tháng
 P = tỷ lệ lạm phát dự kiến
 Un = tỷ lệ thất nghiệp đã hiệu chỉnh yếu tố mùa
 M = những thay đổi trong cơ sở tiền tệ
 Dum = biến giả, lấy giá trị 1 đối với các quan sát bắt đầu từ 1/7/1979.

- (a) Giải thích các kết quả này.
 (b) Đây là tác động của chính sách nói lỏng kiểm soát lãi suất? Các kết quả có ý nghĩa kinh tế không?
 (c) Các hệ số của P_t , Un và M_t nhỏ hơn 0. Bạn có thể cho biết lý do kinh tế căn bản không?
- 15.14.** Tham chiếu hồi quy từng khúc thảo luận trong chương. Giả sử không chỉ có thay đổi trong hệ số góc tại X^* mà đường hồi quy cũng dịch chuyển, như mô tả trong Hình 15.10. Bạn chỉnh (15.11.1) như thế nào để tính đến sự dịch chuyển trong đường hồi quy tại X^* ?



HÌNH 15.10

Hồi quy tuyến tính từng khúc không liên tục.

* Sidney Langer, “Interest Rate Deregulation and Short-Term Interest Rates” (Nói lỏng kiểm soát lãi suất và các mức lãi suất ngắn hạn), bài viết cuối học kỳ không xuất bản.

- 15.15.** Các yếu tố xác định giá 1 ounce cola. Cathy Schaefer, một sinh viên của tôi, đã ước lượng hồi quy sau dựa vào số liệu chéo của 77 quan sát:¹

$$P_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \mu_i$$

- với P_i = giá 1 ounce cola
 D_{1i} = 001 nếu là cửa hàng bán giảm giá
 = 010 nếu là cửa hàng chuỗi xích (trong một hệ thống cửa hàng)
 = 100 nếu là cửa hàng tự chọn
 D_{2i} = 10 nếu là sản phẩm có nhãn
 = 01 nếu là sản phẩm không có nhãn
 D_{3i} = 0001 chai 67,6 ounce (2 lít)
 = 0010 chai 33,8 ounce (*Lưu ý*: 33,8 ounce = 1 lít)
 = 0100 chai 16 ounce
 = 1000 chai 12 ounce.

Sau đây là các kết quả:

$$P_i = 0,0143 - 0,000004D_{1i} + 0,0090D_{2i} + 0,00001D_{3i}$$

$$(0,00001) \quad (0,00011) \quad (0,00000)$$

$$t = (-0,3837) \quad (8,3927) \quad (5,8125)$$

$$R^2 = 0,6033$$

Lưu ý: Các sai số chuẩn chỉ được biểu diễn với 5 chữ số sau dấu phẩy.

- (a) Bình luận cách thức đưa các biến giả vào mô hình.
 (b) Giả sử cách thiết lập biến giả là chấp nhận được, bạn giải thích các kết quả như thế nào?
 (c) Hệ số của D_3 dương và có ý nghĩa thống kê. Bạn hợp lý hóa kết quả này như thế nào?
- 15.16.** Dựa vào số liệu của 101 nước về thu nhập bình quân đầu người tính theo USD (X) và tuổi thọ tính theo năm (Y) trong đầu thập niên 70, Sen và Srivastava đã tính được các kết quả hồi quy sau:^{*}

$$Y_i = -2,40 + 9,39 \ln X_i - 3,36[D_i(\ln X_i - 7)]$$

$$se = (4,73) \quad (0,859) \quad (2,42) \quad R^2 = 0,752$$

với $D_i = 1$ nếu $\ln X_i > 7$ và $D_i = 0$ nếu khác. *Lưu ý*: Khi $\ln X_i = 7$, $X = 1097$ USD (gần đúng).

- (a) Với (các) lý do gì để thiết lập biến thu nhập dưới dạng log?
 (b) Bạn giải thích hệ số 9,39 của $\ln X_i$ như thế nào?
 (c) Đâu có thể là lý do để đưa ra biến làm hồi quy $D_i(\ln X_i - 7)$? Làm sao bạn giải thích hồi quy này bằng lời (*Gợi ý*: hồi quy tuyến tính từng khúc)?
 (d) Giả sử thu nhập bình quân đầu người 1097 USD là đường chia cắt giữa nước nghèo và nước giàu, bạn tính hồi gán cho các nước có thu nhập dưới 1097 USD và hồi gán cho các nước có thu nhập trên 1097 USD như thế nào?
 (e) Bạn rút ra các kết luận tổng quát nào từ kết quả hồi quy trình bày trong bài tập này?

