

Chương 6

LỰA CHỌN DẠNG HÀM SỐ VÀ KIỂM ĐỊNH ĐẶC TRƯNG MÔ HÌNH

Trong Chương 4 và 5 chúng ta đã nghiên cứu sự hồi quy bội trong đó biến phụ thuộc đang quan tâm (Y) quan hệ với nhiều biến độc lập (X s). Sự lựa chọn các biến độc lập sẽ dựa theo lý thuyết kinh tế, trực giác, kinh nghiệm quá khứ, và những nghiên cứu khác. Để tránh sự thiên lệch của biến bị loại bỏ như đã thảo luận trước đây; nhà nghiên cứu thường thêm vài biến giải thích mà ngờ rằng có ảnh hưởng đến biến phụ thuộc. Tuy nhiên; mối quan hệ giữa Y và các biến X nghiên cứu cho đến giờ vẫn giả sử là tuyến tính. Đây hiển nhiên là ràng buộc nghiêm ngặt và không thực tế trên một mô hình. Trong ứng dụng Phần 3.11, chúng ta lưu ý rằng biểu đồ phân tán quan sát được giữa số lượng bản quyền phát hành và chi phí nghiên cứu phát triển (Hình 3.11) cho thấy mối quan hệ theo đường cong. Ta thấy rằng giả thiết tuyến tính đã cho dự đoán xấu trong vài năm. Bên cạnh các sự việc quan sát thực nghiệm của dạng này, thường còn có những lý lẽ lý thuyết tốt cho việc xem xét các dạng hàm tổng quát của mối quan hệ giữa các biến phụ thuộc và độc lập. Ví dụ, lý thuyết kinh tế cho chúng ta biết rằng đường cong chi phí trung bình có dạng chữ U, và do vậy giả thiết tuyến tính là đáng ngờ nếu ta muốn ước lượng đường cong chi phí trung bình.

Trong chương này, chúng ta khảo sát một cách chi tiết đáng kể các cách thành lập và ước lượng các quan hệ phi tuyến. Để có thể vẽ các đồ thị, nhiều cách trình bày chỉ giải quyết duy nhất một biến giải thích. Đây chỉ đơn thuần là một phương cách mang tính sư phạm. Trong các ví dụ và ứng dụng chúng ta sẽ giảm nhẹ ràng buộc này.

Chương này cũng thảo luận vài phương pháp tiến hành các kiểm định đặc trưng mô hình chính thức. Đặc biệt, các phương pháp “tổng quát đến đơn giản” và “đơn giản đến tổng quát” được đề cập trong Chương 1 sẽ được thảo luận, và gọi là thủ tục Ramsey’s RESET (1969).

● 6.1 Ôn Lại Các Hàm Logarit và Hàm Mũ

Các hàm mũ và logarit là hai trong số các hàm được dùng phổ biến nhất trong lập mô hình. Vì lý do này, sẽ hữu ích khi ôn lại những tính chất cơ bản của các hàm này trước khi sử dụng chúng.

Hàm $Y = a^X$ ($a > 0$) là một ví dụ của một hàm mũ. Trong hàm này, a là **cơ số** của hàm và X là **số mũ**. Trong toán học, cơ số thông thường nhất dùng trong một hàm mũ là hằng số toán học e được xác định bởi

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n} \right)^n = 2,71828...$$

Vậy **hàm mũ** chuẩn có dạng $Y = e^X$, và cũng được viết dưới dạng $\exp(X)$. Hàm nghịch của hàm mũ gọi là **hàm logarit**. Logarit cơ số a cho trước (phải là số dương) của một số được định nghĩa là khi lũy thừa logarit của cơ số sẽ cho chính số đó. Ta viết $X = \log_a Y$. Ví dụ, vì $32 = 2^5$, logarit cơ số 2 của 32 là 5. Logarit cơ số e được gọi **logarit tự nhiên** và ký hiệu là $Y = \ln X$, mà không

cần ghi rõ cơ sở. Lưu ý rằng $\ln 1 = 0$ bởi vì $e^0 = 1$. Một số tính chất của hàm mũ và logarit được liệt kê dưới đây.

Tính chất 6.1

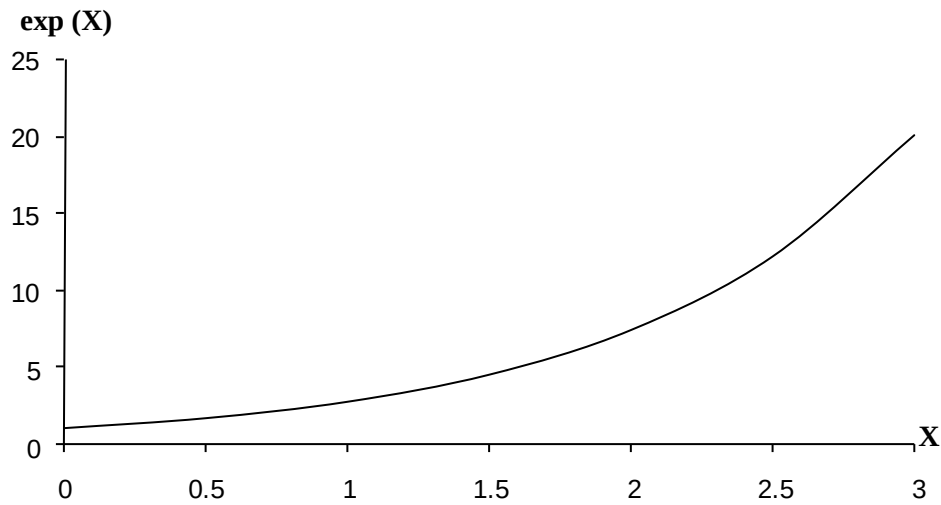
- Hàm logarit và hàm mũ là đơn điệu tăng; nghĩa là, nếu $a > b$, thì $f(a) > f(b)$, và ngược lại.
- Logarit của tích hai số bằng tổng logarit; nghĩa là, $\ln(XY) = \ln X + \ln Y$. Cũng vậy, logarit của tỷ số là hiệu của các logarit. Vậy, $\ln(X/Y) = \ln X - \ln Y$. Theo đó $\ln(1/X) = -\ln X$.
- $\ln(a^X) = X \ln a$. Theo đó $a^X = e^{X \ln a}$.
- $a^X a^Y = a^{X+Y}$ và $(a^X)^Y = a^{XY}$.

Không như đường thẳng, có độ dốc không đổi, hàm số tổng quát $f(X)$, như hàm mũ và logarit, có độ dốc thay đổi. Sự thay đổi của Y theo thay đổi đơn vị của X là tác động cận biên của X lên Y và thường ký hiệu bởi $\Delta Y / \Delta X$ (xem Hình 2.A và phần thảo luận liên quan). Nếu sự thay đổi của X vô cùng nhỏ, ta có độ dốc của tiếp tuyến của đường cong $f(X)$ tại điểm X . Độ dốc giới hạn này được xem là đạo hàm của Y đối với X và được ký hiệu bởi dY/dX . Vậy đạo hàm là tác động cận biên của X lên Y với sự thay đổi rất nhỏ của X . Đó là một khái niệm vô cùng quan trọng trong kinh tế lượng, bởi vì ta luôn hỏi sự thay đổi kỳ vọng của biến phụ thuộc là gì khi ta thay đổi giá trị của một biến độc lập với một lượng rất nhỏ. Các tính chất của các đạo hàm được tóm tắt trong Tính chất 2.A.5 và đáng để nghiên cứu. Tính chất 6.2 liệt kê một ít tính chất của hàm mũ và logarit mà rất hữu ích trong kinh tế lượng. Hình 6.1 minh họa bằng đồ thị hai hàm số này.

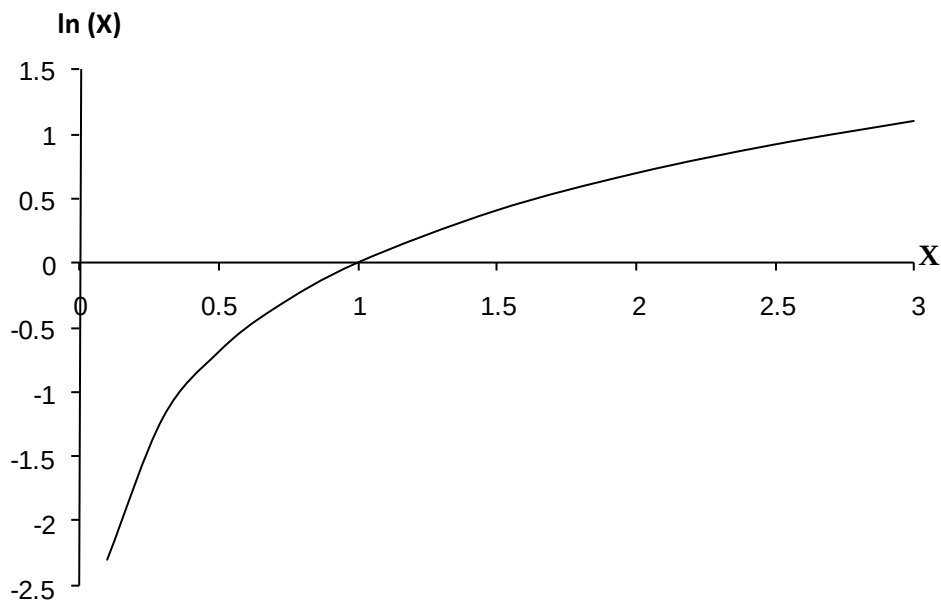
Tính chất 6.2

- Hàm mũ với cơ số e có tính chất đặc biệt là nó bằng với đạo hàm của chính nó. Vậy, nếu $Y = e^X$, thì $dY/dX = e^X$.
- Đạo hàm của e^{aX} là ae^{aX} .
- Đạo hàm của $\ln X$ bằng $1/X$.
- Đạo hàm của a^X bằng $a^X \ln a$. Kết quả này có được từ cơ sở là $a^X = e^{X \ln a}$ và tính chất đạo hàm của $e^{bX} = be^{bX}$.

● Hình 6.1 Đồ Thị của Hàm Mũ và Logarit



a. Đồ thị của $Y = \exp(X)$



b. Đồ thị của $Y = \ln(X)$

Khái Niệm của Độ Co Giãn

Logarit có tương quan rất gần với khái niệm của độ co giãn được dùng trong kinh tế. Ta sẽ thấy trong các phần sau rằng khái niệm này cũng được sử dụng rộng rãi trong kinh tế lượng thực nghiệm. Theo thuật ngữ đơn giản, độ co giãn của Y đối với X được định nghĩa là *phần trăm thay đổi của Y đối với một phần trăm thay đổi của X* cho một thay đổi nhỏ của X . Vậy nếu ΔY là sự thay đổi của Y , phần trăm thay đổi là $100\Delta Y/Y$. Tương tự, $100\Delta X/X$ là phần trăm thay đổi của X . Tỷ số của số đầu đối với số sau là độ co giãn. Điều này đưa đến định nghĩa sau.

● Bảng 6.1 Các Tác Động Cận Biên và Độ Co Giãn của các Dạng Hàm Khác Nhau

Tên	Dạng Hàm	Tác Động Cận Biên (dY/dX)	Độ Co Giãn [(X/Y)(dY/dX)]
Tuyến tính	$Y = \beta_1 + \beta_2 X$	β_2	$\beta_2 X/Y$
Logarit – tuyến tính	$Y = \beta_1 + \beta_2 \ln X$	β_2/X	β_2/Y
Nghịch đảo	$Y = \beta_1 + \beta_2 (1/X)$	$-\beta_2/X^2$	$-\beta_2/(XY)$
Bậc hai	$Y = \beta_1 + \beta_2 X + \beta_3 X^2$	$\beta_2 + 2\beta_3 X$	$(\beta_2 + 2\beta_3 X)X/Y$
Tương tác	$Y = \beta_1 + \beta_2 X + \beta_3 XZ$	$\beta_2 + \beta_3 Z$	$(\beta_2 + \beta_3 Z)X/Y$
Tuyến tính-logarit	$\ln Y = \beta_1 + \beta_2 X$	$\beta_2 Y$	$\beta_2 X$
Nghịch đảo – logarit	$\ln Y = \beta_1 + \beta_2 (1/X)$	$-\beta_2 Y/X^2$	$-\beta_2/X$
Bậc hai – logarit	$\ln Y = \beta_1 + \beta_2 X + \beta_3 X^2$	$Y(\beta_2 + 2\beta_3 X)$	$X(\beta_2 + 2\beta_3 X)$
Log-hai lần (log-log)	$\ln Y = \beta_1 + \beta_2 \ln X$	$\beta_2 Y/X$	β_2
Logistic	$\ln \left[\frac{Y}{1-Y} \right] = \beta_1 + \beta_2 X$	$\beta_2 Y(1-Y)$	$\beta_2(1-Y)X$

ĐỊNH NGHĨA 6.1

Độ co giãn của Y đối với X (ký hiệu là η) là

$$\eta = \frac{\Delta Y}{Y} \div \frac{\Delta X}{X} = \frac{X}{Y} \frac{\Delta Y}{\Delta X} \rightarrow \frac{X}{Y} \frac{dY}{dX} \text{ khi } \Delta X \text{ tiến về } 0. \quad (6.1)$$

Bảng 6.1 có các tác động ứng cận biên (dY/dX) và độ co giãn [(X/Y)(dY/dX)] của một số dạng hàm có thể chọn lựa trong chương này. Lưu ý rằng đôi khi các kết quả này phụ thuộc vào X và/hoặc Y. Để tính toán chúng, người ta thường thay thế giá trị trung bình $\bar{X}$ và giá trị dự đoán tương ứng $\hat{Y}$.

● 6.2 Quan Hệ Logarit-Tuyến Tính

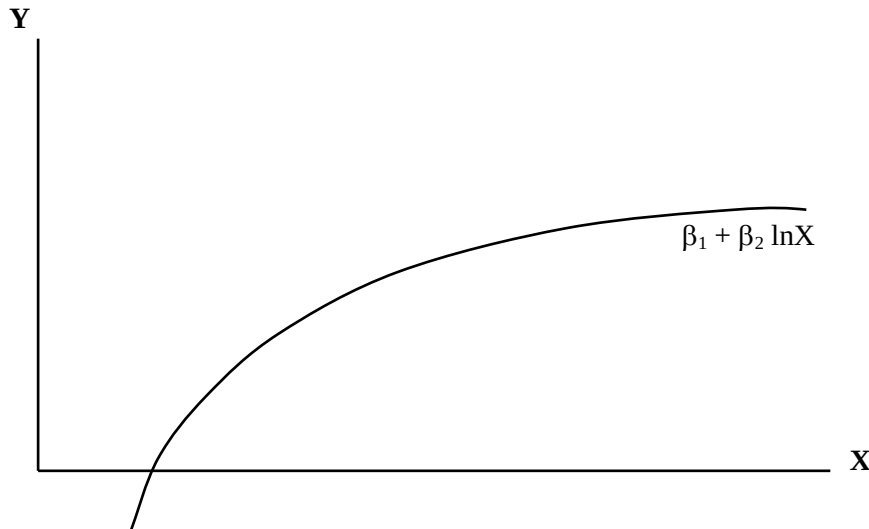
Trong một **mô hình logarit-tuyến tính**, biến phụ thuộc không đổi nhưng biến độc lập thể hiện dưới dạng logarit. Như vậy,

$$Y = \beta_1 + \beta_2 \ln X + u \quad (6.2)$$

Với số dương β_1 và β_2 , Hình 6.2 minh họa đồ thị quan hệ như là một hàm phi tuyến. Quan hệ này cho $\Delta Y/\Delta X = \beta_2/X$. Nếu $\beta_2 > 0$, sự tăng cận biên của Y tương ứng với sự tăng của X là một hàm giảm của X. Ta lưu ý rằng

$$\Delta Y = \beta_2 \frac{\Delta X}{X} = \frac{\beta_2}{100} \left[100 \frac{\Delta X}{X} \right] = \frac{\beta_2}{100} \times \text{thay đổi phần trăm của X}$$

Từ đây sẽ cho một điều là thay đổi một *phần trăm* giá trị biến X sẽ làm thay đổi Y, trung bình, $\beta_2/100$ đơn vị (*không phải phần trăm*).

● Hình 6.2 Dạng Hàm Logarit-Tuyến Tính

Ví dụ, gọi Y là sản lượng lúa mì và X là số mẫu trồng trọt. Vậy $\Delta Y/\Delta X$ là sản lượng cận biên của một mẫu trồng trọt thêm. Ta giả thuyết rằng sản lượng cận biên sẽ giảm khi diện tích tăng. Khi diện tích thấp, ta kỳ vọng rằng vùng đất màu mỡ nhất sẽ được trồng trọt trước tiên. Khi diện tích tăng, những vùng ít màu mỡ hơn sẽ được đem sử dụng; sản lượng có thêm từ những vùng này có thể không cao như sản lượng từ những vùng đất màu mỡ hơn. Điều này đưa ra giả thuyết sự giảm sản lượng cận biên của diện tích lúa mì. Lập công thức logarit-tuyến tính giúp chúng ta có thể hiểu thấu mối quan hệ này.

Ví dụ khác, Gọi Y là giá của một căn nhà và X là diện tích sinh hoạt. Xem xét 2 căn nhà, một căn với diện tích sinh hoạt là 1.300 bộ vuông (square feet) và một căn khác với diện tích sinh hoạt 3.200 bộ vuông. Ta kỳ vọng rằng phần giá tăng thêm mà một người tiêu dùng sẽ sẵn sàng trả cho 100 bộ vuông thêm vào diện tích sinh hoạt sẽ cao khi $X = 1.300$ hơn là khi $X = 3.200$. Điều này là bởi vì căn nhà sau đã rộng sẵn, và người mua có thể không muốn trả thêm nhiều để tăng thêm diện tích. Điều này có nghĩa rằng tác động cận biên của SQFT (diện tích) lên PRICE (giá) kỳ vọng sẽ giảm khi SQFT tăng. Một cách để kiểm định điều này là điều chỉnh một mô hình logarit-tuyến tính và kiểm định giả thuyết $H_0: \beta_2 = 0$ đối lại giả thuyết $H_1: \beta_2 > 0$. Điều này sẽ được nhìn nhận như là một kiểm định một phía. Quy tắc ra quyết định là bác bỏ H_0 nếu $t_c > t_{n-2}^* (0,05)$. Ta lưu ý từ Bảng 6.1 rằng trong mô hình này độ co giãn của Y đối với X là β_2/Y . Ta có thể tính toán độ co giãn tại giá trị trung bình là $\beta_2/\bar{Y}$. Nếu dữ liệu là chuỗi thời gian, độ co giãn đáng quan tâm hơn là độ co giãn tương ứng với quan sát gần đây nhất – với $t = n$. Độ co giãn này là β_2/Y_n .

Mặc dù những ví dụ minh họa này vẫn là các dạng mô hình hồi quy đơn giản, phần mở rộng thêm cho trường hợp đa biến là không phức tạp. Đơn giản là phát ra các logarit của các biến giải thích thích hợp, gọi chúng là Z_1, Z_2 v.v... và hồi quy biến Y theo một hằng số và các biến Z.

● BÀI TOÁN THỰC HÀNH 6.1

Tìm biểu thức độ co giãn của Y đối với X trong các mô hình tuyến tính và phi tuyến và chứng minh các mục trong Bảng 6.1.

● BÀI TOÁN THỰC HÀNH 6.2

Vẽ đồ thị Phương trình (6.2) khi $\beta_2 < 0$ (để đơn giản giả sử rằng $\beta_1 = 0$).

● VÍ DỤ 6.1

Ta đã ước lượng mô hình logarit-tuyến tính sử dụng dữ liệu giá nhà trong Bảng 4.1 (xem Phần Máy Tính Thực Hành 6.1 giới thiệu cách chạy lại các kết quả của ví dụ này và kiểm tra những khẳng định đã thực hiện ở đây). Sự biện luận về sự giảm tác động cận biên áp dụng như nhau cho số phòng ngủ và số phòng tắm. Vì vậy ta đã phát ra các logarit của các biến SQFT, BEDRMS, và BATHS và kế tiếp đã hồi quy biến PRICE theo một hằng số và những số hạng logarit này. Kế đến logarit của BATHS và BEDRMS được loại bỏ mỗi lần từng biến một bởi vì hệ số của chúng rất không có ý nghĩa. Mô hình “tốt nhất” đã được chọn theo các tiêu chuẩn lựa chọn đã thảo luận trong Chương 4. Các phương trình ước lượng của mô hình tuyến tính tốt nhất và mô hình logarit-tuyến tính tốt nhất sẽ được trình bày tiếp sau, với các trị thống kê t trong ngoặc.

$$\widehat{\text{PRICE}} = 52,351 + 0,139 \text{ SQFT}$$

(1,4) (7,4)

$$\bar{R}^2 = 0,806 \quad \text{d.f.} = 12$$

$$\widehat{\text{PRICE}} = -1,749,974 + 299,972 \ln(\text{SQFT}) - 145,094 \ln(\text{BEDRMS})$$

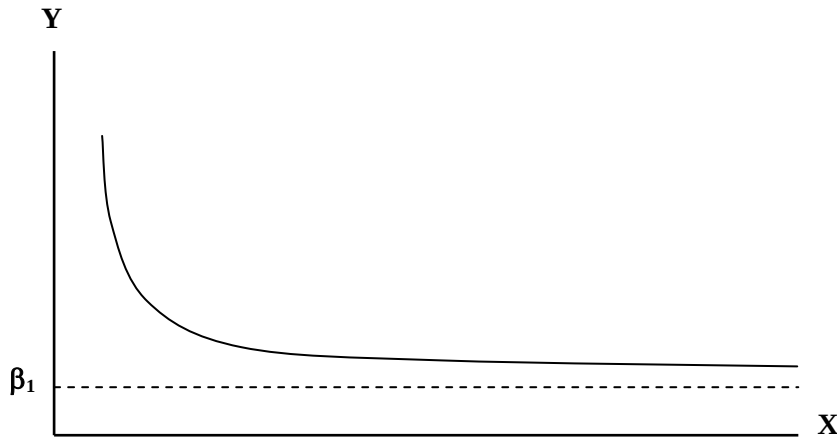
(-6,8) (7,5) (-1,7)

$$\bar{R}^2 = 0,826 \quad \text{d.f.} = 11$$

Ta lưu ý rằng giá trị $\bar{R}^2$ hơi cao hơn đối với mô hình logarit-tuyến tính. Mô hình này cũng có các trị thống kê lựa chọn mô hình thấp nhất. Tuy nhiên, hệ số cho logarit của BEDRMS chỉ có ý nghĩa ở mức 11,48 phần trăm. Nếu số hạng này bị loại bỏ, các trị thống kê lựa chọn sẽ xấu đi đáng kể, và do đó ta đã chọn giữ nó lại. Hệ số hồi quy cho $\ln(\text{SQFT})$ có ý nghĩa cao, vậy ủng hộ cho giả thuyết rằng tác động cận biên của diện tích sinh hoạt giảm khi số bộ vuông tăng. Hệ số cho logarit của BEDRMS có giá trị âm giống như đối với mô hình tuyến tính, nhưng tác động của hệ số này là yếu về mặt thống kê.

● BÀI TOÁN THỰC HÀNH 6.3

Tính độ co giãn từng phần của PRICE đối với SQFT cho các mô hình ước lượng logarit-tuyến tính và tuyến tính khi SQFT là 1.500, 2.000 và 2.500. Làm thế nào chúng so sánh với nhau?

● Hình 6.3 Quan Hệ Nghịch Đảo**● 6.3 Biến Đổi Nghịch Đảo**

Một dạng hàm thường được sử dụng để ước lượng đường cong nhu cầu là hàm biến đổi nghịch đảo:

$$Y = \beta_1 + \beta_2 \left(\frac{1}{X} \right) + u$$

Bởi vì đường cong nhu cầu đặc thù dốc xuống, ta kỳ vọng β_2 là dương. Lưu ý rằng khi X trở nên lớn, Y tiệm cận gần với β_1 (xem Hình 6.3). Dấu và độ lớn của β_1 sẽ xác định đường cong có cắt trục X hay không.

● BÀI TOÁN THỰC HÀNH 6.4

Vẽ đồ thị hàm nghịch đảo với $\beta_2 < 0$, $\beta_1 > 0$.

● 6.4 Thích Hợp Đường Cong Đa Thức

Các nhà nghiên cứu rất thường dùng một đa thức để liên hệ một biến phụ thuộc với một biến độc lập. Mô hình này có thể là

$$Y = \beta_1 + \beta_2 X + \beta_3 X^2 + \beta_4 X^3 + \dots + \beta_{k+1} X^k + u$$

Thủ tục ước lượng bao gồm tạo các biến mới X^2 , X^3 , v.v... qua các phép biến đổi và kể đến hồi quy Y theo một số hạng hằng số, theo X , và theo các biến đã biến đổi này. Mức đa thức (k) bị ràng buộc bởi số quan sát. Nếu $k = 3$, ta có quan hệ bậc ba; và nếu $k = 2$, ta có công thức bậc hai. Các công thức bậc hai thường được sử dụng để điều chỉnh các hàm chi phí có dạng chữ U và các quan hệ phi tuyến khác. Một đường cong bậc ba thường được làm thích hợp gần đúng với hình dạng trong Hình 6.9 (xem phần mô hình logit). Nhìn chung, bậc đa thức lớn hơn 2 nên tránh. Một trong các lý do là thực tế mỗi số hạng đa thức đồng nghĩa với việc mất đi thêm một bậc tự do. Như đã đề cập trong Chương 3, sự mất đi bậc tự do nghĩa là giảm sự chính xác của các ước lượng các thông số và giảm khả năng của các kiểm định. Cũng vậy, ta đã thấy trong Chương 5 rằng mối tương quan cao có thể có giữa X , X^2 , và X^3 làm cho các hệ số riêng lẻ kém tin cậy hơn.

Sử dụng các tính chất về đạo hàm (xem Tính chất 2.A.5), ta có thể cho thấy rằng tác động cận biên của X lên Y được xác định bởi

$$dY/dX = \beta_2 + 2\beta_3X + 3\beta_4X^2 + \dots + k\beta_{k+1}X^{k-1}$$

Một trường hợp đặc biệt của dạng hàm đa thức là mô hình bậc hai

$$Y = \beta_1 + \beta_2X + \beta_3X^2 + u$$

Tác động cận biên của X lên Y, nghĩa là độ dốc của quan hệ bậc hai, được xác định bởi $dY/dX = \beta_2 + 2\beta_3X$. Lưu ý rằng tác động cận biên của X lên Y phụ thuộc vào giá trị của X mà tại đó ta tính tác động cận biên. Một giá trị phổ biến được dùng là giá trị trung bình, $\bar{X}$. Như đã cho thấy trong phụ lục Chương 2, khi $dY/dX = 0$, hàm số sẽ hoặc đạt cực đại hoặc cực tiểu. Giá trị X tại đó xảy ra điều này sẽ có được từ việc giải điều kiện $\beta_2 + 2\beta_3X = 0$ khi $X_0 = -\beta_2/(2\beta_3)$. Để xác định xem hàm đạt cực tiểu hay cực đại, ta cần phải tính đạo hàm bậc hai, $d^2Y/dX^2 = 2\beta_3$. Nếu $\beta_3 < 0$, hàm số sẽ đạt cực đại tại X_0 , và nếu β_3 dương, hàm đạt cực tiểu tại X_0 . Tiếp theo ta trình bày hai ví dụ: một hàm chi phí trung bình có quan hệ dạng chữ U (Hình 6.4) và một hàm sản xuất có quan hệ dạng đường cong lõm (hump-shaped) (Hình 6.5).

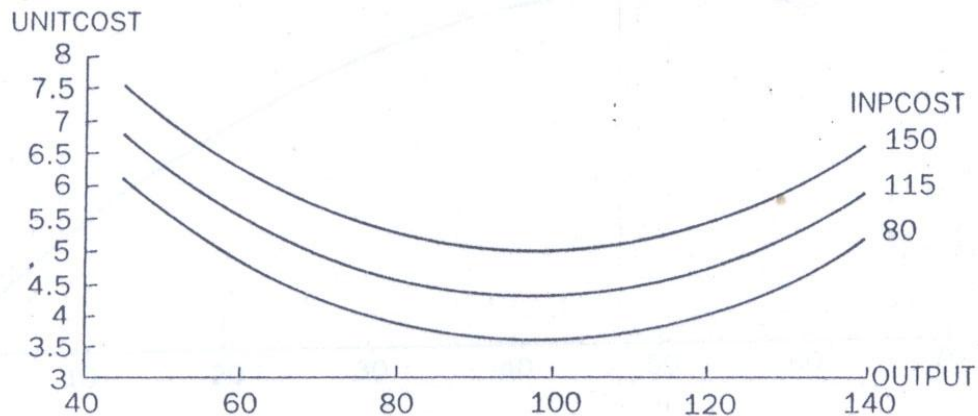
● VÍ DỤ 6.2

DATA6-1 đã mô tả trong Phụ lục D có dữ liệu về chi phí đơn vị (UNITCOST) của một công ty sản xuất trên một thời đoạn 20 năm, một chỉ số xuất lượng của công ty (OUTPUT), và một chỉ số chi phí nhập lượng của công ty (INPCOST). Trước hết ta có bình phương hai biến độc lập và kế đến hồi quy UNITCOST theo một hằng số, OUTPUT, OUTPUT², INPCOST, và INPCOST² (xem Phần Máy Tính Thực Hành 6.2 để biết thêm chi tiết về điều này). Bởi vì INPCOST² có hệ số vô cùng không có ý nghĩa, nó bị loại bỏ và mô hình được ước lượng lại. Các kết quả được cho sau đây, với các trị thống kê t trong ngoặc.

$$\begin{aligned} \widehat{\text{UNITCOST}} &= 10,522 - 0,175 \text{ OUTPUT} + 0,000895 \text{ OUTPUT}^2 \\ &\quad \begin{matrix} (14,3) & (-9,7) & (7,8) \end{matrix} \\ &\quad + 0,0202 \text{ INPCOST} \\ &\quad \begin{matrix} (14,454) \end{matrix} \\ \bar{R}^2 &= 0,978 \quad \text{d.f.} = 16 \end{aligned}$$

Lưu ý rằng đối với mô hình này $\hat{\beta}_1, \hat{\beta}_3 > 0$ và $\hat{\beta}_2 < 0$, giải thích cho quan hệ dạng chữ U. Mô hình giải thích 97,8 phần trăm sự thay đổi trong chi phí trung bình. Dễ dàng chứng minh rằng tất cả các hệ số hồi quy đều vô cùng có ý nghĩa. Lưu ý rằng những gì ta có trên đây là một họ các đường cong chi phí trung bình được di chuyển theo các mức chỉ số chi phí nhập lượng. Cũng rất hữu ích khi vẽ đồ thị hàm chi phí đơn vị cho một chỉ phí nhập lượng tiêu biểu. Hình 6.4 là hàm chi phí trung bình có dạng chữ U ước lượng cho một dây xuất lượng và 3 mức chi phí nhập lượng khác nhau (80, 115, và 150). Chúng đạt giá trị nhỏ nhất tại chỉ số xuất lượng có mức 98 (hãy xác minh).

● Hình 6.4 Các Hàm Chi Phí Trung Bình Ước Lượng



● VÍ DỤ 6.3

DATA6-2 đã mô tả trong Phụ lục D có dữ liệu hàng năm về việc sản xuất cá ngừ trắng (*Thunnus Alalunga*) trong vùng Basque của Tây Ban Nha. Biến xuất lượng (phụ thuộc) là tổng số mẻ cá theo đơn vị ngàn tấn và biến nhập lượng (độc lập) là nỗ lực đánh cá được đo lường bằng tổng số ngày đánh cá (đơn vị là ngàn). Mô hình ước lượng là (trị thống kê t trong ngoặc)

$$\widehat{\text{Catch}} = 1,642 \text{ Effort} - 0,01653 \text{ Effort}^2$$

(17,1) (-8,0)

$$\bar{R}^2 = 0,660 \quad \text{d.f.} = 32$$

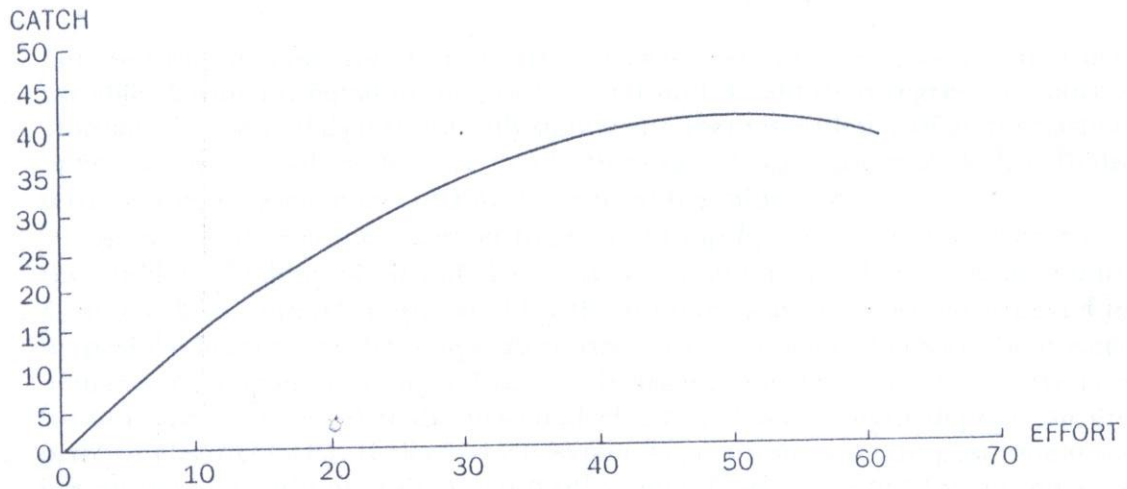
Phần Máy Tính Thực Hành 6.3 có thể được dùng để xác minh điều này. Lưu ý rằng, bởi vì mẻ cá không thể có được khi không có nỗ lực, β_1 về lý thuyết phải bằng 0 cho mô hình này. Ta hẳn thấy rằng $\hat{\beta}_2 > 0$ và $\hat{\beta}_3 < 0$; do đó, hàm sản xuất sẽ có đồ thị như Hình 6.5 với giá trị cực đại đạt được khi nỗ lực là 50.

● BÀI TOÁN THỰC HÀNH 6.5⁺

Sử dụng dữ liệu giá nhà, hãy ước lượng quan hệ bậc hai sau giữa giá và bộ vuông:

$$\text{PRICE} = \beta_1 + \beta_2 \text{SQFT} + \beta_3 \text{SQFT}^2 + u$$

● Hình 6.5 Hàm Sản Xuất Ước Lượng



Diễn giải về mặt kinh tế của giả thuyết $\beta_3 = 0$ là gì? Kiểm định giả thuyết này đối lại với giả thuyết $H_1: \beta_3 < 0$. Bạn có kết luận gì về tác động cận biên của SQFT lên PRICE? So sánh mô hình này, theo các tiêu chuẩn lựa chọn, với mô hình logarit-tuyến tính được ước lượng trong Ví dụ 6.1 (xem Phần Máy Tính Thực Hành 6.4).

● BÀI TOÁN THỰC HÀNH 6.6

Hãy ước lượng mô hình $PRICE = \beta_1 + \beta_2 \ln SQFT + \beta_3 BATHS + u$, và so sánh các kết quả với các kết quả trong Bảng 4.2 và trong Bài Toán Thực Hành 6.5.

● BÀI TOÁN THỰC HÀNH 6.7

Với quan hệ $Y = \beta_1 + \beta_2 X + \beta_3 X^2$, hãy xác minh độ dốc và độ co giãn cho trong Bảng 6.1.

● 6.5 Các Số Hạng Tương Tác

Tác động cận biên của một biến giải thích đôi khi có thể phụ thuộc vào một biến khác. Để minh họa, Klein và Morgan (1951) đã đề xuất một giả thuyết về sự tương tác của thu nhập và tài sản trong việc xác định các dạng tiêu dùng. Họ biện luận cho rằng xu hướng tiêu dùng biên tế cũng sẽ phụ thuộc vào tài sản – một người giàu hơn có thể có xu hướng biên tế khác để tiêu dùng ngoài khoản thu nhập. Để thấy điều này, gọi $C = \alpha + \beta Y + u$. Giả thuyết là β , xu hướng tiêu dùng biên tế, phụ thuộc vào tài sản (A). Một cách đơn giản cho phép thực hiện là giả sử rằng $\beta = \beta_1 + \beta_2 A$. Thay thế biểu thức này vào hàm tiêu dùng, ta thu được $C = \alpha + (\beta_1 + \beta_2 A)Y + u$. Điều này biến đổi thành mô hình $C = \alpha + \beta_1 Y + \beta_2 (AY) + u$. Số hạng AY được xem là **số hạng tương tác** bởi vì nó bao gộp sự tương tác giữa các tác động của thu nhập và tài sản. Nhằm mục đích ước lượng, ta tạo ra một biến mới Z , bằng với tích của Y và A , và kể đến hồi quy C theo một hằng số, Y , và Z . Nếu β_2 có ý nghĩa về mặt thống kê, thì có dấu hiệu về sự tương tác giữa thu nhập và tài sản. Lưu ý rằng trong ví dụ này, $\Delta C / \Delta Y = \beta_1 + \beta_2 A$. Để xác định tác động cận biên của Y lên C , ta cần có giá trị của A .

Ví dụ thứ hai, xét quan hệ $E_t = \alpha + \beta T_t + u_t$, trong đó E_t là số kilowatt giờ tiêu thụ điện và T_t là nhiệt độ tại thời điểm t . Nếu mô hình này được ước lượng cho mùa hè, ta kỳ vọng β sẽ

đương bởi vì, khi nhiệt độ tăng vào mùa hè, thì nhu cầu dùng máy lạnh sẽ cao hơn và do đó tiêu thụ điện sẽ tăng. Tuy nhiên, ta có thể giả thuyết rằng tác động cận biên của T lên E có thể phụ thuộc vào giá điện (P_t). Nếu giá điện là đắt, người tiêu dùng có thể hoãn bật máy lạnh hoặc tắt sớm hơn. Một cách để kiểm định tác động này là giả sử rằng $\beta = \beta_1 + \beta_2 P_t$. Vậy ta đang giả sử rằng tác động cận biên của nhiệt độ lên tiêu thụ điện phụ thuộc vào giá. Thay biểu thức này vào quan hệ, ta có

$$E_t = \alpha + (\beta_1 + \beta_2 P_t)T_t + u_t = \alpha + \beta_1 T_t + \beta_2 (P_t T_t) + u_t$$

Để ước lượng các thông số, ta cho $Z_t = P_t T_t$ và hồi quy E theo một hằng số, T , và Z . Sự ý nghĩa của β_2 là dấu hiệu của một tác động tương hỗ giữa nhiệt độ và giá. Lưu ý rằng $\Delta E / \Delta P = \beta_2 T$; nghĩa là, tác động cận biên của P lên E phụ thuộc vào nhiệt độ. Nếu ta cho α cũng phụ thuộc vào P , mô hình trở thành

$$E_t = \alpha_1 + \alpha_2 P_t + \beta_1 T_t + \beta_2 (P_t T_t) + u_t$$

Trong các chương sau, ta có vài ví dụ về các tác động tương hỗ như vậy.