¹ Cathy Schaefer, "Price Per Ounce of Cola Beverage as a Function of Place of Purchase, Size of Container, and Branded or Unbranded Product" (Giá 1 ounce đồ uống cola biểu diễn bằng hàm số của nơi mua, kích thước côngtenơ và sản phẩm có nhãn hiệu hay không có), đề án cuối học kỳ không xuất bản.

^{*} Ashish Sen & Muni Srivastava, *Regression Analysis: Theory, Methods, and Applications* (Phân tích hồi quy: Lý thuyết, Phương pháp và Ứng dụng), Springer-Verlag, New York, 1990, trang 92. Ký hiệu đã được thay đổi.

15.17. Xem xét mô hình sau:

$$Y_i = \beta_1 + \beta_2 D_i + u_i$$

với $D_i = 0$ đối với 20 quan sát đầu và $D_i = 1$ đối với 30 quan sát còn lại. Bạn cũng được cho biết rằng $\text{var}(u_i^2) = 300$.

- (a) Bạn giải thích β_1 và β_2 như thế nào?
- (b) Các giá trị trung bình của hai nhóm bằng bao nhiêu?
- (c) Bạn tính phương sai của $(\beta_1 + \beta_2)$ như thế nào? *Lưu ý:* Bạn được cho biết rằng $\text{cov}(\beta_1, \beta_2) = -15$.

Bài tập

- 15.18. Sử dụng số liệu trong Phụ lục 15A, Mục 15A.1, kiểm định giả thiết rằng các phương sai của sai số trong hai giai đoạn 1958-IV đến 1966-III và 1966-IV đến 1971-II bằng nhau. Xem Chương 11 về các phương pháp khác nhau trong kiểm định tính thuần nhất (không đổi) của phương sai.
- 15.19. Kiểm định giả thiết cho rằng σ_2^2 và σ_3^2 ước lượng từ (8.8.6) và (8.8.7) bằng nhau. Bạn có thể sử dụng kiểm định của Barlett về tính thuần nhất của phương sai đã thảo luận trong Chương 11.
- 15.20. **Kiểm định Chow đã chỉnh sửa** (khi các quan sát nhỏ hơn số các tham số cần ước lượng). Tham chiếu các hồi quy (15.6.1) và (15.6.2). Giả thiết rằng n_2 , số các quan sát trong giai đoạn II, nhỏ hơn hay cùng lắm là bằng số các tham số cần ước lượng. Trong trường hợp này, Chow đưa ra chỉnh sửa sau về kiểm định của mình: Gọi $S_1 = \text{RSS}$ từ hồi quy tổng hợp; $S_2 = \text{RSS}$ từ hồi quy trong giai đoạn I (giả thiết rằng $n_1 > \text{số các tham số}$). Bây giờ, sử dụng kiểm định F sau:

$$F = \frac{(S_1 - S_2) / n_2}{S_2 / (n_1 - k)}$$

trong đó có n_2 và $(n_1 - k)$ bậc tự do.

Nếu giá trị F có ý nghĩa thống kê, ta bác bỏ giả thiết cho rằng n_2 quan sát cuối đến từ mô hình đã tạo ra hồi quy trong giai đoạn I dựa vào n_1 quan sát. Nếu nó không có ý nghĩa, ta không bác bỏ giả thiết này.

Sử dụng số liệu trong Bảng 15.2 để kiểm định giả thiết cho rằng hai quan sát cuối đến từ cùng một tổng thể đã tạo ra 16 quan sát đầu.