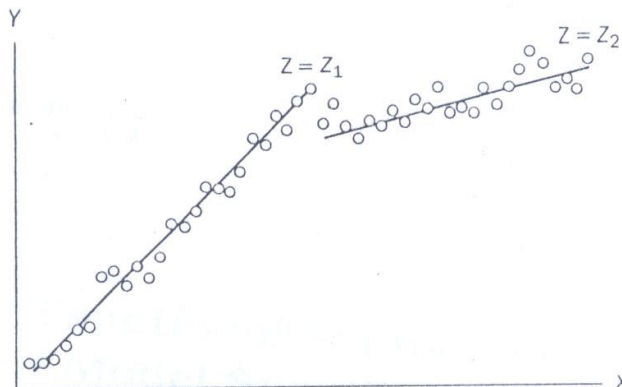
Phi Tuyến Giả Tạo

Để nhận biết sự phi tuyến có thể có, ta có thể thử vẽ đồ thị Y theo một biến độc lập cụ thể (X) và quan sát xem có sự phi tuyến nào xảy ra hay không. Đây là thủ tục nguy hiểm bởi vì nó có thể dẫn đến đặc trưng sai mô hình nghiêm trọng. Ví dụ, giả sử rằng Y là tuyến tính với X , Z , và số hạng tương tác XZ , vậy ta có

$$Y = \beta_1 + \beta_2 X + \beta_3 Z + \beta_4 (XZ) + u \quad \text{và} \quad \Delta Y / \Delta X = \beta_2 + \beta_4 Z$$

Trong tính toán tác động cận biên của X lên Y , ta xem Z là cố định. Lưu ý rằng tác động cận biên của X lên Y , nghĩa là độ dốc, phụ thuộc vào Z . Biểu đồ phân tán quan sát thực nghiệm, giữa Y và X có thể nhìn giống như Hình 6.6, có vẻ như là quan hệ logarit-tuyến tính giữa Y và X . Trong thực tế, điều này là do hai quan hệ tuyến tính giữa Y và X với các giá trị khác nhau của Z (Z_1 và Z_2). Vậy, thay vì vẽ đồ thị thực nghiệm quan sát biến Y theo mỗi biến X , bạn nên cố gắng mô hình hoá quá trình phát dữ liệu (DGP) dùng lý thuyết và trực giác về hành vi cơ bản và kẻ đến tiến hành kiểm định đặc trưng. Trong Phần 6.13, 6.14, và 6.15, ta thảo luận vài phương pháp để kiểm định các đặc trưng hồi quy.

● Hình 6.6 Một Ví Dụ của Phi Tuyến Giả Tạo



● 6.6 Hiện Tượng Trễ Trong Hành Vi (Các Mô Hình Động)

Các tác động kinh tế và các biến khác hiếm khi xảy ra tức thời; phải tốn thời gian để người tiêu dùng, nhà sản xuất, và các tác nhân kinh tế khác phản ứng. Lý thuyết kinh tế vĩ mô cho ta biết rằng tổng sản lượng quốc dân (GNP) cân bằng (Y) được xác định bởi một số biến ngoại sinh, đặc biệt, bởi chi tiêu chính phủ (G), thuế (T), cung tiền (M), xuất khẩu (X) v.v.... Bởi vì hiệu ứng cân bằng chỉ giảm được sau một khoảng thời gian, các mô hình kinh tế lượng dùng dữ liệu dạng chuỗi thời gian thường được thành lập với **hiện tượng trễ trong hành vi**. Một ví dụ của mô hình như vậy cho như sau:

$$Y_t = \beta_1 + \beta_2 G_t + \beta_3 G_{t-1} + \beta_4 M_t + \beta_5 M_{t-1} + \beta_6 T_t + \beta_7 T_{t-1} + \beta_8 X_t + \beta_9 X_{t-1} + u_t$$

Thủ tục ước lượng ở đây hoàn toàn đơn giản. Đơn giản ta tạo các biến có hiệu ứng trễ G_{t-1} , M_{t-1} , T_{t-1} và X_{t-1} và hồi quy Y_t theo các biến này *dùng quan sát từ 2 đến n*. Bởi vì G_{t-1} và các biến khác không được định nghĩa cho $t = 1$, ta mất quan sát thứ nhất trong ước lượng. Tuy nhiên, một số vấn đề phát sinh trong mô hình này bởi vì các biến độc lập tương quan với nhau và cũng do bởi vì bậc tự do bị mất khi có nhiều hiệu ứng trễ hơn thêm vào. Những vấn đề này được thảo luận chi tiết trong Chương 10.

Hiện tượng trễ trong hành vi có thể có dạng hiện tượng trễ trong biến phụ thuộc. Mô hình có thể có dạng

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 X_t + \beta_4 X_{t-1} + u_t$$

Ví dụ, gọi Y_t là chi tiêu tại thời điểm t và X_t là thu nhập. Bởi vì người tiêu dùng có xu hướng duy trì mức tiêu chuẩn sống thường lệ, ta có thể kỳ vọng sự tiêu dùng của họ liên quan mật thiết với sự tiêu dùng trước đây của họ. Vì vậy, chúng ta có thể kỳ vọng là Y_t cũng phụ thuộc vào Y_{t-1} . Cụ thể hơn, xem phương trình sau:

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 (X_t - X_{t-1}) + u_t$$

Vì “các tập quán thói quen” nên nói chung người tiêu dùng miễn cưỡng thay đổi lối sống của họ, và do đó chúng ta kỳ vọng mức tiêu thụ tại thời điểm t (Y_t) phụ thuộc vào mức tiêu thụ ở giai đoạn trước đó (Y_{t-1}). Tuy nhiên, nếu mức thu nhập (X_t) thay đổi, người tiêu dùng sẽ điều chỉnh hành vi tiêu dùng của họ tương ứng với sự tăng hoặc giảm thu nhập. Do vậy chúng ta sẽ dùng mô hình động được xây dựng ở trên và kỳ vọng rằng tất cả các hệ số sẽ có giá trị dương.

● VÍ DỤ 6.4

Tập dữ liệu DATA6-3 (xem Phụ lục D) là dữ liệu về chi tiêu tiêu dùng cá nhân đầu người của Vương Quốc Anh (C , đo bằng bảng Anh) và thu nhập tùy dụng đầu người (nghĩa là, thu nhập cá nhân trừ thuế, ký hiệu là DI , và cũng được tính theo đơn vị bảng Anh). Để điều chỉnh tác động của lạm phát, cả hai biến này được biểu diễn theo giá trị thực (còn được gọi là *giá không đổi*). Mô hình động ước lượng được trình bày dưới đây (xem Phần Thực Hành Máy Tính 6.5), với trị thống kê t trong ngoặc đơn.

$$\hat{C}_t = -46,802 + 1,022C_{t-1} + 0,706 (DI_t - DI_{t-1})$$

(-2.07) (123.0) (9.93)

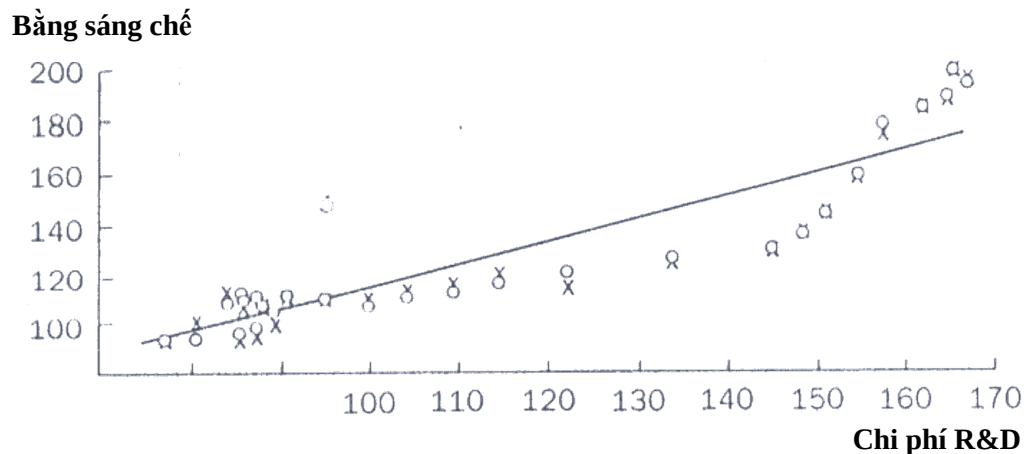
$$\bar{R}^2 = 0,998 \quad df = 38$$

Mặc dù mô hình đạt được sự thích hợp rất tốt và các ước lượng có vẻ hợp lý, mô hình này có một số trở ngại. Như sẽ thấy ở Chương 10 và 13 rằng mô hình này vi phạm tính độc lập chuỗi của Giả thiết 3.6 và Giả thiết 3.4 là các biến độc lập không được tương quan với các số hạng sai số. Đặc trưng sai này sẽ làm cho các trị ước lượng bị thiên lệch. Chúng ta sẽ xem xét lại mô hình này trong các chương 10 và 13.

● 6.7 Ứng dụng: Quan Hệ Giữa Số Bằng Sáng Chế Và Chi Tiêu R&D (đã duyệt lại)

Trong Phần 3.11, chúng ta đã ước lượng mô hình hồi quy tuyến tính đơn giữa số bằng sáng chế và chi tiêu cho R&D và biết rằng mô hình này là hoàn toàn không đủ vì biểu đồ phân tán của các giá trị quan sát cho thấy một quan hệ đường cong (Xem Hình 3.11). Chúng ta cũng chỉ ra rằng có hiện tượng trễ giữa chi tiêu thực cho hoạt động nghiên cứu và phát triển và hiệu quả của các chi tiêu này về mặt số bằng sáng chế. Ở đây chúng ta sẽ ước lượng **mô hình phi tuyến động** và so sánh các kết quả. Tuy nhiên, vì chưa có lý thuyết về kinh tế hay các lý thuyết khác về số năm của hiện tượng trễ này hoặc về dạng hàm số cần sử dụng, nên một cách tùy ý chúng ta cho độ trễ này lên đến 4 năm. Bốn biến trễ được tạo ra gồm R&D(t-1), R&D(t-2), R&D(t-3), và R&D(t-4). Các biến này sau đó sẽ được bình phương lên và một mô hình bậc hai với tất cả các biến được ước lượng.

● Hình 6.7 So Sánh Mô Hình Động và Mô Hình Tĩnh (đường liền là mô hình tĩnh, x là giá trị quan sát thực, và o là mô hình động)



Vì vậy, đây là một bài tập “khớp đường cong” thuần túy thay vì là một bài tập dựa trên lý thuyết kinh tế. Báo cáo có chú giải in ra từ máy tính ở bảng 6.2 cần được tìm hiểu kỹ lưỡng (xem Phần Thực Hành Máy Tính 6.6 để chạy lại bảng 6.2). Hình 6.7 vẽ số bằng sáng chế thật, các giá trị gán từ mô hình tĩnh ở Chương 3 (đường thẳng liền), và các giá trị từ mô hình động cuối cùng. Chúng ta nhận thấy rằng mô hình động thể hiện rất tốt diễn biến thực tế, ngay cả trong những năm các chi phí R&D tụt lại và trong những năm từ 1988-1993 khi mô hình tuyến tính hoàn toàn không thể hiện được. Do đó mô hình phi tuyến động là một đặc trưng tốt hơn so với **mô hình tĩnh** tuyến tính đơn giản.

● Bảng 6.2 Kết Quả Máy Tính Có Kèm Chú Giải Cho Phần Ứng Dụng ở Phần 6.7

MODEL 1: OLS estimates using the 34 observations 1960-1993
Dependent variable: PATENTS

	VARIABLE	COEFFICIENT	STDERROR	T STAT	2Prob(t> T)
0)	const	34.5711	6.3579	5.438	0.000006 ***
3)	R&D	0.7919	0.0567	13.966	0.000000 ***
Mean of dep. var.					
		119.238	S.D. of dep. variable		29.306
Error Sum of Sq (ESS)		3994.3003	Std Err of Resid. (sgmahat)		11.1724
Unadjusted R-squared		0.859	Adjusted R-squared		0.855
F-statistic (1, 32)		195.055	p-value for F()		0.000000
Durbin-Watson stat.		0.234	First-order autocorr. coeff		0.945

MODEL SELECTION STATISTICS

SGMASQ	124.822	AIC	132.146	FPE	132.164
HQ	136.255	SCHWARZ	144.56	SHIBATA	131.301
GCV	132.623	RICE	133.143		

● Bảng 6.2 (tiếp theo)

[phát các biến trễ]

R&D1 = R&D(-1)

R&D2 = R&D(-2)

R&D3 = R&D(-3)

R&D4 = R&D(-4)

$$sq_R\&D = (R\&D)^2$$

$$sq_R\&Di = (R\&Di)^2$$

for i = 1,2,3, and 4

[Ước lượng mô hình tổng quát với tất cả các biến giải thích bằng cách sử dụng chỉ các quan sát từ 1964-1993, vì các biến trễ không được định nghĩa trong giai đoạn từ 1960-1963]

MODEL 2: OLS estimates using 30 observations 1964-1993
Dependent variable: PATENTS

	VARIABLE	COEFFICIENT	STDERROR	T STAT	2Prob(t> T)
0)	const	85.3526	22.1027	3.862	0.001051 ***
3)	R&D	-0.0477	1.1251	-0.042	0.966638
4)	R&D1	0.6033	2.0562	0.293	0.772387
5)	R&D2	0.0001794	2.1850	0.000	0.999935
6)	R&D3	-0.5869	2.0522	-0.286	0.777989
7)	R&D4	-0.1837	1.0994	-0.167	0.869055
8)	sq_R&D	-0.0007326	0.0049	-0.150	0.882674
9)	sq_R&D1	-0.0018	0.0089	-0.197	0.845884
10)	sq_R&D2	0.0017	0.0098	0.177	0.861555
11)	sq_R&D3	-0.0007564	0.0092	-0.082	0.935597
12)	sq_R&D4	0.0071	0.0051	1.405	0.176209

Mean of dep. var.		123.330	S.D. of dep. variable		28.795
Error Sum of Sq (ESS)		223.3789	Std Err of Resid. (sgmahat)		3.4288
Unadjusted R-squared		0.991	Adjusted R-squared		0.986
F-statistic (1, 32)		202.626	p-value for F()		0.000000
Durbin-Watson stat.		1.797	First-order autocorr. coeff		0.101

MODEL SELECTION STATISTICS

SGMASQ	11.7568	AIC	15.5026	FPE	16.0676
--------	---------	-----	---------	-----	---------

HQ	18.2719	SCHWARZ	25.9139	SHIBATA	12.9063
GCV	18.5633	RICE	27.9224		

Excluding the constant, p-value was highest for variable 5 (R&D2)

[Lưu ý rằng có hiện tượng đa cộng tuyến rất cao giữa các biến giải thích. Các giá trị hiện hành và trễ của chi phí R&D cũng như R&D và các bình phương của chúng được kỳ vọng là tương quan chặt với nhau. Như vậy, không có gì ngạc nhiên, trừ số hạng hằng số, tất cả đều không có ý nghĩa. Như đã đề cập ở chương trước, điều này không có nghĩa rằng các biến này là “không quan trọng”, mà chỉ có nghĩa rằng hiện tượng đa cộng tuyến có thể là những biến ẩn cần được đưa vào mô hình. Theo phương pháp đơn giản hóa mô hình dựa trên dữ liệu, chúng ta nên loại các biến thừa. Bước đầu tiên, chúng ta loại bỏ các biến với giá trị p-values trên 0,9. Đó là các biến R&D, R&D2, và sq_R&D3.]

MODEL 3: OLS estimates using 30 observations 1964-1993

● Bảng 6.2 (tiếp theo)

Dependent variable: PATENTS

	VARIABLE	COEFFICIENT	STDERROR	T STAT	2Prob(> T)	
0)	const	84.8409	19.0579	4.452	0.000200	***
4)	R&D1	0.6043	0.6351	0.952	0.351669	
6)	R&D3	-0.7352	0.5233	-1.405	0.174012	
7)	R&D4	-0.0745	0.5134	-0.145	0.886004	
8)	sq_R&D	-0.0009491	0.0012	-0.824	0.418554	
9)	sq_R&D1	-0.0017	0.0034	-0.496	0.624855	
10)	sq_R&D2	0.0016	0.0025	0.641	0.527835	
12)	sq_R&D4	0.0066	0.0020	3.364	0.002799	***
Mean of dep. var.		123.330	S.D. of dep. variable		28.795	
Error Sum of Sq (ESS)		223.6243	Std Err of Resid. (sgmahat)		3.1882	
Unadjusted R-squared		0.991	Adjusted R-squared		0.988	
F-statistic (1, 32)		334.799	p-value for F()		0.000000	

MODEL SELECTION STATISTICS

SGMASQ	10.1647	AIC	12.7064	FPE	12.8753
HQ	14.3197	SCHWARZ	18.4628	SHIBATA	11.4297
GCV	13.861	RICE	15.9732		

Excluding the constant, p-value was highest for variable 7 (R&D4).

Comparison of Model 2 and Model 3 is given below: Null hypothesis is: the regression parameters are zero for the variables R&D, R&D2, and sq_R&D3.

Test statistic: $F(3,19) = 0.006957$, with p-value = 0.999173

Of the 8 model selection statistics, 8 have improved

[Trong kiểm định F Wald cho các biến bị loại ra, p-value đạt giá trị cao cho thấy rằng chúng ta không thể bác bỏ giả thuyết không cho rằng các hệ số của các biến này tất cả đều bằng không ngay cả tại mức ý nghĩa cao đến 0,9. Như vậy, loại bỏ chúng là hợp lý. Hơn nữa, tất cả tám trị thống kê chọn mô hình đều giảm, điều đó có nghĩa có một sự cải thiện về độ thích hợp của mô hình. Mặc dù nhiều giá trị p-value giảm, chỉ có duy nhất một giá trị đủ nhỏ để có ý nghĩa – đó là giá trị của biến số 12. Điều này có nghĩa phải loại bỏ thêm. Tiếp theo, chúng ta loại bỏ biến R&D4, sq_R&D1, và sq_R&D2, các biến này ứng với giá trị p-value lớn hơn 0,5]

MODEL 4: OLS estimates using 30 observations 1964-1993

Dependent variable: PATENTS

	VARIABLE	COEFFICIENT	STDERROR	T STAT	2Prob(> T)	
0)	const	82.8545	12.0355	6.884	0.000000	***
4)	R&D1	0.4771	0.3278	1.455	0.158001	
6)	R&D3	-0.6370	0.2388	-2.667	0.013227	**
8)	Sq_R&D	-0.0011	0.0010000	-1.146	0.262479	
12)	Sq_R&D4	0.0065	0.0006784	9.609	0.000000	***

● Bảng 6.2 (tiếp theo)

Mean of dep. var.	123.330	S.D. of dep. variable	28.795
Error Sum of Sq (ESS)	223.5118	Std Err of Resid. (sgmahat)	3.0562
Unadjusted R-squared	0.990	Adjusted R-squared	0.989
F-statistic (1, 32)	637.338	p-value for F()	0.000000
Durbin-Watson stat.	1.844	First-order autocorr. coeff	0.078

MODEL SELECTION STATISTICS

SGMASQ	9.34047	AIC	10.8631	FPE	10.8972
HQ	11.7057	SCHWARZ	13.7206	SHIBATA	10.3783
GCV	11.2086	RICE	11.6756		

Excluding the constant, p-value was highest for variable 8 (sq_R&D).
Comparison of Model 3 and Model 4:

Null hypothesis is: the regression parameters are zero for the variables R&D4, sq_R&D1, and sq_R&D2.