- 15.21. Sử dụng phương pháp luận thảo luận trong Chương 8, so sánh các hồi quy hạn chế và không hạn chế (15.10.2) và (15.10.3), tức là, kiểm định giá trị của các hạn chế được áp đặt.
- 15.22. Tham chiếu các số liệu trong Phụ lục 15A, Mục 15A.2 và hồi quy (15.10.2). Xây dựng một mô hình hồi quy để kiểm định giả thiết cho rằng độ dốc cũng như tung độ gốc của hồi quy lợi nhuận theo doanh thu trong quý II của năm khác với các quý còn lại. Trình bày các phép tính cần thiết.
- 15.23. **Loại bỏ yếu tố mùa trong số liệu.** Ví dụ minh họa trong Mục 15.10 đã chỉ ra các biến giả có thể được sử dụng như thế nào để tính đến các tác động của mùa. Sau khi ước lượng hồi quy (15.10.2), ta tìm thấy rằng chỉ có biến giả gắn với quý II có tác động mùa. Do vậy, phương pháp loại bỏ yếu tố mùa khỏi số liệu sẽ trừ đi giá trị 1.322,8938 (triệu USD), giá trị của hệ số của biến giả trong quý II, từ các số liệu lợi nhuận và doanh thu

trong quý II của mỗi năm và chạy hồi quy lợi nhuận theo doanh thu sử dụng số liệu vừa biến đổi.

- (a) Biến đổi số liệu ở trên và chạy hồi quy. Đừng đưa vào hồi quy này bất cứ biến giả nào. (Tại sao?)
- (b) So sánh hệ số của biến doanh thu trong hồi quy ước lượng bằng cách sử dụng số liệu đã biến đổi với hệ số trong (15.10.2). Hai hệ số này có được mong đợi là đồng nhất về mặt thống kê hay không? Tại sao?
- (c) Giả sử trong (a) bạn đã đưa vào mô hình các biến giả. Điều gì sẽ xảy ra đối với các hệ số của các biến giả?
- 15.24. Trong hồi quy tiết kiệm - thu nhập (15.7.4), giả sử rằng thay cho việc sử dụng các giá trị 1 và 0 cho biến giả D_i , bạn sử dụng $Z_i = a + bD_i$, với $D_i = 1$ và 0, $a = 2$, $b = 3$. So sánh hai kết quả.
- 15.25. Tiếp tục với hồi quy tiết kiệm - thu nhập (15.7.4), giả sử bạn cho $D_i = 1$ đối với các quan sát trong giai đoạn II và $D_i = 0$ đối với các quan sát trong giai đoạn I. Các kết quả trong (15.7.4) thay đổi như thế nào?
- 15.26. Làm thế nào bạn tính được các sai số chuẩn của các hệ số ước lượng đối với hồi quy (15.7.5) và (15.7.6)? Bạn cần phải biết thêm thông tin gì, nếu có, để tính ra các kết quả bằng số?
- 15.27. Bảng sau cho biết số liệu hàng quý (chưa hiệu chỉnh theo mùa) về doanh thu bán cổ phần quỹ hỗ tương của ngành quỹ hỗ tương (quỹ đầu tư chung) trong giai đoạn 1968-1973.

**Doanh thu bán cổ phần quỹ hỗ tương
(triệu USD)**

Năm	Quý			
	I	II	III	IV
1968	1.564	1.65 4	1.60 7	1.994
1969	2.129	1.65 8	1.42 8	1.503
1970	1.381	1.03 9	975	1.230
1971	1.304	1.28 8	1.10 8	1.446
1972	1.398	1.17 6	1.09 9	1.219
1973	1.382	888	933	1.156

Nguồn: 1974 Mutual Fund Fact Book (Sách Sự kiện về Quỹ hỗ tương 1974), Viện Công ty Đầu tư, Washington, D.C. (Số liệu được làm tròn theo triệu đô la).

Xem xét mô hình sau:

$$\text{Doanh thu}_t = \alpha_1 + \alpha_2 D_2 + \alpha_3 D_3 + \alpha_4 D_4 + u_t$$

với $D_2 = 1$ đối với quý II, 0 nếu khác

$D_3 = 1$ đối với quý III, 0 nếu khác

$D_4 = 1$ đối với quý IV, 0 nếu khác

- Ước lượng hồi quy trên.
- Bạn giải thích các hệ số α như thế nào?
- Làm thế nào bạn sử dụng các hệ số α ước lượng để loại bỏ yếu tố mùa trong số liệu doanh thu?

15.28. Sử dụng số liệu của bài tập 15.27 nhưng với mô hình sau:

$$\text{Doanh thu}_t = \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \alpha_4 D_4 + u_t$$

với D là các biến giả nhận giá trị 1 hoặc 0 trong các quý I đến IV.