Test statistic: $F(3,22) = 0.324242$, with p-value = 0.807788

Of the 8 model selection statistics, 8 have improved.

[Trong trường hợp này cũng vậy, trong kiểm định F Wald cho các biến bị loại ra, p-value đạt giá trị cao cho thấy rằng chúng ta không thể bác bỏ giả thuyết không cho rằng các hệ số của các biến này tất cả đều bằng không ngay cả tại mức ý nghĩa cao đến 0,8. Vì vậy, việc loại bỏ chúng là hợp lý. Thêm nữa, tất cả tám trị thống kê chọn mô hình đều giảm, điều đó có nghĩa có một sự cải thiện về độ thích hợp của mô hình. Vẫn còn hai biến (sq_R&D và R&D1) có giá trị trên 15%. Chúng ta tiếp tục loại bỏ các biến này, nhưng từng biến một, và đi đến một mô hình cuối cùng trong đó tất cả các hệ số có ý nghĩa ở mức dưới 2%]

MODEL 5: OLS estimates using 30 observations 1964-1993

Dependent variable: PATENTS

	VARIABLE	COEFFICIENT	STDERROR	T STAT	2Prob(> T)	
0)	const	91.3464	6.4046	14.263	0.000000	***
6)	R&D3	-0.2951	0.1175	-2.512	0.018286	**
12)	sq_R&D4	0.0059	0.0005486	10.675	0.000000	***

Mean of dep. var.	123.330	S.D. of dep. variable	28.795
Error Sum of Sq (ESS)	258.6727	Std Err of Resid. (sgmahat)	3.0952
Unadjusted R-squared	0.989	Adjusted R-squared	0.988
F-statistic (1, 32)	1241.43	p-value for F()	0.000000
Durbin-Watson stat.	1.665	First-order autocorr. coeff	0.166

MODEL SELECTION STATISTICS

SGMASQ	9.58047	AIC	10.5315	FPE	10.5385
HQ	11.0143	SCHWARZ	12.1155	SHIBATA	10.3469
GCV	10.645	RICE	10.778		

Of the 8 model selection statistics, 7 have improved.

● Bảng 6.2 (tiếp theo)

[Tính các trị dự báo và sai số phần trăm tuyệt đối cho từng dự báo]

Obs	R&D	PATENT S	Predicted value	Prediction error	Absolute percent error
1964	76.83	93.2	93.1259	0.0740826	0.0794878
1965	80	100.4	93.8292	6.57081	6.54463
1966	84.82	93.5	94.8126	-1.31258	1.40383
1967	86.84	93	97.9126	-4.91264	5.28241
1968	88.81	98.7	102.306	-3.606	3.65394
1969	88.28	104.4	103.795	0.605085	0.579583
1970	85.29	109.4	107.851	1.5492	1.41609
1971	83.18	111.1	109.3	1.80002	1.62018
1972	85.07	105.3	111.483	-6.1826	5.87141
1973	86.72	109.6	111.815	-2.21525	2.02121
1974	85.45	107.4	109.399	-1.99891	1.86118
1975	83.41	108	106.76	1.24028	1.14841
1976	87.44	110	108.135	1.86509	1.69554
1977	90.11	109	110.169	-1.16945	1.07289
1978	94.5	109.3	109.491	-0.191014	0.174761
1979	99.28	108.9	106.285	2.61523	2.4015
1980	103.64	113	109.529	3.4713	3.07194
1981	108.77	114.5	111.009	3.49072	3.04867
1982	113.96	118.4	114.344	4.05551	3.42526
1983	121.72	112.4	118.482	-6.0819	5.41094
1984	133.33	120.6	122.149	-1.54888	1.28431
1985	144.78	127.1	126.998	0.101834	0.0801211
1986	148.39	133	131.477	1.52261	1.14482
1987	150.9	139.8	138.761	1.03908	0.743265
1988	154.36	151.9	152.722	-0.821732	0.540969
1989	157.19	166.3	170.303	-4.00303	2.40711
1990	161.86	176.7	175.76	0.9403	0.532145
1991	164.54	178.4	179.138	-0.737635	0.413472
1992	166.7	187.2	184.487	2.71267	1.44908
1993	165.2	189.4	188.272	1.12779	0.595455

[Trừ một số năm (1965, 1967, 1972 và 1983), tất cả các sai số phần trăm tuyệt đối đều nhỏ hơn 5 phần trăm. Thật ra, hầu hết các giá trị này đều nhỏ hơn 2 phần trăm. Cũng như vậy, so sánh với mô hình thống kê tuyến tính có R bình phương hiệu chỉnh bằng 0,855, mô hình cuối cùng này có giá trị tương ứng là 0,988.]

● 6.8 Quan hệ tuyến tính-logarit (hay là mô hình bán logarit)

Tất cả các quan hệ phi tuyến được thảo luận trước đây có biến phụ thuộc Y xuất hiện dưới dạng tuyến tính. Chỉ có những biến độc lập phải trải qua mọi sự biến đổi. Cũng sẽ lưu ý là, mặc dù chúng ta sử dụng log và bình phương của các biến độc lập, các mô hình đều *tuyến tính theo các hệ số*. Bây giờ, chúng ta khảo sát một vài mô hình trong đó biến độc lập xuất hiện ở dạng biến đổi.

Giả sử chúng ta có một biến P tăng với một tốc độ không đổi. Cụ thể hơn, đặt $P_t = (1 + g)P_{t-1}$, với g là tốc độ tăng trưởng không đổi giữa thời đoạn $t - 1$ và t . P có thể là dân số và g là tốc độ tăng dân số. Bằng cách thay thế lặp lại ta có $P_t = P_0(1+g)^t$. Sử dụng dữ liệu về P_t , chúng ta muốn ước lượng tốc độ tăng trưởng g . Mỗi quan hệ này không có dạng tuyến tính thuận lợi đã

được dùng trong các phần trước. Tuy nhiên, có thể chuyển quan hệ này thành dạng tuyến tính được. Lấy logarit của hai vế (và dùng Tính chất 6.1), chúng ta có $\ln P_t = \ln P_0 + t \ln(1 + g)$. Đặt $Y_t = \ln P_t$, $X_t = t$, $\beta_1 = \ln P_0$ và $\beta_2 = \ln(1 + g)$. Khi đó, mỗi quan hệ có thể được viết lại như sau $Y_t = \beta_1 + \beta_2 X_t$. Vì Y và X có lẽ không thỏa mãn một cách chính xác mỗi quan hệ, chúng ta cộng thêm một số hạng sai số u_t , làm cho mỗi quan hệ giống với mô hình hồi quy đơn giản của Phương trình (3.1). Mô hình biến đổi trở thành

$$\ln P_t = \beta_1 + \beta_2 t + u_t \quad (6.3)$$

Lấy hàm số mũ phương trình này, ta có mô hình gốc là

$$P_t = e^{\beta_1 + \beta_2 t + u_t} \quad (6.4)$$

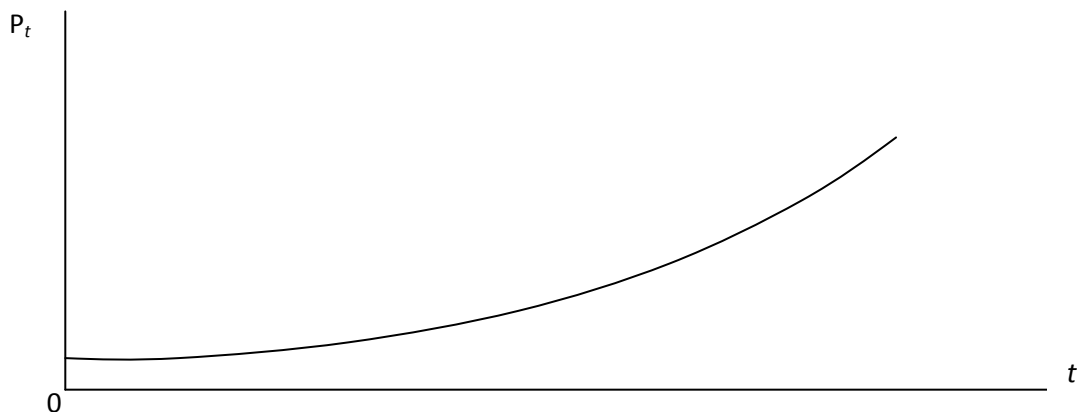
Phương trình (6.4) là một quan hệ hàm số mũ và được minh họa trong Hình 6.8. Cần lưu ý là số hạng nhiễu trong Phương trình (6.4) có thể tăng lên gấp nhiều lần. Phương trình (6.3) là tuyến tính khi biến phụ thuộc ở dạng logarit. Với $\ln P_t$ thuộc trục tung, công thức trở thành phương trình đường thẳng. Bước đầu tiên để ước lượng tốc độ tăng trưởng (g) là chuyển các quan sát $P_1, P_2, \dots, P_n$ bằng cách sử dụng phép biến đổi logarit vì vậy chúng ta có $Y_t = \ln P_t$. Kể đến chúng ta hồi quy Y_t theo một số hạng không đổi và thời gian t . Chúng ta có

$$\widehat{\ln P_0} = \hat{\beta}_1 \quad \text{và} \quad \widehat{\ln(1 + g)} = \hat{\beta}_2$$

Giải được g và P_0 , ta có

$$\hat{P}_0 = e^{\hat{\beta}_1} \quad \text{và} \quad \hat{g} = e^{\hat{\beta}_2} - 1 \quad (6.5)$$

● Hình 6.8 Hàm Dạng Hàm Số Mũ



Bất kỳ giả thuyết nào về g đều có thể thể hiện (có một số ngoại lệ không đáng kể) thành một giả thuyết tương đương theo β_2 . Do biến phụ thuộc được biến đổi ở dạng log, mô hình này được gọi là **mô hình tuyến tính-logarit**, hoặc đôi khi còn gọi là **mô hình bán logarit**. Nếu mô hình này được viết dưới dạng $\ln P_t = \beta_1 + \beta_2 X_t + u_t$, β_2 là tác động biên tế của X lên $\ln P_t$ không phải lên P_t . β_2 được gọi là **tốc độ tăng trưởng tức thời**. Lấy đạo hàm hai vế theo X_t (xem Tính chất 6.2 về đạo hàm), ta có

$$\beta_2 = \frac{d(\ln P_t)}{dX_t} = \frac{1}{P_t} \frac{dP_t}{dX_t} \quad (6.6)$$

Số hạng dP_t/P_t có thể được diễn dịch như là *thay đổi* của P_t chia cho P_t . Khi nhân với 100, β_2 cho *phần trăm thay đổi của P_t trên một đơn vị thay đổi của X_t* . Để tính độ co giãn của P theo X , xem Bảng 6.1.

Lấy giá trị kỳ vọng của hai vế phương trình (6.4), ta có

$$E(P_t) = e^{\beta_1 + \beta_2 t} E(e^{u_t}) \quad (6.7)$$

Có thể thấy là $E(e^{u_t}) = e^{\sigma^2/2} \neq 1$, và do đó nếu chúng ta dự báo P_t bằng cách dùng biểu thức $e^{\beta_1 + \beta_2 t}$, giá trị dự đoán sẽ thiên lệch, không nhất quán và không hiệu quả. Biểu thức phù hợp trong trường hợp này là

$$\hat{P}_t = \exp[\hat{\beta}_1 + \hat{\beta}_2 t + (\hat{\sigma}^2/2)] \quad (6.8)$$

với $\hat{\sigma}^2$ là phương sai mẫu của các số hạng sai số và $\exp$ là hàm số mũ. $\hat{P}_t$ là một ước lượng nhất quán của $E(P_t)$.

Cần có một điều chỉnh tương tự trong Phương trình (6.5) vì $E(e^{\hat{\beta}_2}) = e^{\beta_2 + [\text{Var}(\hat{\beta}_2)/2]}$. Do đó, một ước lượng không thiên lệch của g được tính bởi

$$\tilde{g} = \exp[\hat{\beta}_2 - 1/2 \widehat{\text{Var}}(\hat{\beta}_2)] - 1$$

Có thể có được một khoảng dự báo hiệu chỉnh của P_t . Trước đây, chúng ta đã định nghĩa $Y_t = \ln(P_t)$. Đặt $\hat{Y}_t$ là dự báo của $\ln(P_t)$ trong mô hình tuyến tính logarit và $s_t = s(\hat{Y}_t)$ là sai số chuẩn được ước lượng tương ứng. Vậy, khoảng tin cậy của Y_t là $\hat{Y}_t \pm t^* s_t$, với t^* là điểm trên phân phối t sao cho $P(t > t^*) = \text{một nửa của mức ý nghĩa}$ (tham khảo Phần 3.9 về các khoảng tin cậy của dự báo). Lấy hàm số mũ (nghĩa là ngược với lấy log) và hiệu chỉnh để thiên lệch giống như trong Phương trình (6.8), chúng ta có khoảng tin cậy hiệu chỉnh cho việc dự báo P_t là $\exp[\hat{Y}_t \pm t^* s_t + (\sigma^2/2)]$, với $\hat{\sigma}^2$ là phương sai mẫu của các số hạng sai số. Cần chỉ ra là khoảng tin cậy này sẽ không đối xứng qua $P_t = \exp[\hat{Y}_t + (\sigma^2/2)]$. Tham khảo Nelson (1973, trang 161-165) để thảo luận thêm về các dự báo điểm và các khoảng tin cậy của chúng khi biến phụ thuộc được biến đổi sang log.

● VÍ DỤ 6.5

Mô hình tuyến tính-logarit được sử dụng rộng rãi trong lý thuyết về vốn nhân lực trong đó lý thuyết cho rằng logarit của thu nhập hoặc lương được sử dụng như là một biến phụ thuộc. Để phát triển lý thuyết này, giả sử là tỷ suất lợi nhuận của một năm học tập thêm là r . Vậy, đối với thời đoạn thứ nhất, lương $w_1 = (1 + r)w_0$. Đối với hai năm học tập công thức này là $w_2 = (1 + r)^2 w_0$. Đối với s năm, chúng ta có $w_s = (1 + r)^s w_0$. Lấy logarit, chúng ta có (tham khảo Tính chất 6.1c).

$$\ln(w_s) = s \ln(1 + r) + \ln(w_0) = \beta_1 + \beta_2 s$$

Vì vậy chúng ta có một quan hệ tuyến tính-logarit giữa lương và số năm học tập. Cũng lý luận tương tự đối với số năm kinh nghiệm. Tuổi của một nhân viên có vẻ như có một loại tác động khác. Chúng ta kỳ vọng thu nhập thấp khi một người còn trẻ, và lương sẽ tăng khi người này tuổi càng lớn hơn, nhưng thu nhập lại giảm sau khi về hưu. Tương quan dạng đường cong lồi này có thể được kiểm định bằng một công thức bậc hai với AGE và AGE². Để tổng quát hóa, chúng ta có thể muốn kiểm định xem học vấn và kinh nghiệm có cùng một dạng tác động bậc hai không. Vì vậy, một mô hình tổng quát có dạng như sau:

$$\ln(\text{WAGE}) = \beta_1 + \beta_2 \text{EDUC} + \beta_3 \text{EXPER} + \beta_4 \text{AGE} + \beta_5 \text{EDUC}^2 + \beta_6 \text{EXPER}^2 + \beta_7 \text{AGE}^2 + u \quad (6.9)$$

DATA6-4 chứa dữ liệu về lương tháng, học vấn tính bằng số năm sau lớp tám, kinh nghiệm tính bằng số năm và tuổi của mẫu gồm 49 cá nhân. Trước tiên chúng ta ước lượng mô hình tuyến tính-logarit trước đó nhưng lại tìm được một số các hệ số hồi quy tuyến tính không có ý nghĩa. Như trước đây, chúng ta thực hiện việc đơn giản hóa tập dữ liệu bằng cách loại bỏ các biến lần lượt mỗi lần một biến (xem Bài Thực hành Máy tính phần 6.7 để tính lại các kết quả này) đến khi các trị thống kê chọn mô hình trở nên xấu hơn. Các kết quả mô hình cuối cùng được trình bày ở đây với trị thống kê t trong dấu ngoặc.

$$\widehat{\ln(\text{WAGE})} = 7,023 + 0,005 \text{ EDUC}^2 + 0,024 \text{ EXPER} \quad (6.10)$$

(76,0) (4,3) (3,9)

$$\bar{R}^2 = 0,33 \quad \text{d.f.} = 46$$

Cả trình độ học vấn bình phương và kinh nghiệm đều rất có ý nghĩa ở mức dưới 0,001. Ý nghĩa của hệ số kinh nghiệm 0,024 là, giữa hai nhân viên có cùng trình độ học vấn, nếu người nào có nhiều hơn một năm kinh nghiệm so với người còn lại thì sẽ được kỳ vọng là có lương cao hơn, trung bình khoảng 2,4 phần trăm (xem Phương trình 6.6 cho phần dịch này). Lưu ý là EDUC có tác động bậc hai với tác động biên tế tăng theo trình độ học vấn. Tuy nhiên, không nên quá xem trọng các kết quả này vì phép đo độ thích hợp khá thấp ngay cả đối với tập dữ liệu chéo. Rõ ràng cần thực hiện nhiều công việc nữa trước khi chúng ta có được những con số chính xác. Chúng ta sẽ nhắc lại mô hình này trong những chương sau và sẽ có nhiều kết quả đáng tin cậy hơn.

Tansel (1994) có một ứng dụng rộng rãi mô hình lương dạng logarit. Vì vậy cần nghiên cứu mô hình này cẩn thận.

● BÀI TẬP THỰC HÀNH 6.8

Sử dụng dữ liệu trong DATA6-4, ước lượng cả mô hình tổng quát trong Phương trình (6.9) và mô hình cuối cùng trong Phương trình (6.10). Thực hiện một kiểm định Wald sử dụng hai mô hình này. Hãy phát biểu *giả thuyết không* và giả thuyết ngược lại và kết luận của bạn dưới dạng văn viết.

Giả sử lương được tính bằng hàng trăm đôla. Việc này sẽ ảnh hưởng đến các hệ số hồi quy như thế nào? Nếu có bất kỳ hệ số nào thay đổi, hãy viết lại các giá trị mới trong Phương trình (6.10)

● BÀI TẬP THỰC HÀNH 6.9

Tính tác động biên tế (dY/dX) và độ co giãn $(X/Y)(dX/dY)$ của mô hình $\ln Y = \beta_1 + \beta_2 X + \beta_3 X^2 + u$

● BÀI TẬP THỰC HÀNH 6.10

Tính tác động biên tế và độ co giãn cho mô hình $\ln Y = \beta_1 + \beta_2 X + \beta_3 (XZ) + u$.

● BÀI TẬP THỰC HÀNH 6.11

Xét mô hình tuyến tính logarit $\ln Y = \beta_1 + \beta_2 X + \beta_3 Z + \beta_4 X^2 + \beta_5 XZ + u$, với X và Z là các biến giải thích. Tìm một biểu thức đại số của độ co giãn của Y theo X . Hãy trình bày cách bạn sử dụng kiểm định Wald để kiểm tra xem các số hạng phi tuyến X^2 và XZ có ý nghĩa thống kê hay không.

● 6.9 So Sánh Các Giá Trị R^2 Giữa Các Mô Hình

Trong Ví dụ 6.5, nếu chúng ta đã sử dụng WAGES như biến phụ thuộc thay vì logarit của biến này, R^2 hiệu chỉnh sẽ là 0,338. Vì R^2 của mô hình tuyến tính-logarit là 0,333, như vậy có phải là mô hình tuyến tính ít nhiều tốt hơn về mức độ thích hợp? Câu trả lời là chắc chắn không, bởi vì thật là không đúng khi so sánh các giá trị R^2 khi mà các biến phụ thuộc là khác nhau. Trong trường hợp tuyến tính, mô hình giải thích 33,8 phần trăm thay đổi của Y , trong khi trong trường hợp tuyến tính-logarit, mô hình giải thích 33,3 phần trăm thay đổi trong $\ln(Y)$. Để sự so sánh là hợp lý, các biến phụ thuộc phải giống nhau.

Tuy nhiên, có một cách so sánh độ thích hợp bằng cách thử sai. Các biến trong trường hợp tuyến tính-logarit như sau:

Bước 1 Ước lượng mô hình tuyến tính-logarit như cách làm thông thường và tính được giá trị thích hợp cho mô hình $\ln(Y)$.

Bước 2 Từ những giá trị này, tạo giá trị trung bình ước lượng cho Y bằng cách phép tính nghịch của logarit, và bảo đảm là thiên lệch hiệu chỉnh như trong Phương trình (6.8). Vậy, chúng ta sẽ có

$$\hat{Y}_t = \exp[\widehat{\ln(Y_t)} + \hat{\sigma}^2/2] \quad (6.11)$$

Bước 3 Tính bình phương của tương quan giữa Y_t và $\hat{Y}_t$. Tương quan này có thể so sánh được với R^2 hiệu chỉnh của một mô hình tuyến tính.

Bước 4 Tính tổng bình phương sai số và phương sai của phần dư bằng cách sử dụng các mối quan hệ

$$ESS = \sum (Y_t - \hat{Y}_t)^2 \text{ và } \hat{\sigma}^2 = \frac{ESS}{n - k}$$

Bước 5 Dùng ESS, tính các trị thống kê lựa chọn mô hình đối với mô hình mới. Các trị thống kê này có thể so sánh được với các trị thống kê của mô hình tuyến tính.

● VÍ DỤ 6.6

Sử dụng dữ liệu trong DATA6-4 và mô hình tuyến tính-logarit được ước lượng trong Ví dụ 6.5, chúng ta đã tiến hành các bước này và đã tính đại lượng R^2 mới và các trị thống kê lựa chọn mô hình (xem chi tiết trong Bài thực hành máy tính 6.8). Kết quả tìm được là R^2 bằng 0,37, lớn hơn

rất nhiều so với giá trị này trong mô hình tuyến tính. Tất cả các trị thống kê lựa chọn mô hình của mô hình tuyến tính-logarit đều thấp hơn so với mô hình tuyến tính. Vì vậy, theo các tiêu chuẩn này, mô hình tuyến tính-logarit có ưu thế hơn một chút.

● 6.10 Mô hình Log-hai lần (hay Log-Log)

Mô hình **Log-hai lần** (hay **Log-Log**) rất phổ biến trong ước lượng các hàm sản xuất cũng như hàm nhu cầu. Nếu Q là số lượng đầu ra của một quá trình sản xuất, K là số lượng vốn đầu vào (số giờ máy), và L là số lượng lao động đầu vào (số giờ nhân công lao động), thì tương quan giữa đầu ra và đầu vào là phương trình hàm sản xuất viết như sau $Q = F(K, L)$. Một đặc trưng chung của dạng hàm này là **hàm sản xuất Cobb-Douglas**, rất nổi tiếng trong lý thuyết kinh tế vi mô. Hàm này có dạng tổng quát sau:

$$Q_t = cK_t^\alpha L_t^\beta$$

với c , α và β là những thông số chưa biết. Lấy logarit hai vế (xem Tính chất 6.1) và thêm vào số hạng sai số, chúng ta có được hàm kinh tế lượng ($\beta_1 = \ln c$):

$$\ln Q_t = \beta_1 + \alpha \ln K_t + \beta \ln L_t + u_t$$

Nếu chúng ta chỉ thay đổi K nhưng giữ L không đổi, thì chúng ta có (sử dụng Tính chất 6.2c)

$$\alpha = \frac{\Delta(\ln Q)}{\Delta(\ln K)} = \frac{(1/Q) \Delta Q}{(1/K) \Delta K} = \frac{K \Delta Q}{Q \Delta K}$$

$100\Delta(\ln Q) = 100\Delta Q/Q$ là phần trăm thay đổi theo Q . Do đó, α là phần trăm thay đổi của Q chia cho phần trăm thay đổi của K . Đây là **độ co giãn của đầu ra theo vốn**. Tương tự như vậy, β là **độ co giãn của đầu ra theo lao động**. Vì vậy, các hệ số hồi quy trong mô hình log-hai lần đơn giản là các độ co giãn tương ứng, có giá trị không đổi. Lưu ý, vì tính chất này, các giá trị bằng số của các hệ số của các biến độc lập khác nhau thì có thể so sánh được trực tiếp. Bảng 6.3 tóm tắt diễn dịch của các hệ số hồi quy trong các mô hình có logarit của các biến.

● Bảng 6.3 Diễn dịch Các tác động biên tế trong các mô hình liên quan đến Logarit

Mô hình	Dạng hàm số	Tác động biên tế	Diễn dịch
Tuyến tính	$Y = \beta_1 + \beta_2 X$	$\Delta Y = \beta_2 \Delta X$	Một đơn vị thay đổi trong X sẽ làm Y thay đổi β_2 đơn vị
Logarit-tuyến tính	$Y = \beta_1 + \beta_2 \ln X$	$\Delta Y = \frac{\beta_2}{100} \left(100 \frac{\Delta X}{X} \right)$	Một phần trăm thay đổi trong X sẽ làm Y thay đổi $\beta_2/100$ đơn vị
Tuyến tính-logarit	$\ln Y = \beta_1 + \beta_2 X$	$100 \frac{\Delta X}{X} = 100\beta_2 \Delta X$	Một đơn vị thay đổi trong X sẽ làm Y thay đổi $100\beta_2$ phần trăm
logarit-hai lần	$\ln Y = \beta_1 + \beta_2 \ln X$	$100 \frac{\Delta Y}{Y} = \beta_2 \left(100 \frac{\Delta X}{X} \right)$	Một phần trăm thay đổi trong X sẽ làm Y thay đổi β_2 phần trăm

Chúng ta có thể có được kết quả thú vị từ mô hình này. Giả sử số lượng vốn và lao động đầu vào tăng gấp đôi. Lúc này đầu ra là

$$Q_1 = c(2K)^\alpha (2L)^\beta = 2^{\alpha+\beta} Q$$

Nếu $\alpha + \beta = 1$, $Q_1 = 2Q$. Vì vậy, đầu ra cũng sẽ tăng gấp đôi nếu $\alpha + \beta = 1$. Đây là điều kiện rất phổ biến về **lợi nhuận không đổi theo qui mô**. Nếu các độ co giãn ước lượng là $P_t = \exp[\hat{Y}_t + (\hat{\sigma}^2/2)]$, chúng thể hiện **lợi nhuận tăng theo qui mô**, và $\hat{\alpha} + \hat{\beta} < 1$ cho thấy **lợi nhuận giảm theo qui mô**. Một kiểm định thông thường đối với lợi nhuận không đổi theo qui mô rất thú vị. *Giả thuyết không* là $H_0: \alpha + \beta = 1$ và giả thuyết đối là $H_1: \alpha + \beta \neq 1$. Trong Phần 4.4, chúng ta phát triển ba kiểm định cho các giả thuyết liên quan đến tổ hợp tuyến tính của các hệ số hồi quy. Để áp dụng Phương pháp 2, định nghĩa $\beta_2 = \alpha + \beta - 1$. Theo *giả thuyết không*, $\beta_2 = 0$. Giải được β , chúng ta có $\beta = \beta_2 + 1 - \alpha$. Thay vào mô hình, ta có

$$\begin{aligned}\ln Q_t &= \beta_1 + \alpha \ln K_t + (\beta_2 + 1 - \alpha) \ln L_t + u_t \\ &= \beta_1 + \alpha (\ln K_t - \ln L_t) + \ln L_t + \beta_2 \ln L_t + u_t\end{aligned}$$

Mô hình này không thể ước lượng được như dạng ở trên vì số hạng $\ln L_t$ không có hệ số. Để ước lượng, các biến như vậy phải được chuyển sang về bên trái. Vì vậy, ta có

$$\ln Q_t - \ln L_t = \beta_1 + \alpha (\ln K_t - \ln L_t) + \beta_2 \ln L_t + u_t$$

Đặt $Y_t = \ln Q_t - \ln L_t$, $X_{t1} = \ln K_t - \ln L_t$, và $X_{t2} = \ln L_t$, mô hình trở thành

$$Y_t = \beta_1 + \alpha X_{t1} + \beta_2 X_{t2} + u_t$$

Để ước lượng mô hình, chúng ta biến đổi các biến ban đầu để tạo ra các biến mới và sau đó hồi quy Y_t theo một số hạng không đổi, X_{t1} và X_{t2} . Kiểm định cần đối với lợi nhuận không đổi theo qui mô chỉ đơn giản là một kiểm định t về hệ số của X_{t2} .

● BÀI TẬP THỰC HÀNH 6.12⁺

Mô tả các bước thực hiện một kiểm định tương tự sử dụng Phương pháp 1 và 3 được mô tả trong Phần 4.4

● BÀI TẬP THỰC HÀNH 6.13

Giả định về lợi nhuận không đổi theo qui mô vẫn được giữ; nghĩa là $\alpha + \beta = 1$. Theo giả thiết này, hãy mô tả bằng cách nào có thể ước lượng được hàm sản xuất Cobb-Douglas.

Ví Dụ Thực Nghiệm: Một Hàm Sản Xuất Nông Nghiệp

Carrasco-Tauber và Moffitt (1992) đã ước lượng một hàm sản xuất loại Cobb-Douglas liên hệ giá trị của sản lượng nông nghiệp (ở dạng log-hai lần) với lao động, đất, nhà, máy móc thiết bị, các đầu vào khác, phân bón và thuốc trừ sâu. Sau đó, họ đã sử dụng hàm sản xuất ước lượng để tính sản lượng biên tế ẩn (được đánh giá bằng trung bình hình học) của mỗi loại đầu vào nông nghiệp. Dữ liệu năm 1987 của các tiểu bang ở Mỹ, trừ Alaska và Hawaii. Tất cả các biến tính bằng hàng ngàn đôla mỗi nông trại, trừ lao động tính bằng ngàn ngày trên mỗi nông trại. Mô hình ước lượng được cho ở đây, với các trị thống kê t trong ngoặc đơn.

$$\begin{aligned}\widehat{\ln Q} &= 4,461 + 0,227 \ln(\text{lao động}) + 0,159 \ln(\text{đất \& nhà}) \\ &\quad (2,11) \quad (2,12) \quad (2,01) \\ &\quad + 0,274 \ln(\text{máy móc thiết bị}) + 0,402 \ln(\text{các đầu vào khác})\end{aligned}$$

$$\begin{array}{cc} (2,42) & (8,55) \\ + 0,082 \ln(\text{phân bón}) + 0,136 \ln(\text{thuốc trừ sâu}) \\ (0,85) & (2,00) \end{array}$$

Trừ độ co giãn của phân bón, các đầu vào khác có ý nghĩa thống kê ở mức 5 phần trăm. Các sản phẩm biên tế ước lượng đối với các đầu vào là \$44,54 mỗi ngày đối với lao động, \$0,04 cho mỗi đôla đất và nhà, \$1,25 cho mỗi đôla máy móc, \$1,29 cho mỗi đôla của các đầu vào khác, \$4,91 cho mỗi đôla phân bón và \$5,66 cho mỗi đôla thuốc trừ sâu. Các tác giả đã ước lượng một số mô hình thay thế bằng cách sử dụng dạng hàm số không được thảo luận trong chương này và đã thu được các đại lượng sản lượng biên tế khác nhau đối với một số đầu vào. Các độc giả quan tâm có thể tham khảo chi tiết trong các bài báo của những tác giả này.

● 6.11 Ứng Dụng: Ước Lượng Độ Co Giãn Của Giao Thông Bằng Xe Buýt

Vì mô hình log-hai lần cho các hệ số hồi quy có độ co giãn không đổi, đây là một hàm rất thông dụng trong ước lượng hàm nhu cầu. Chúng ta minh họa mô hình log-hai lần bằng cách xem lại các yếu tố quyết định của giao thông bằng xe buýt đã tìm hiểu trong Phần 4.6. Tập dữ liệu trong tập tin DATA4-4, và Bài thực hành máy tính Phần 6.9 có hướng dẫn để tính toán các kết quả được trình bày ở đây.

Mô hình cùng với tất cả các biến giải thích sẽ có hệ số không ý nghĩa (ở mức 10%) đối với logarit của FARE, GASPRICE, POP, DENSITY, và LAND AREA, trong đó log của hàm *mật độ dân số* ít có ý nghĩa nhất (nghĩa là có giá trị p cao nhất). Khi biến này được loại bỏ ra ngoài và mô hình được ước lượng lại thì hệ số đối với hàm INCOME, POP, và LANDAREA trở nên có ý nghĩa ở mức dưới 0,001. Nguyên nhân chủ yếu về mặt lý thuyết cho sự thay đổi nghiêm trọng này là do giảm tính đa cộng tuyến và làm tăng bậc tự do kết hợp với một mô hình nhỏ gọn hơn có thể cải thiện được độ chính xác của các hệ số. Chúng ta tiếp tục loại bớt những biến khác mà hệ số của chúng không có ý nghĩa cho đến khi chỉ còn lại những hệ số có ý nghĩa mà thôi. Như trong trường hợp tuyến tính, $\ln(\text{FARE})$ cũng bị loại bỏ. Mô hình sau đây là mô hình cuối cùng với sai số chuẩn để trong ngoặc đơn (không giống như trị thống kê t thông thường):

$$\begin{array}{l} \widehat{\ln(\text{BUSTRAVL})} = 45,846 - 4,730 \ln(\text{INCOME}) + 1,820 \ln(\text{POP}) \\ \quad (9,614) \quad (1,021) \quad (0,236) \\ \quad - 0,971 \ln(\text{LANDAREA}) \\ \quad (0,207) \\ \bar{R}^2 = 0,609 \quad \text{d.f} = 36 \end{array}$$

Một câu hỏi thú vị khác được nêu ra là biến *du lịch bằng xe buýt* có tính chất co giãn hay không co giãn. Nếu giá trị bằng số của độ co giãn này thấp hơn 1 (bỏ qua dấu) thì chúng ta có thể kết luận rằng biến *sử dụng xe buýt* là không co giãn. Nếu nó cao hơn 1 thì có nghĩa là biến có tính co giãn. *Giả thuyết không* chính thức sẽ được áp dụng đối với hệ số này và giả thuyết ngược lại sẽ có tính hai phía. Trị thống kê kiểm định đối với mỗi biến co giãn là

$$\frac{4.73 - 1}{1.021} = 3.65 \quad \frac{1.82 - 1}{0.236} = 3.47 \quad \frac{0.971 - 1}{0.207} = -0.14$$

Từ bảng tra t được trình bày ở mặt trong của trang bìa đầu, chúng ta có giá trị tới hạn với bậc tự do 36 và mức ý nghĩa 0,002 (đối với kiểm định hai phía) nằm giữa 3,307 và 3,385. Vì trị

thống kê t đối với hệ số của biến *thu nhập* và *dân số* tính toán được cao hơn khoảng này nên chúng ta có thể kết luận rằng tính cơ giãn của các biến số trên là có ý nghĩa. Tuy nhiên, ngược lại thì hệ số đối với biến *diện tích đất* là không khác 1, ngay cả với mức 0,8 (giá trị tới hạn nằm trong khoảng 0,225 và 0,256 và giá trị bằng số cao hơn giá trị quan sát được). Trong trường hợp này, chúng ta có thể kết luận rằng biến *diện tích đất* có tính chất **cơ giãn đơn vị**.

● BÀI TẬP THỰC HÀNH 6.14

Thực hiện kiểm định Wald tương tự như bài tập thực hành 6.8

● 6.12 Những Mô Hình Khác *

Mô Hình Logit *

Trong vài trường hợp, biến phụ thuộc có thể nhận giá trị giữa 0 và 1. Ví dụ ta có thể liên hệ giữa phân số của số người bỏ phiếu cho một vị tổng thống nào đó với các yếu tố quyết định của nó. Một cách khác, có thể liên hệ giữa phân số của số người mua xe hơi trong một thời đoạn xác định nào đó với các yếu tố quyết định của nó. Nếu một mô hình hồi quy thông thường nào đó được sử dụng trong những trường hợp như vậy thì không có gì có thể bảo đảm rằng giá trị dự đoán trước sẽ nằm trong khoảng 0 và 1. Để bảo đảm không xảy ra những trường hợp như vậy, người ta thường áp dụng một dạng hàm như sau (được gọi là **đường cong Logistic**):

$$\ln\left[\frac{P}{1-P}\right] = \alpha + \beta X + u$$

trong đó P giá trị của biến phụ thuộc nằm trong khoảng 0 và 1. Mô hình này thường được gọi là mô hình Logit. Rút P từ phương trình trên (bằng cách lấy hàm số mũ lần thứ nhất hai vế phương trình), ta có

$$P = \frac{1}{1 + e^{-(\alpha + \beta X + u)}}$$

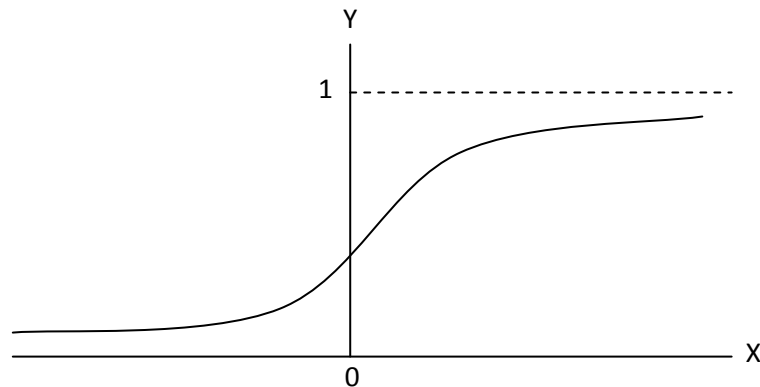
Dễ dàng nhận thấy rằng, nếu giá trị $\beta > 0$ thì P sẽ tiến đến giá trị 0 khi $X \rightarrow -\infty$, và giá trị 1 khi $X \rightarrow \infty$. Vì thế, giá trị P không bao giờ vượt ra khỏi phạm vi [0, 1]. Đường cong Logistic sẽ có hình dáng như trình bày trong hình 6.9. Đường cong này cũng được sử dụng để khớp với dạng *đường cong tăng trưởng*. Ví dụ, doanh số bán hàng của một sản phẩm mới (như tivi có độ nét cao) có thể tăng nhanh trong thời gian đầu nhưng sau đó giảm dần rồi ngưng hẳn. Mô hình Logit được ước lượng dựa trên cách tính hồi quy của hàm $\ln[P/(1 - P)]$ theo một hằng số và biến X. Những mô hình dưới dạng như vậy được mở rộng và phân tích đầy đủ hơn ở chương 12.