- Bạn ước lượng phương trình trên như thế nào?
- Phương trình trên có vi phạm quy tắc số biến giả phải bằng số phân loại (quý) trừ 1 hay không?

* 15.29. **Hồi quy có vẻ như không có quan hệ** (seemingly unrelated regressions - SURE).

Tham chiếu các hàm đầu tư G.M. và Westinghouse trong (15.12.1) và (15.12.2). Mặc dù các hàm này có thể được ước lượng riêng rẽ, sẽ hiệu quả hơn (về mặt thống kê) nếu các hàm này cùng được ước lượng, do hai công ty hoạt động trong cùng một thị trường vốn và một cú sốc (ví dụ như gia tăng lãi suất) sẽ có nhiều khả năng tác động đến cả hai công ty. Do vậy, có nhiều khả năng là số hạng sai số u đối với G.M. và số hạng sai số u' đối với Westinghouse tại cùng một thời điểm sẽ có tương quan (trường hợp này được gọi là **tương quan cùng thời điểm**). Trong trường hợp này, Zellner đã chỉ ra rằng ước lượng hai phương trình đồng thời, mặc dù chúng có vẻ như không có quan hệ, có thể cải thiện tính hiệu quả của các ước lượng so với trường hợp từng hàm được ước lượng riêng rẽ.¹ Do vậy, nó được viết tắt là SURE.

Phương pháp thực tế để ước lượng SURE khá phức tạp, nhưng các phần mềm kinh tế lượng tiêu chuẩn đều có chức năng SURE.

- Sử dụng bất cứ phần mềm kinh tế lượng nào, tìm các ước lượng SURE của các tham số của hai hàm đầu tư.
- Tính cả các ước lượng OLS của hai hàm hồi quy riêng rẽ.
- So sánh các sai số chuẩn của các hệ số hồi quy khác nhau tính theo phương pháp OLS và SURE. Phương pháp nào tốt hơn và tại sao?

15.30. Tham chiếu ví dụ 15.5 trong Mục 15.8. Áp dụng kiểm định Chow cho số liệu trong mục này và so sánh các kết quả của bạn với các kết quả trong Ví dụ 15.5. Bạn thích phương pháp nào? Và tại sao?

* Không bắt buộc.

¹ A. Zellner, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", *Journal of the American Statistical Association* (Một phương pháp ước lượng hiệu quả các hồi quy có vẻ như không có quan hệ và kiểm định sự thiên lệch tổng hợp, *Tạp chí của Hiệp hội Thống kê Hoa Kỳ*), tập 57, 1962, trang 348-368.

15.31. Sử dụng số liệu đầu tư trong Bảng 15.4, chạy hồi quy tổng hợp cho Westinghouse và G.E sử dụng mô hình (15.12.3) và so sánh các kết quả của bạn với hồi quy tổng hợp G.M. và Westinghouse. Nếu có sự khác biệt trong các kết quả của bạn, thì điều gì có thể là lý do cho sự khác biệt này?

Bạn chạy hồi quy tổng hợp cho G.M., G.E. và Westinghouse sử dụng kỹ thuật biến giả như thế nào?

15.32. Sử dụng số liệu trong Bảng 15.4, ước lượng các hàm hồi gán cho G.M và Westinghouse với cả các biến giả tung độ gốc và hệ số góc chênh lệch để phân biệt hai công ty và bình luận các kết quả của bạn đối chiếu với các kết quả trong (15.12.5).

15.33. So sánh các kết quả của hồi quy (15.13.2) với các kết quả tính được từ việc chạy mô hình tuyến tính, tức là, với Y là hàm tuyến tính của X_2 và D_i .