Phép biến đổi Box – Cox *

Trong mô hình sau đây, người ta đã sử dụng phép biến đổi được gọi phép biến đổi Box – Cox [xem Box and Cox (1964)]:

$$\frac{Y^\lambda - 1}{\lambda} = \alpha + \beta \frac{X^\lambda - 1}{\lambda} + u$$

● Hình 6.9 Đồ Thị Đường Cong Logistic



Có thể chứng minh được rằng khi giá trị $\lambda = 0$ thì mô hình có thể rút gọn bằng dạng log – hai lần $\ln Y = \alpha + \beta \ln X + u$. Trong trường hợp $\lambda = 1$, chúng ta có được mô hình dạng tuyến tính $Y - 1 = \alpha + \beta (X - 1) + u$ hay $Y = \alpha^* + \beta X + u$, trong đó $\alpha^* = \alpha - \beta + 1$. Khi λ nhận các giá trị khác, chúng ta sẽ có được mô hình phi tuyến tính. Mô hình này có thể ước lượng bằng thủ tục ước lượng thích hợp nhất bằng cách sử dụng chương trình tối ưu hoá phi tuyến tính. Đồ thị hàm số sẽ có nhiều dạng một cách linh động, và người ta có thể kiểm định với λ bằng 0 hay bằng 1, hay với các giá trị khác. Nếu biết trước được phạm vi giá trị của λ từ -1 đến $+1$, chúng ta có thể chọn một giá trị cho λ và dạng các biến mới là $Y^* = (Y^\lambda - 1)/\lambda$ và $X^* = (X^\lambda - 1)/\lambda$. Sau đó chúng ta hồi quy Y^* theo biến X^* và theo các số hạng hằng số và nhận được tổng bình phương sai số. Chúng ta lập lại quy trình này với các giá trị khác nhau của λ và chọn ra giá trị nhỏ nhất trong số các kết quả tổng bình phương sai số. Quy trình dò tìm này có thể thực hiện bằng chương trình hồi quy tuyến tính mà không cần đến chương trình hồi quy phi tuyến tính. Phần mở rộng cho phương pháp này là sử dụng đẳng thức $X^* = (X^\mu - 1)/\mu$, thử kết quả với các giá trị λ và μ (từ -1 đến $+1$), và chọn ra tổ hợp mà cho kết quả tổng bình phương sai số ESS là nhỏ nhất. Muốn biết thêm chi tiết về phép biến đổi Box – Cox, xin tham khảo tác giả Kim và Hill (1993), Showalter (1994) và tác giả Wooldridge (1992).

Tính Phi Tuyến Trong Các Thông Số *

Chúng ta đã xem xét nhiều phương pháp mà trong đó tính chất phi tuyến trong các biến có thể giải quyết tương tự như trong trường hợp hồi quy tuyến tính, tức là các biến này sẽ được biến đổi một cách thích hợp, và như vậy chúng ta sẽ có được một mô hình tuyến tính với hệ số hồi quy chưa biết. Tuy nhiên, cũng có những trường hợp mà cách thức trên không thể thực hiện được. Hàm Box – Cox là một ví dụ cho trường hợp mà mối quan hệ là phi tuyến tính trong các thông số và cũng không thể biến đổi thành dạng tuyến tính ngoại trừ một vài trường hợp đặc biệt kể trên. Một ví dụ khác là hàm sản xuất của các biến thay thế có hệ số co giãn không đổi (CES), được cho như sau:

$$Q = \gamma [\delta K^{-\rho} + (1 - \delta) L^{-\rho}]^{-\nu/\rho} e^u \quad (\gamma > 0, 0 < \delta < 1, \nu > 0, \rho \geq -1)$$

Trong đó thông số chưa biết γ , δ , ν , và ρ có tính chất phi tuyến. Chúng chỉ có thể xác định bằng thủ tục ước lượng thích hợp nhất hoặc bằng các phương pháp bình phương tối thiểu phi tuyến. Tuy nhiên, trong trường hợp này tác giả Kmenta (1986, trang 515) đã áp dụng cách tính bình phương bậc hai gần đúng như sau:

$$\ln Q = \ln \gamma + \gamma \delta \ln K + \nu(1 - \delta) \ln L - \frac{1}{2} \rho \nu \delta (1 - \delta) [\ln K - \ln L]^2 + u$$

$$= \beta_1 + \beta_2 \ln K + \beta_3 \ln L + \beta_4 (\ln K - \ln L)^2 + u$$

Giá trị ước lượng của các thông số ban đầu trong hàm sản xuất CES có thể lấy kết quả từ giá trị ước lượng của các β .

Mặc dù có thể dễ dàng biến đổi các biến và đưa chúng vào trong mô hình nhưng nên tránh việc áp dụng các phương pháp biến đổi mà không phân biệt ý nghĩa ứng dụng của các phương pháp này. Điều cần thiết là hãy xem xét những điểm lý thuyết cơ bản trong các phép biến đổi và giữ cho mô hình càng đơn giản càng tốt.

● 6.13 Phương Pháp Mô Hình Hoá “Từ Tổng Quát Đến Đơn Giản” Hendry/Lse

Như đã phát biểu trước đây, sự hình thành mô hình kinh tế lượng mức chấp nhận được là rất cần thiết đối với những kết luận rút ra từ mô hình đó. Trong các phần và chương trước, chúng ta đã thảo luận về những tiêu chuẩn dùng để đánh giá xem thể nào là một mô hình “tốt”. Quá trình đầu tiên hình thành mô hình được dựa trên lý thuyết kinh tế. Đây là những kiến thức riêng của nhà nghiên cứu về những hành vi cơ bản, về những nghiên cứu khác tương tự, ...v.v. Nhà phân tích cũng có thể có những ý kiến tổng quan về các tác động có thể của tính chất phi tuyến cũng như sự tương tác giữa các biến. Vì không thể có một phương pháp thống nhất chung trong việc xác định mối quan hệ giữa biến phụ thuộc và biến giải thích, nên nhà nghiên cứu thường đưa ra các mô hình thay thế khác và sau đó thực hiện hàng loạt các kiểm định chẩn đoán các mô hình này.

Trong việc đánh giá xem một mô hình kinh tế lượng có được chấp nhận hay không thì dấu của các hệ số hồi quy ước lượng là một trong những đại lượng rất quan trọng, và điều cần thiết là nhà nghiên cứu phải có được những nhận định ban đầu về các giá trị kỳ vọng sẽ nhận được, ít nhất là đối với những biến quan trọng. Ví dụ, giả sử chúng ta đang ước lượng cho mối quan hệ của nhu cầu và kết quả có độ co giãn ước lượng về giá là dương. Đây là một dấu hiệu rõ ràng cho những đặc trưng sai có thể có trong thành phần xác định hoặc cấu trúc sai số (hoặc cả hai) và/hoặc phương pháp luận kinh tế lượng có lỗi sai. Mặc dù giá trị $\bar{R}^2$ là một đại lượng rất hữu dụng khi dùng để đánh giá tính thích hợp, nhưng tin tưởng vào đại lượng này quá mức là một điều không nên. Thông thường những cuộc nghiên cứu chéo đều cho kết quả $\bar{R}^2$ thấp hơn so với các nghiên cứu theo chuỗi thời gian, trong đó hầu hết các biến đều cho thấy xu hướng và có mối tương quan cao giữa chúng. Mặc dù giá trị $\bar{R}^2$ thấp sẽ cho thấy khả năng một số biến bị loại bỏ nhưng người ta cũng không khuyến khích chọn lựa một mô hình dựa trên tiêu chuẩn cực đại giá trị $\bar{R}^2$.

Trong chương 4, chúng ta đã đề cập đến 8 tiêu chuẩn chọn lựa mô hình như là những đại lượng hữu ích dùng để đánh giá xem mô hình này có “tốt” hơn mô hình kia hay không. Một tiêu chuẩn khác nữa cũng thường được dùng để đánh giá mô hình là khả năng dự báo giá trị biến phụ thuộc của mô hình. Nếu giá trị dự báo là một trong những mục tiêu quan trọng đối với nhà kinh tế lượng thì khả năng dự báo của mô hình cần thiết phải được xem xét thận trọng (xin xem thêm ở chương 11).

Trong các phần ứng dụng và ví dụ về tính chất phi tuyến và về các biến có độ trễ trình bày trong chương này, chúng ta đã bắt đầu với một mô hình không có giới hạn tổng quát và sau đó giảm đi bằng cách loại dần từng biến có hệ số ý nghĩa thấp nhất. Cách tiếp cận này, được gọi là phương pháp đi “từ tổng quát đến đơn giản”, được tác giả Hendry (1985) rất tán thành cũng như nhiều nhà kinh tế lượng khác thuộc trường kinh tế London (xin tham khảo thêm ở tác giả Hendry và Richards, 1982, 1983; tác giả Gilbert, 1986, 1989). Cách tiếp cận của họ còn được gọi

là **phương pháp Hendry/ LSE**. Mặc dù tác giả Hendry đã nhấn mạnh đến việc mô hình hoá theo chuỗi thời gian nhưng nguyên lý cũng có thể áp dụng tương tự trên dữ liệu chéo. Ý tưởng cơ bản trong thuật ngữ của tác giả Hendry là có một **quy trình phát ra các dữ liệu (DGP)** nằm ẩn dưới các giá trị của các biến số về kinh tế và công việc của nhà nghiên cứu là dùng các lý thuyết kinh tế, khả năng trực quan, và kinh nghiệm cũng như thông qua một số kiểm định đánh giá mô hình để xem mô hình có thể cải thiện được hay không. Để làm được điều này, nhà nghiên cứu sẽ bắt đầu với một mô hình động tổng quát cho phép họ sử dụng **nhiều tham số hơn**, nghĩa là có nhiều độ trễ và biến (bao gồm cả những số hạng phi tuyến có thể có) hơn những gì mà người ta thường tiến hành và sau đó, nhà nghiên cứu sẽ thực hiện **phép đơn giản hoá dựa trên dữ liệu** bằng kiểm định Wald và kiểm định t. Ví dụ 6.5 và ứng dụng trong phần 6.7 về *chi phí bằng sáng chế* và *chi phí R&D* là những ví dụ cụ thể cho phương pháp luận này. Trong chương 7, 9, và 10, chúng ta sẽ có nhiều ví dụ hơn về phương pháp tiếp cận *từ tổng quát đến đơn giản*. Sự đơn giản hoá mô hình dựa trên dữ liệu dẫn tới kết quả là mô hình đó có **các đặc trưng xúc tích hơn**, có nghĩa là mô hình với ít thông số hơn. Lợi thế của kết quả xúc tích này là (1) tăng độ chính xác cho các giá trị ước lượng vì đã làm giảm tính đa cộng tuyến, (2) có nhiều bậc tự do hơn và vì thế mà giá trị ước lượng sẽ đáng tin cậy hơn, (3) năng lực kiểm định sẽ cao hơn, và (4) mô hình đơn giản hơn, người ta sẽ dễ dàng lĩnh hội hơn là mô hình phức tạp.

● 6.14 Mô Hình Hoá “Từ Đơn Giản Đến Tổng Quát” Bằng Cách Sử Dụng Kiểm Định Nhân Tử Lagrange

Về mặt nguyên tắc thì người ta vẫn ưa thích cách tiếp cận *từ tổng quát đến chi tiết* hơn, nhưng những ứng dụng thực tế thuần túy theo phương pháp này có thể đem lại một số phiền toái. Ví dụ, việc đưa thêm một vài biến mới có tính chất trễ so với các biến hiện tại trong mô hình sẽ khiến cho các biến độc lập khác trở nên tương quan với nhau cao hơn. Như đã trình bày trong các chương trước, khi mức độ tương quan trở nên cao hơn thì việc đo lường các ảnh hưởng riêng lẻ của các biến độc lập trở nên khó khăn hoặc không thể được. Tương tự, thông thường thì nhà nghiên cứu sẽ tự tin hơn với một mô hình đặc trưng cơ bản hơn là việc xây dựng một mô hình tổng quát. Để tránh những vấn đề trên, việc mô hình hoá dựa trên cách tiếp cận **“từ đơn giản đến tổng quát”** sẽ được bắt đầu từ phía ngược lại với đặc trưng cơ bản về những gì mà nhà nghiên cứu cảm thấy tự tin và rồi câu hỏi đặt ra là liệu có nên đưa thêm biến vào trong mô hình hay không. Một công cụ chẩn đoán thường được các nhà nghiên cứu sử dụng trong cách tiếp cận này là **kiểm định với nhân tử Lagrange (LM)**. Tuy nhiên, trước khi đi sâu vào vấn đề trên, điều cần thiết là phải có được cách nhìn tổng quát đầy đủ về những phương pháp khác nhau đối với kiểm định đặc trưng mô hình.

Có nhiều phương pháp chính thức được dùng trong việc kiểm định giả thuyết, nhưng phương pháp được dùng thường xuyên nhất bao gồm kiểm định bằng nhân tử Lagrange (LM), **kiểm định tỷ số thích hợp (LR)**, và kiểm định Wald. Trong các phần được trình bày trước đây thì kiểm định được sử dụng nhiều nhất là kiểm định Wald. Trong phần này, người ta sẽ tập trung vào kiểm định LM như là một phương án thay thế trong việc kiểm định các đặc điểm của mô hình. Kiểm định LR sẽ được thảo luận thêm trong phần phụ lục. Trong tất cả cách tiếp cận trên, người ta đã xây dựng lên hai mô hình, *mô hình giới hạn* và *mô hình không giới hạn*. Mô hình giới hạn được thiết lập bằng cách đưa các giới hạn tuyến tính hoặc phi tuyến tính vào trong các thông số của mô hình và tương ứng với giả thuyết không. Mô hình không giới hạn là một giải pháp thay thế khác.

Trong các chương trước đây, chúng ta đã dùng phương pháp Wald để kiểm định giả thuyết giữa mô hình không giới hạn và mô hình giới hạn, trong đó chúng ta có thể bỏ qua một số biến.

Điều này được trình bày nhiều trong các ứng dụng theo chuỗi thời gian và chéo. Kiểm định Wald bắt đầu với một phương án thay thế (mô hình không giới hạn) và đưa ra câu hỏi là *giả thuyết không* (mô hình giới hạn) có được ưa thích hơn hay không. Kiểm định tỷ số thích hợp là một phép so sánh trực tiếp giữa hai giả thuyết. Nguyên lý của nhân tử Lagrange bắt đầu với *giả thuyết không* và hỏi xem có sự dịch chuyển về hướng các giải pháp ưa thích hơn. Hay nói cách khác là thủ tục LM sẽ xác định được một mô hình đơn giản hơn nhưng xúc tích hơn và câu hỏi đặt ra là có cần thiết phải thêm biến mới hay không. Vì vậy mà chúng ta sẽ bắt đầu với cách thiết lập cơ bản và tiến hành kiểm định việc thêm các biến mới thay vì bắt đầu với một mô hình đầy đủ và kiểm định xem có nên bỏ bớt đi một số các biến hay không. Phương pháp LM thì khá tổng quát nên có thể áp dụng vào trong những tình huống khác sẽ được mô tả trong các chương sau. Cả hai phương pháp từ tổng quát đến đơn giản và từ đơn giản đến tổng quát là những phương pháp rất hữu dụng và người ta đề nghị rằng nên sử dụng cả hai phương pháp để có được một kết luận vững chắc hơn.

Có nhiều bài nghiên cứu đề cập đến kiểm định LM như của tác giả Aitcheson và Silvey (1958), Silvey (1959), Berndt và Savin (1977), Godfrey (1978), Buse (1982), và của tác giả Engle (1982, 1984). Tất cả các bài nghiên cứu trên đòi hỏi người đọc phải có kiến thức về đại số tuyến tính. Trong chương này, chúng ta sẽ được trình bày một cách tóm lược các loại kiểm định cùng với khả năng ứng dụng của chúng. Một vài kết quả lý thuyết khác cũng sẽ được tóm tắt trong phần phụ lục của chương cùng với một bảng giải thích các thuật ngữ về *nhân tử Lagrange* và *tỷ số thích hợp*. Một sự so sánh ba phương pháp này dựa trên cấp số nhân cũng được đưa thêm vào trong phần phụ lục và được diễn giải bằng ví dụ đơn giản. Trong chương, chúng ta sẽ xác định các bước cần thiết để thực hiện một kiểm định LM và ứng dụng chúng với dữ liệu thực tế. Mặc dù kiểm định LM là kiểm định trên mẫu lớn, nhưng người ta nhận thấy nó cũng rất hữu ích ngay cả trong trường hợp số mẫu quan sát chỉ cỡ 30. Kiểm định Wald cũng có thể áp dụng đối với trường hợp số mẫu quan sát nhỏ. Kiểm định tỷ số thích hợp đôi khi được áp dụng cho cỡ mẫu nhỏ. Những điểm này sẽ được trình bày chi tiết hơn trong phần phụ lục.

Những giả thuyết với ngụ ý bác bỏ bớt các biến là trường hợp đặc biệt của **giả thuyết lồng vào nhau**. Ở dạng lồng vào nhau, mô hình giới hạn trở thành một mô hình con của một mô hình không giới hạn khác. Các **giả thuyết không lồng vào nhau** là những mô hình hoàn toàn khác nhau mà trong đó một mô hình không thể trở thành mô hình con của một mô hình khác. Ví dụ, việc bác bỏ một số biến và thêm một số biến khác sẽ hình thành nên mô hình không lồng vào nhau. Trong cuốn sách này, chúng ta chỉ tập trung vào các giả thuyết lồng vào nhau. Độc giả có quan tâm đến kiểm định giả thuyết không lồng vào nhau có thể tham khảo các bài viết của tác giả Davidson và MacKinnon (1981, 1982) và của MacKinnon (1983) nhưng chú ý rằng chúng đòi hỏi phải có kiến thức về đại số tuyến tính.

Kiểm Định Nhân Tử Lagrange Khi Thêm Các Biến

Thủ tục kiểm định LM sẽ dễ hiểu hơn trong trường hợp nhà phân tích kinh doanh hay kinh tế muốn biết có nên đưa thêm biến vào trong mô hình hay không (ví dụ đưa các số hạng phi tuyến tính và số hạng tương tác). Hãy xem xét mô hình không giới hạn và giới hạn sau đây:

$$(R) \quad Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_m X_m + u$$

$$(U) \quad Y = \beta_1 + \beta_2 X_2 + \dots + \beta_m X_m + \beta_{m+1} X_{m+1} + \dots + \beta_k X_k + v$$

Trong mô hình U, có $(k - m)$ biến mới $X_{m+1}, X_{m+2}, \dots, X_k$ (ví dụ, các biến phi tuyến tính) đã được thêm vào. Điểm thú vị ở *giả thuyết không* này là hệ số hồi quy của các biến thêm vào là bằng 0. Các bước thực hiện kiểm định LM như sau:

- Bước 1** $H_0: \beta_{m+1} = \beta_{m+2} = \dots = \beta_k = 0$. H_1 : trong số các β trên tồn tại ít nhất một β khác không.
- Bước 2** Ước lượng giá trị R của mô hình giới hạn
- Bước 3** Thu được phần dư ước lượng của mô hình như sau

$$\hat{u}_R = Y - \hat{\beta}_1 - \hat{\beta}_2 X_2 - \hat{\beta}_3 X_3 - \dots - \hat{\beta}_m X_m$$

Giả sử các đặc trưng “đúng” thuộc về mô hình U; trong trường hợp này thì nên đưa các biến $X_{m+1}, X_{m+2}, \dots, X_k$ vào trong mô hình. Các ảnh hưởng của chúng sẽ được quan sát bằng phần dư $\hat{u}_R$. Vì vậy, $\hat{u}_R$ được xem như có liên hệ với những biến bị loại bỏ. Nói cách khác, nếu chúng ta hồi quy giá trị $\hat{u}_R$ theo các biến trên, chúng ta sẽ có được sự thích hợp tốt, một chỉ số chứng tỏ rằng ít nhất có một số biến trong số $X_{m+1}, X_{m+2}, \dots, X_k$ nên được đưa thêm vào trong mô hình. Lý luận này dẫn đến bước tiếp theo.

- Bước 4** Hồi quy biến $\hat{u}_R$ theo hằng số và tất cả các biến X, bao gồm các biến trong mô hình giới hạn. Điều này có nghĩa là hồi quy theo tất cả các biến độc lập trong mô hình không giới hạn. Chúng ta sẽ xem bước hồi quy thứ hai này là hồi quy phụ. Tác giả Engle (1982) đã chứng minh rằng, đối với các mẫu quan sát lớn, cỡ mẫu (n) nhân với giá trị R^2 không hiệu chỉnh trong hồi quy phụ này sẽ có phân phối Chi-square với bậc tự do tương đương với *số giới hạn trong giả thuyết không* (điều này được trình bày trong phần phụ lục 6.A.3 đối với trường hợp hồi quy đơn). Vì thế, trong bài toán của chúng ta, $nR^2 \sim \chi^2_{k-m}$. Lý do đưa các biến ban đầu vào trong hồi quy phụ là để giá trị thống kê kiểm định có được dạng thuận tiện. Nếu $nR^2 > \chi^2_{k-m}(\alpha)$ thì vị trí của một điểm nào đó trên đồ thị phân phối χ^2_{k-m} mà diện tích bên phải điểm đó bằng (α) , chính là mức ý nghĩa mà chúng ta sẽ bác bỏ *giả thuyết không* cho rằng tất cả hệ số hồi quy mới đều bằng không. Một cách khác, chúng ta sẽ tính giá trị $p = P(\chi^2_{k-m} > nR^2)$ và bác bỏ giả thuyết không nếu giá trị p thấp hơn mức ý nghĩa. Nói cách khác, chúng ta có thể kết luận rằng ít nhất có một số trong các biến mới nên được đưa thêm vào trong mô hình. Các giá trị p của mỗi hệ số riêng lẻ cũng có thể kết luận rằng biến nào nên được đưa vào.

Sử Dụng Hồi Quy Phụ Để Xác Định Các Biến Đưa Thêm Vào Mô Hình Cơ Bản

Hồi quy phụ cung cấp các thông tin về những biến mới đang được xem xét trở thành đối tượng đưa thêm vào trong mô hình cơ bản. *Thực tế, hệ số ước lượng và các trị số thống kê liên quan đến các biến mới sẽ được đưa thêm vào trong hồi quy phụ của kiểm định LM cũng giống với các đại lượng có được từ mô hình không giới hạn hoàn toàn. Mô hình không giới hạn này sẽ được bắt đầu cùng với phương pháp từ tổng quát đến đơn giản.* Điểm này đã được chứng minh đầy đủ bởi tác giả Ramanathan (1986) và cũng được diễn giải một cách thực nghiệm qua ví dụ 6.7. Mặc dù không có một hướng dẫn lý thuyết rõ ràng nào để có thể chọn ra các biến từ danh sách hồi quy phụ, nhưng người ta có thể sử dụng một quy tắc đơn giản là đưa những biến nào có hệ số hồi quy hoặc giá trị p nhỏ hơn 0,5. Quy tắc này tỏ ra bảo thủ hơn quy tắc chỉ chọn những biến có ý nghĩa thực sự. Như sẽ trình bày trong ví dụ 6.7 và 6.8, việc sử dụng tiêu chuẩn mức ý nghĩa một cách nghiêm ngặt sẽ có khuynh hướng bỏ qua các biến có thể có ý nghĩa so với những biến có

mức ý nghĩa tối thiểu bị loại bỏ. Thủ tục này tương đương với việc ước lượng toàn bộ mô hình tổng quát rồi loại bỏ tất cả các biến có giá trị p tương ứng cao hơn 0,5.

● VÍ DỤ 6.7

Phương pháp kiểm định LM được diễn giải trước tiên với dữ liệu để ước lượng nhu cầu sử dụng truyền hình cáp trình bày trong phần DATA4-8. Dữ liệu này thuộc dạng dữ liệu chéo thu thập từ 40 thành phố (các biến được định nghĩa sau đây và sẽ được mô tả chi tiết hơn trong phần phụ lục 1).

sub = số người đăng ký trên mỗi hệ thống (tính bằng đơn vị ngàn)
home = số lượng nhà mà hệ thống đi qua
inst = chi phí lắp đặt tính bằng đô la
svc = chi phí dịch vụ hằng tháng của mỗi hệ thống tính bằng đô la
tv = số lượng tín hiệu truyền hiệu tải bởi mỗi hệ thống cáp
age = tuổi thọ của mỗi hệ thống tính theo năm
air = số lượng tín hiệu truyền hình tự do nhận được
y = thu nhập tính bằng đô la trên mỗi đầu người

Bảng 6.4 trình bày các kết quả máy tính riêng phần có kèm giải thích, cho độc giả thấy các bước vừa mô tả. Để có được toàn bộ kết quả, hãy thực hành bài tập máy tính phần 6.10.

Mặc dù các trị thống kê kiểm định LM cho trong ví dụ đều cho thấy có ý nghĩa, nhưng đôi khi phép kiểm định cũng có thể cho các dấu trái ngược. Điểm này sẽ được trình bày trong ví dụ tiếp theo.

● Bảng 6.4 Kết Quả Máy Tính Riêng Phần Có Kèm Giải Thích Trong Ví Dụ 6.7

[Trước tiên, hãy ước lượng mô hình cơ bản bằng cách hồi quy biến sub theo biến constant, home, inst, svc, tv, age, air, và biến y. Sau đó phát ra phần dư u_t . Hồi quy phụ trình bày ở đây sẽ hồi quy phần dư u_t theo các biến trong mô hình cơ bản và cộng tất cả các số hạng bình phương của các biến, biểu diễn dưới dạng sq_x (ví dụ $sq_home = home^2$).]

Dependent variable: Ut

	VARIABLE	COEFFICIENT	STDERROR	T STAT	2Prob(t > T)	
	E					
0)	const	-481.4363	264.2862	-1.822	0.080496	*
2)	home	0.0339	0.0839	0.404	0.689961	
3)	inst	0.9184	2.1242	0.432	0.669195	
4)	svc	10.1055	19.1942	0.526	0.603188	
5)	tv	-1.4180	2.6542	-0.534	0.597895	
6)	age	-2.5507	1.4623	-1.744	0.093391	*
7)	air	23.8229	5.2392	4.547	0.000121	***
8)	y	0.0829	0.0526	1.576	0.127509	
9)	sq_home	0.0002207	0.0002839	0.778	0.444146	
10)	sq_inst	-0.0210	0.0655	-0.321	0.750748	
11)	sq_svc	-0.7790	1.2854	-0.606	0.549977	
12)	sq_tv	0.0484	0.1017	0.476	0.637925	
13)	sq_age	0.1393	0.0734	1.898	0.069252	*
14)	sq_air	-1.5823	0.3732	-4.240	0.000267	***
15)	sq_y	-4.547e-006	2.8346e-006	-1.604	0.121287	

Unadjusted R-squared 0.550

Adjusted R-squared

0.298

[Trị thống kê LM = số lần quan sát nhân với giá trị chưa hiệu chỉnh $R^2 = 21,992652$.]

Chi-square (7): area to right of 21.992652 = 0.002548.

[*Giả thuyết không* đối với kiểm định LM là hệ số của tất cả bảy biến bình phương được đưa thêm vào mô hình sẽ bằng 0] (vì vậy, bậc tự do là 7). Giá trị p bằng 0,002548 cho thấy chúng ta “an toàn” khi quyết định bác bỏ *giả thuyết không* và kết luận rằng có ít nhất một vài trong số các biến được đưa thêm vào thuộc về mô hình. (Sử dụng máy tính cầm tay để kiểm tra trị thống kê kiểm định LM và thực hiện kiểm định Chi-square bằng cách sử dụng mức ý nghĩa 1% với bảng phân phối Chi-square).

● Bảng 6.4 (Tiếp theo)

Hồi quy phụ sẽ giúp chúng ta quyết định những biến mới nào sẽ được đưa thêm vào trong mô hình. Tuy nhiên, người ta cũng không có những hướng dẫn nào về mặt lý thuyết đối với việc chọn lựa trong thực tế. Vì thế, chúng ta sẽ sử dụng quy tắc tùy ý là đưa những biến mới có giá trị p nhỏ hơn 0,5 vào trong mô hình, tương đương với mức ý nghĩa 50%. Quy tắc này bảo thủ hơn cả khi ta sử dụng điểm ngưỡng 10% mà chúng ta sử dụng lâu nay và nó cũng được thiết kế để cực tiểu hoá các sai lệch có thể có từ các biến bị bỏ qua với nguyên nhân không đưa đủ biến vào trong mô hình. Theo quy tắc “0,5”, trị bình phương của biến *home*, *age*, *air*, và *y* được đưa thêm vào trong mô hình cơ bản. Điều này sẽ được thực hiện tiếp theo đây. Chú ý rằng biến phụ thuộc hiện tại là *sub*. Lỗi mà người ta hay phạm phải ở điểm này là điều chỉnh biến u_t như là biến phụ thuộc hoặc đưa biến này vào trong nhóm các biến độc lập. Điều này rõ ràng là sai và không có ý nghĩa.]

Dependent variable: sub

	VARIABLE	COEFFICIENT	STDERROR	T STAT	2Prob(t > T)	
0)	const	-407.0791	211.7804	-1.922	0.064813	*
2)	home	0.4319	0.0792	5.451	0.000008	***
3)	inst	-0.1821	0.3957	-0.460	0.648969	
4)	svc	0.2123	1.9666	0.108	0.914822	
5)	tv	0.6962	0.5292	1.315	0.199029	
6)	age	-1.0718	1.2305	-0.871	0.391149	
7)	air	18.1986	4.8824	3.727	0.000868	***
8)	y	0.0757	0.0476	1.591	0.122767	
9)	sq_home	0.0002240	0.0002689	0.833	0.411944	
13)	sq_age	0.1174	0.0580	2.025	0.052478	*
14)	sq_air	-1.5579	0.3383	-4.605	0.000082	***
15)	sq_y	-4.049e-006	2.5562e-006	-1.584	0.124383	
Mean of dep. Var.		24.509	S.D. of dep. Variable		33.537	
Error Sum of Sq (ESS)		2307.1870	Std Err of Resid. (sgmahat)		9.0774	
Unadjusted R-squared		0.947	Adjusted R-squared		0.927	
F-statistic(11,28)		45.8496	p-value for F()		0.000000	
Durbin-Watson stat.		1.943	First-order autocorr. coeff		0.001	

MODEL SELECTION STATISTICS

SGMASQ	82.3995 AIC	105.099 FPE	107.119
HQ	126.229 SCHWARZ	174.438 SHIBATA	92.2875
GCV	117.714 RICE	144.199	

Excluding the constant, p-value was highest for variable 4 (svc)

[Phần cuối của thủ tục là làm gọn mô hình dựa trên dữ liệu mà chúng ta nhận được trước đây. Điều này được thực hiện bằng cách loại bỏ liên tiếp các biến có giá trị p cao nhất, nhưng phải loại bỏ từng biến một. Để tránh làm lẫn nếu trình bày quá nhiều kết quả không cần thiết nên trên trang tài liệu này chỉ đưa ra mô hình cuối cùng.]

● Bảng 6.4 (Tiếp theo)

	VARIABLE	COEFFICIENT	STDERROR	T STAT	2Prob(t > T)	
0)	const	-562.6761	158.0817	-3.559	0.001185	***
2)	home	0.4960	0.0283	17.525	0.000000	***
6)	age	-1.5575	0.9037	-1.723	0.094460	*
7)	air	17.3047	4.3410	3.986	0.000364	***
8)	y	0.1108	0.0348	3.186	0.003215	***
13)	sq_age	0.1392	0.0438	3.181	0.003251	***
14)	sq_air	-1.4177	0.2919	-4.856	0.000030	***
15)	sq_y	-5.948e-006	1.8798e-006	-3.164	0.003399	***
Mean of dep. Var.		24.509	S.D. of dep. Variable		33.537	
Error Sum of Sq (ESS)		2521.9340	Std Err of Resid. (sgmahat)		8.8775	
Unadjusted R-squared		0.943	Adjusted R-squared		0.930	
F-statistic(11,28)		74.9412	p-value for F()		0.000000	
Durbin-Watson stat.		2.069	First-order autocorr. coeff		-0.051	

MODEL SELECTION STATISTICS

SGMASQ	78.8104 AIC	94.0571 FPE	94.5725
HQ	106.275 SCHWARZ	131.852 SHIBATA	88.2677
GCV	98.513 RICE	105.081	

[Để làm rõ sự tương phản giữa phương pháp *từ đơn giản đến tổng quát* này với phương pháp Hendry/ LSE mô hình hoá *từ tổng quát đến đơn giản*, chúng ta sẽ ước lượng mô hình tổng quát nhất bao quát được số hạng tuyến tính và bình phương bậc hai. Một chú ý thú vị là các hệ số và sai số chuẩn của bình phương các số hạng thêm vào cũng *giống* như các số hạng trong hồi quy phụ trình bày ở trên. Muốn biết thêm cách chứng minh về mặt lý thuyết rằng tại sao trường hợp này luôn luôn xảy ra, hãy tham khảo tác giả Ramanathan (1986).]

Dependent variable: sub

	VARIABLE	COEFFICIENT	STDERROR	T STAT	2Prob(t > T)	
0)	const	-488.2440	264.2862	-1.847	0.076556	*
2)	home	0.4394	0.0839	5.238	0.000020	***
3)	inst	0.3920	2.1242	0.185	0.855089	
4)	svc	12.1443	19.1942	0.633	0.532671	
5)	tv	-0.6615	2.6542	-0.249	0.805230	
6)	age	-1.3571	1.4623	-0.928	0.362229	

7)	air	18.7117	5.2392	3.572	0.001475	***
8)	y	0.0845	0.0526	1.608	0.120423	
9)	sq_home	0.0002207	0.0002839	0.778	0.444146	
10)	sq_inst	-0.0210	0.0655	-0.321	0.750748	
11)	sq_svc	-0.7790	1.2854	-0.606	0.549977	
12)	sq_tv	0.0484	0.1017	0.476	0.637925	
13)	sq_age	0.1393	0.0734	1.898	0.069252	*
14)	sq_air	-1.5823	0.3732	-4.240	0.000267	***
15)	sq_y	-4.547e-006	2.8346e-006	-1.604	0.121287	

● Bảng 6.4 (Tiếp theo)

Error Sum of Sq (ESS)	2216.6660	Std Err of Resid. (sgmahat)	9.4163
Unadjusted R-squared	0.949	Adjusted R-squared	0.921

[Theo chiến lược giản lược mô hình dựa trên dữ liệu, chúng ta lần lượt loại bỏ các biến có hệ số không ý nghĩa. Mô hình cuối cùng được xác định theo cách này giống như mô hình tìm được trước đây theo phương pháp từ đơn giản đến tổng quát. Như vậy, trong ví dụ này, hai phương pháp là tương đương. Vì điều này không phải lúc nào cũng xảy ra, người ta đề nghị sử dụng cả hai phương pháp và thực hiện kiểm tra chéo. Tuy nhiên, nếu cần phải chọn một trong hai cách tiếp cận, cách tiếp cận Hendry/LSE thường được sử dụng hơn vì biện pháp tiếp cận này chắc chắn hơn và không phụ thuộc vào quy tắc 0,5 chủ quan khi chọn các biến từ việc hồi quy phụ. Tuy nhiên, trong chương 8, 9, và 10 chúng ta sẽ thấy rằng kiểm định LM là một thủ tục kiểm định cực kỳ mạnh trong những tình huống khác]

● BÀI TẬP THỰC HÀNH 6.15

Trong ví dụ 6.7, chúng ta loại bỏ các biến dựa trên mức ý nghĩa của các hệ số hồi quy của chúng. Bắt đầu từ mô hình tổng quát nhất theo phương pháp Hendry/LSE và loại bỏ từng biến một như trước đây, nhưng giữ lại biến thu nhập (income), phí dịch vụ hàng tháng (monthly service charge), và phí lắp đặt (installation fee) cho đến cuối cùng bởi vì chúng là các số đo về thu nhập và giá trong phương trình đường cầu và vì vậy có ý nghĩa về mặt lý thuyết. So sánh mô hình cuối cùng thu được (về mặt tiêu chí chọn lựa và mức ý nghĩa của các hệ số) với mô hình cuối cùng ở bảng 6.4. Bạn thấy những khác biệt gì? Bạn sẽ đề nghị sử dụng mô hình nào để thực hiện diễn dịch cuối cùng? Hãy sử dụng mô hình đó để diễn dịch các kết quả.

● VÍ DỤ 6.8

Ví dụ minh họa thứ hai này sẽ trình bày cách thức áp dụng kiểm định LM cho bài tập được nghiên cứu ở ví dụ 6.5, nghĩa là, trong mô hình tuyến tính lôgarit về tiền lương. Bảng 6.5 trình bày kết quả máy tính có chú thích về trường hợp này (xem chi tiết ở Phần Thực Hành Máy Tính 6.11). Giá trị R^2 không hiệu chỉnh của hồi quy phụ chỉ bằng 0,079, với trị thống kê nR^2 bằng 3,86. Theo giả thuyết không các số hạng bậc hai có hệ số bằng 0, giá trị này tuân theo phân phối Chi bình phương với 3 bậc tự do. Giá trị p-value bằng 0,28 cho thấy rằng chúng ta không thể bác bỏ giả thuyết H_0 một cách an toàn. Điều này hàm ý rằng không một biến mới nào có hệ số có ý nghĩa. Tuy nhiên, lưu ý rằng giá trị p-value của hệ số của biến $EDUC^2$ có ý nghĩa tại mức 7,33%, đây là mức ý nghĩa chấp nhận được. Vì vậy, hồi quy phụ đề nghị biến này được đưa vào mô hình (quy tắc p-value 0,5 cũng sẽ chọn biến này và loại tất cả các biến còn lại). Ngược lại, kiểm định nR^2 cho thấy không có biến nào cần đưa vào mô hình. Do đó, kiểm định LM đưa ra các kết luận trái ngược nhau về mức độ quan trọng của việc thêm một biến mới vào mô hình ban đầu.

● Bảng 6.5 Báo Cáo Có Chú Giải Một Phần In Từ Máy Tính Cho Ví Dụ 6.8

[Ước lượng hồi quy phụ]

	VARIABLE	COEFFICIENT	STDERROR	T STAT	2Prob(t> T)
0)	const	0.4934	0.8092	0.610	0.545334
2)	EDUC	-0.1576	0.0864	-1.824	0.075224 *
3)	EXPER	-0.0088	0.0245	-0.361	0.719991
4)	AGE	-0.0008179	0.0338	-0.024	0.980822
7)	sq_EDUC	0.0115	0.0063	1.837	0.073294 *
8)	sq_EXPER	0.0004293	0.0011	0.384	0.703130
9)	sq_AGE	0.0000211	0.0003814	0.055	0.956041

Unadjusted R-squared = 0.079

Value of the LM statistic = 3.861657

Chi-square (3): area to the right of 3.861657 = 0.276796

[Lưu ý rằng p-value cho biết không thể bác bỏ *giả thuyết không*, nhưng hệ số của biến EDUC² có ý nghĩa tại mức ý nghĩa 7,33%]

Trong ví dụ này, phương pháp từ tổng quát đến đơn giản sẽ tốt hơn vì sẽ tránh được sự mơ hồ. Tuy nhiên, nếu chúng ta sử dụng quy tắc p-value 0,5 trong việc chọn biến, hai phương pháp là như nhau.

Ví dụ trên giải thích rằng, mặc dù kiểm định LM là một công cụ chẩn đoán hữu ích trong việc xây dựng một khung phân tích từ đơn giản đến tổng quát, trong một số trường hợp tính hữu ích của chúng bị hạn chế. Tuy nhiên, chúng ta sẽ thấy ở các chương 8, 9, và 10 rằng kiểm định LM rất mạnh trong nhiều tình huống.

● 6.15 Thủ Tục RESET Ramsey Để Xác Định Sai Số Đặc Trưng Hồi Quy

Ramsey (1969) đề ra một phương pháp khác để kiểm định đặc trưng của mô hình. Nó được gọi là RESET (**kiểm định sai số đặc trưng hồi quy**). Việc áp dụng thủ tục này cũng dễ dàng như việc áp dụng kiểm định LM được mô tả ở phần trước. Các bước của thủ tục RESET được thực hiện như sau:

Bước 1: Ước lượng mô hình theo thủ tục OLS và lưu các giá trị được thích hợp $\hat{Y}_t$.