PHỤ LỤC 15A

15A.1 MA TRẬN SỐ LIỆU CHO HỒI QUY (15.8.2)

Năm và quý	Tỷ lệ thất nghiệp UN (%)	Tỷ lệ việc làm còn trống V (%)	D	DV	Năm và quý	Tỷ lệ thất nghiệp UN (%)	Tỷ lệ việc làm còn trống V (%)	D	DV
1958-IV	1,915	0,510	0	0	1965-I	1,201	0,997	0	0
1959-I	1,876	0,541	0	0	-II	1,192	1,035	0	0
-II	1,842	0,541	0	0	-III	1,259	1,040	0	0
-III	1,750	0,690	0	0	-IV	1,192	1,086	0	0
-IV	1,648	0,771	0	0	1966-I	1,089	1,101	0	0
1960-I	1,450	0,836	0	0	-II	1,101	1,058	0	0
-II	1,393	0,908	0	0	-III	1,243	0,987	0	0
-III	1,322	0,968	0	0	-IV	1,623	0,819	1	0,819
-IV	1,260	0,998	0	0	1967-I	1,821	0,740	1	0,740
1961-I	1,171	0,968	0	0	-II	1,990	0,661	1	0,661
-II	1,182	0,964	0	0	-III	2,114	0,660	1	0,660
-III	1,221	0,952	0	0	-IV	2,115	0,698	1	0,698
-IV	1,340	0,849	0	0	1968-I	2,150	0,695	1	0,695
1962-I	1,411	0,748	0	0	-II	2,141	0,732	1	0,732
-II	1,600	0,658	0	0	-III	2,167	0,749	1	0,749
-III	1,780	0,562	0	0	-IV	2,107	0,800	1	0,800
-IV	1,941	0,510	0	0	1969-I	2,104	0,783	1	0,783
1963-I	2,178	0,510	0	0	-II	2,056	0,800	1	0,800
-II	2,067	0,544	0	0	-III	2,170	0,794	1	0,794
-III	1,942	0,568	0	0	-IV	2,161	0,790	1	0,790
-IV	1,764	0,677	0	0	1970-I	2,225	0,757	1	0,757
1964-I	1,532	0,794	0	0	-II	2,241	0,746	1	0,746

-II	1,455	0,838	0	0	-III	2,366	0,739	1	0,739
-III	1,409	0,885	0	0	-IV	2,324	0,707	1	0,707
-IV	1,296	0,978	0	0	1971-I	2.516*	0.583*	1	0.583
									*
					-II	2.909*	0.524*	1	0.524
									*

* Các ước lượng sơ khởi.

Nguồn: Damodar Gujarati, “The Behaviour of Unemployment and Unfilled Vacancies: Great Britain, 1958-1971”, *The Economic Journal* (Hành vi của thất nghiệp và chỗ làm việc còn trống: Anh Quốc, 1958-1971, *Tạp chí Kinh tế*), tập 82, 3/1972, trang 195-202.

15A.2 MA TRẬN SỐ LIỆU CHO HỒI QUY (15.10.2)

Năm và quý	Lợi nhuận (triệu USD)	Doanh thu (triệu USD)	D_2	D_3	D_4
1965-I	10.503	114.862	0	0	0
-II	12.092	123.968	1	0	0
-III	10.834	121.454	0	1	0
-IV	12.201	131.917	0	0	1
1966-I	12.245	129.911	0	0	0
-II	14.001	140.976	1	0	0
-III	12.213	137.828	0	1	0
-IV	12.820	145.465	0	0	1
1967-I	11.349	136.989	0	0	0
-II	12.615	145.126	1	0	0
-III	11.014	141.536	0	1	0
-IV	12.730	151.776	0	0	1
1968-I	12.539	148.862	0	0	0
-II	14.849	158.913	1	0	0
-III	13.203	155.727	0	1	0
-IV	14.947	168.409	0	0	1
1969-I	14.151	162.781	0	0	0
-II	15.949	176.057	1	0	0
-III	14.024	172.419	0	1	0
-IV	14.315	183.327	0	0	1
1970-I	12.381	170.415	0	0	0
-II	13.991	181.313	1	0	0
-III	12.174	176.712	0	1	0
-IV	10.985	180.370	0	0	1

Lưu ý: $D_2 = 1$ đối với quý II, 0 nếu khác

$D_3 = 1$ đối với quý III, 0 nếu khác

$D_4 = 1$ đối với quý IV, 0 nếu khác

Nguồn: Số liệu về lợi nhuận và doanh thu là cho toàn bộ trung tâm công nghiệp chế tạo và được lấy từ *Báo cáo tài chính hàng quý đối với các công ty công nghiệp chế tạo*, Ủy ban Thương mại Liên bang Hoa Kỳ và Ủy ban Chứng khoán và Giao dịch Hoa Kỳ