Bước 2: Thêm các biến $\hat{Y}_t^2$, $\hat{Y}_t^3$, và $\hat{Y}_t^4$ vào mô hình ở bước 1 và ước lượng mô hình mới

Bước 3: Thực hiện kiểm định F Wald cho việc loại bỏ ba biến mới trong bước 2. Nếu *giả thuyết không* cho rằng các biến mới không có hiệu ứng bị bác bỏ, đó chính là dấu hiệu của sai số đặc trưng.

Cơ sở của thủ tục RESET Ramsey là các phần dư ước lượng ($\hat{u}_t$) mà đại diện cho các tác động biến bị loại bỏ có thể được tính xấp xỉ bằng tổ hợp tuyến tính của các lũy thừa của các giá trị được thích hợp. Nếu các lũy thừa này có các tác động có ý nghĩa, thì mô hình gốc được coi như đã bị đặc trưng sai. Tuy nhiên, nhược điểm chính của phương pháp RESET là kiểm định sẽ không chỉ ra được loại đặc trưng sai và cũng không gợi ý dạng hàm thích hợp cần sử dụng. Tuy vậy, kiểm định này bổ sung cho các kiểm định Wald và LM được ứng dụng để kiểm định các tác động động và phi tuyến đặc thù. Điểm này được minh họa trong ví dụ dưới đây.

● VÍ DỤ 6.9

Trong ví dụ 6.2, chúng ta đã sử dụng tập dữ liệu DATA6-1 để ước lượng hàm chi phí trung bình của một công ty sản xuất. Đầu tiên chúng ta sử dụng thủ tục RESET để kiểm định xem quan hệ tuyến tính đã đủ thể hiện bản chất bài toán chưa (xem Phần Thực Hành Máy Tính 6.12 về các bước để chạy lại các kết quả của ví dụ này). Như vậy, chúng ta hồi quy biến UNITCOST theo số hạng hằng số, OUTPUT, và INPCOST, và lưu các trị ước lượng của Y ($\hat{Y}$). Tiếp theo chúng ta tiến hành hồi quy biến UNITCOST theo các biến trên và thêm các lũy thừa của trị ước lượng Y và thực hiện kiểm định F Wald cho các lũy thừa của $\hat{Y}$. Trị thống kê tính toán F bằng 3,7447, trị này, theo giả thuyết không là các biến được thêm vào không tác động đến UNITCOST, có phân phối F với 3 bậc tự do ở tử số và 14 ($=20 - 6$) bậc tự do ở mẫu số. Giá trị p-value tương ứng là 0,036407, có nghĩa rằng các hệ số của các biến được thêm vào có ý nghĩa kết hợp dưới mức 5%. Nói cách khác, thủ tục RESET chỉ ra sự đặc trưng sai mô hình. Trong ví dụ 6.2, chúng ta thêm vào số hạng bậc hai đối với biến OUTPUT và nhận thấy biến đó có một tác động có ý nghĩa (điều này cũng chẳng có gì ngạc nhiên bởi vì lý thuyết cho chúng ta thấy đường cong chi phí trung bình có dạng tổng quát hình chữ U). Điều này đòi hỏi trước tiên phải hồi quy biến UNITCOST theo một số hạng hằng số, OUTPUT, INPCOST, và OUTPUT² và lưu các trị ước lượng Y như trước đó. Sau đó thêm lũy thừa của trị Y ước lượng vào làm biến giải thích và sử dụng kiểm định F Wald đối với các biến được thêm vào này. Trị thống kê F là 0,4826 với trị p-value bằng 0,7. Vì giá trị này quá cao, chúng ta không thể bác bỏ giả thuyết không rằng các biến được thêm vào không có ảnh hưởng đến biến UNITCOST. Như vậy, phương pháp RESET cho rằng mô hình cuối cùng trong Ví dụ 6.2 có thể không bị đặc trưng sai.

● BÀI TẬP THỰC HÀNH 6.16

Áp dụng thủ tục RESET để kiểm định đặc trưng sai trong mô hình cuối cùng ở phần bài tập ví dụ 6.7.

● BÀI TẬP THỰC HÀNH 6.17

Làm tương tự cho mô hình cuối cùng trong Ví dụ 6.5

Tóm Tắt

Mô hình hồi quy tuyến tính đơn cũng có thể được sử dụng để giải quyết các quan hệ không tuyến tính, với điều kiện là mô hình tuyến tính trong các thông số. Các dạng hàm khác nhau thường được sử dụng là mô hình bán lôgarít hoặc tuyến tính-lôgarít, quan hệ lôgarít-tuyến tính, mô hình lôgarít hai lần, và phép biến đổi nghịch đảo.

Bình phương và lũy thừa cao hơn của các biến độc lập, hoặc độ trễ của các biến, dễ dàng được xem xét trong mô hình miễn là các hệ số hồi quy chưa biết dường như có dạng tuyến tính. Chỉ cần biến đổi dữ liệu thích hợp và đưa chúng vào trong mô hình. Tác động cận biên của một biến có thể được tạo ra để phụ thuộc vào một biến giải thích khác thông qua các số hạng tương tác. Một số mô hình không thể biến đổi được về dạng có các thông số tuyến tính. Trong những trường hợp như vậy, thủ tục ước lượng bao gồm phương pháp bình phương nhỏ nhất phi tuyến hoặc phương pháp thích hợp cực đại.

Việc so sánh giá trị R^2 của các mô hình sẽ không đúng trừ phi chúng có cùng các biến phía bên tay trái của mô hình. Nếu các biến phụ thuộc khác nhau, chúng ta có thể sử dụng các mô hình khác để dự đoán giá trị của cùng biến đó và kể đó tính hệ số tương quan của các giá trị tiên đoán và quan sát của biến này. Các hệ số tương quan này có thể được so sánh với nhau giữa các

mô hình. Tuy nhiên cần lưu ý rằng các dự báo về mức độ của biến độc lập được tạo ra từ các mô hình tuyến tính-lôgarít và lôgarít hai lần là thiên lệch và không nhất quán và cần có hệ số hiệu chỉnh.

Ba phương pháp thường được sử dụng nhất trong *kiểm định giả thuyết lồng vào nhau* - nghĩa là, trong các giả thuyết mà trong đó mô hình giới hạn là tập con của một mô hình không giới hạn tổng quát hơn. Đó là các *kiểm định Wald*, *kiểm định tỉ số thích hợp* (LR), và *kiểm định nhân tử Lagrange* (LM). Phương pháp Wald (còn được gọi là phương pháp lập mô hình từ “tổng quát đến đơn giản” Hendry/LSE) lập mô hình với nhiều biến độc lập và các độ trễ của chúng và kể đến sẽ hỏi liệu rằng có loại bớt một số biến không. Kiểm định LM liên quan đến việc lập mô hình cơ bản và tiếp theo là liệu có nên thêm biến nào khác nào mô hình không. Đây là phương pháp “từ đơn giản đến tổng quát”. Cả hai phương pháp sử dụng sự phán đoán và đều hữu dụng, tùy vào tình huống. Kiểm định LR xem hai mô hình tương đương.

Mặc dù một cách tiệm cận (nghĩa là, với cỡ mẫu lớn) ba kiểm định này tương đương, kiểm định LM hữu dụng trong các tình huống tổng quát hơn. Nó cũng hữu dụng trong việc kiểm định các tác động phi tuyến và sự tồn tại của các số hạng tương tác. Kiểm định LM được tiến hành theo ba bước: (1) hồi quy biến phụ thuộc theo một nhóm biến độc lập cơ bản, bao gồm cả số hạng hằng số; (2) xác định các phần dư từ thủ tục OLS được thực hiện ở Bước (1); và (3) hồi quy các phần dư theo tất cả các giá trị của X trong Bước (1), cũng như các biến mới (m về số lượng), mà có thể gồm các số hạng phi tuyến hoặc tích chéo (bình phương và tương tác) của các biến độc lập.

Nếu tích của cỡ mẫu (n) và R^2 không hiệu chỉnh từ phép hồi quy phụ này (nghĩa là, nR^2) lớn hơn $\chi^2_m(\alpha)$, điểm nằm trên phân phối Chi bình phương với m bậc tự do, về phía phải sao cho phần diện tích là α (mức ý nghĩa), thì *giả thuyết không* cho rằng tất cả m biến được thêm vào có hệ số bằng 0 bị bác bỏ. Nếu giả thuyết bị bác bỏ, trị t -values trong Bước (3) sẽ giúp xác định các biến có thể được thêm vào mô hình cơ bản. Ngay cả nếu kiểm định nR^2 không bác bỏ được *giả thuyết không* về các hệ số bằng 0, trị thống kê t của phép hồi quy phụ có thể gợi ý rằng một số biến nên được thêm vào. Sau đó các biến này có thể được thêm vào mô hình cơ bản để thực hiện các tập ước lượng mới. Trong các chương sau chúng ta sẽ thấy rằng các nguyên tắc của thủ tục kiểm định LM có thể áp dụng được trong các trường hợp tổng quát hơn.

Kiểm định sai số đặc trưng hồi quy Ramsey (RESET) cũng có thể được sử dụng để kiểm định đặc trưng của mô hình. Đầu tiên mô hình được ước lượng và các trị ước lượng của Y ($\hat{Y}$) được lưu lại. Các biến $\hat{Y}_1^2$, $\hat{Y}_1^3$, và $\hat{Y}_1^4$ được thêm vào mô hình và kiểm định F kết hợp được thực hiện cho các hệ số. Nếu các hệ số là có ý nghĩa kết hợp, đây sẽ là dấu hiệu của đặc trưng sai mô hình. Tuy nhiên, thủ tục này không xác định bản chất của đặc trưng sai. Dù vậy, phương pháp RESET có thể là một phương pháp bổ sung hữu ích cho kiểm định Wald, LM và LR về đặc trưng mô hình.

Thuật ngữ

Auxiliary Regression

Base

Box – Cox transformation

Cobb – Douglas production function

Constant returns to scale

Data – based simplification

Hồi quy phụ

Cơ sở

Phép biến đổi Box – Cox

Hàm sản xuất Cobb – Douglas

Lợi nhuận không đổi theo quy mô

Đơn giản hóa dựa trên dữ liệu

Data generation process
Decreasing returns to scale
Derivative
Double-log model
Dynamic model
Elasticity
Elasticity of output with respect to capital
Elasticity of output with respect to labor
Exponent
Exponential function
General to simple approach
Hendry/LSE approach
Increasing returns to scale
Instantaneous rate of growth
Interaction terms
Lagrange multiplier (LM) test
Lags in behavior
Likelihood ratio (LR) test
Linear-log model
Logarithmic function
Logistic curve
Logit model
Log-linear model
LSE approach
Marginal effect
Natural logarithm
Nested hypothesis
Nonlinearity in parameters
Nonnested hypothesis
Overparametrized
Parsimonious specification
Polynomial curve-fitting
Reciprocal transformation
Regression specification error test (RESET)
Semilog model
Simple to general approach
Spurious nonlinearity
Static model
Trend-fitting
Unitary elastic

Quá trình phát dữ liệu
Lợi nhuận giảm dần theo quy mô
Đạo hàm
Mô hình lôgarít hai lần
Mô hình động
Độ co giãn
Độ co giãn của sản lượng theo vốn
Độ co giãn của sản lượng theo lao động
Số mũ e
Hàm số mũ
Phương pháp từ tổng quát đến đơn giản
Phương pháp Hendry/LSE
Lợi nhuận tăng dần theo quy mô
Tỉ lệ tăng trưởng tức thời
Số hạng tương tác
Kiểm định nhân tử Lagrange (LM)
Độ trễ về hành vi
Kiểm định tỉ số thích hợp
Mô hình lôgarít-tuyến tính
Hàm lôgarít
Đường cong Logistic
Mô hình Logit
Mô hình tuyến tính-lôgarít
Phương pháp LSE
Tác động cận biên
Lôgarít cơ số e
Giả thuyết lồng vào nhau
Tính phi tuyến của các thông số
Giả thuyết không lồng vào nhau
Quá nhiều thông số
Đặc trưng xúc tích
Thích hợp bằng đường cong đa thức
Phép biến đổi nghịch đảo
Kiểm định sai số đặc trưng hồi quy
Mô hình bán lôgarít
Phương pháp từ đơn giản đến tổng quát
Phi tuyến giả tạo
Mô hình tĩnh
Thích hợp đường xu hướng
Co giãn đơn vị

6.A PHỤ LỤC CHI TIẾT VỀ CÁC KIỂM ĐỊNH LR, WALD VÀ LM

Phụ lục này cung cấp các chi tiết lý thuyết về kiểm định Wald, tỉ số thích hợp và nhân tử Lagrange. Tuy nhiên, trước khi xem phần này bạn nên đọc phần 2.A.3 về nguyên tắc thích hợp cực đại và phần 3.A.5 về ứng dụng của nó đối với mô hình hồi quy tuyến tính đơn. Mặc dù ba kiểm định này có thể ứng dụng trong nhiều trường hợp nhưng ở đây chúng ta vẫn nên tập trung vào các vấn đề hồi quy, đặc biệt là mô hình sau đây:

$$y_t = \beta x_t + u_t \quad (6.A.1)$$

Những chữ viết thường thể hiện các độ lệch của các biến so với giá trị trung bình tương ứng. Như đã trình bày ở phần 4.A.1, lợi ích của cách tiếp cận này là loại bỏ hằng số. Theo sự giả định này, các giá trị u tuân theo phân phối chuẩn có trung bình là 0, phương sai σ^2 , logarit của hàm

thích hợp đối với tập hợp các quan sát $y_1, y_2, \dots, y_n$ và tham số β chưa biết được viết như sau (quá trình này tương tự như quá trình trong phần 3.A.5).

$$\ln L = -n \ln \sigma - n \ln(\sqrt{2\pi}) - \frac{\sum (y_t - \beta x_t)^2}{2\sigma^2} \quad (6.A.2)$$

Giả thuyết không mà chúng ta đang xem xét có dạng $\beta = \beta_0$ và giả thuyết ngược lại $\beta \neq \beta_0$. Khi $\beta_0 = 0$, điều này tương đương với câu hỏi liệu biến số x có thuộc về mô hình không. Mỗi thủ tục kiểm định được thảo luận riêng rẽ, và sau đó thực hiện so sánh các phương pháp về mặt hình học. Xem lại chứng minh của Buse (1982) và Engle (1982) cũng như Ramanathan (1993) để biết chi tiết hơn về ba kiểm định này.

6.A.1 Kiểm định Tỉ số Thích Hợp

Trong thống kê, thủ tục kiểm định cổ điển dựa trên *tỉ số thích hợp*, mà theo những cụm từ đơn giản, nó được định nghĩa như tỉ số của giá trị lớn nhất của hàm thích hợp với *giả thuyết không* bị chia bởi giá trị lớn nhất của nó khi không bị giới hạn. Đặc biệt hơn, cho $\hat{\beta}$ là ước lượng thích hợp cực đại của tham số. Hàm thích hợp được đánh giá từ những giá trị này được diễn đạt bởi $L(\hat{\beta})$, bỏ qua σ^2 . Hãy đặt hàm thích hợp theo giả thuyết $\beta = \beta_0$ là $L(\beta_0)$. Tỉ số thích hợp được xác định như sau:

$$\lambda = \frac{L(\beta_0)}{L(\hat{\beta})}$$

Bởi vì mẫu số dựa trên mô hình không giới hạn, nên giá trị của nó không thể nhỏ hơn giá trị của tử số. Vì thế, $0 \leq \lambda \leq 1$. Nếu giả thuyết này đúng, bằng trực giác chúng ta kỳ vọng λ gần bằng 1. Nếu λ cách xa 1 thì LR theo *giả thuyết không* khác với LR theo mô hình không giới hạn, đó là giả thuyết ngược lại. Điều này cho thấy rằng chúng ta nên bác bỏ *giả thuyết không* nếu λ quá nhỏ. Kiểm định LR được thành lập như là một kiểm định bác bỏ *giả thuyết không* nếu $\lambda \leq K$, với K được xác định bởi điều kiện, theo *giả thuyết không*, $0 \leq \lambda \leq K$ tương đương với mức ý nghĩa (α); nghĩa là, $P(0 \leq \lambda \leq K | \beta = \beta_0) = \alpha$.

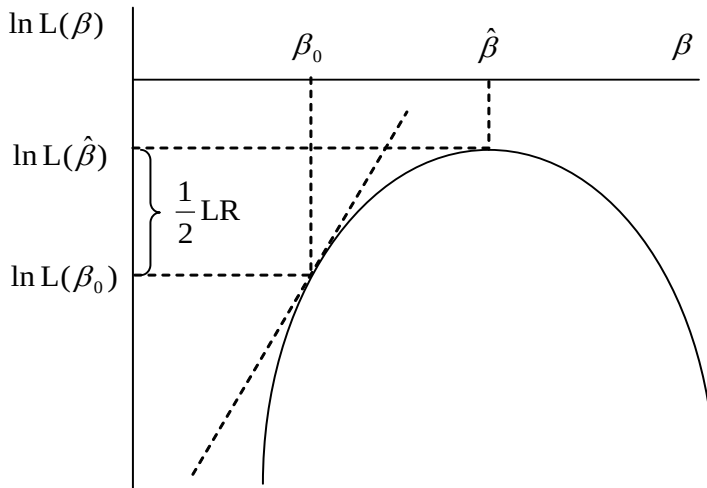
Trong một số trường hợp, vùng tới hạn $\lambda \leq K$ có thể chuyển sang một hình thức khác liên quan đến thống kê mẫu phổ biến như là thống kê t hay F . Trong những tình huống này, kiểm định LR giảm xuống thành kiểm định t -, F -, hay χ^2 . Ví dụ cho những trường hợp này, người đọc tham khảo Mood, Graybill và Boes (1974) và cả Ramanathan (1993), Chương 9. Những kiểm định khác trình bày ở Chương 2 có thể xuất phát từ nguyên tắc tỉ số thích hợp. Khi λ không thể chuyển sang dạng thống kê khác có phân bố phổ biến, thì phép thử tiến hành trên một lượng mẫu lớn thường được sử dụng. Điều đó có thể chỉ ra rằng (xem Ramanathan. 1993, trang 228), đối với kích cỡ mẫu lớn, thống kê

$$LR = -2 \ln \lambda = 2 \ln L(\hat{\beta}) - 2 \ln L(\beta_0) \quad (6.A.3)$$

có phân bố chi-square với bậc tự do tương đương với số giới hạn, bậc tự do bằng 1 như trong ví dụ của chúng ta. Ý tưởng đằng sau kiểm định này có thể được trình bày một cách hình học. Ở hình 6.A.1, logarit của hàm thích hợp được vẽ khi chỉ có duy nhất một tham số trong mô hình.

Hình vẽ nằm bên dưới trực bởi vì log của hàm thích hợp (nó là một mật độ phân bố nhỏ hơn 1) là số âm. Điểm $\hat{\beta}$ tương ứng với trường hợp khi hàm thích hợp đạt giá trị cực đại và β_0 tương ứng với *giả thuyết không*. Kiểm định LR dựa trên hiệu số về tung độ, chính là bằng một nửa LR. Nếu khoảng cách theo tung độ lớn, *giả thuyết không* bị bác bỏ.

● **Hình 6.A.1 Biểu diễn hình học của các kiểm định Wald, LR, và LM**



● **VÍ DỤ 6.A.1**

Nguyên tắc kiểm định tỉ lệ thích hợp được minh họa cho kiểm định giả thuyết $\beta = 0$ trong phương trình (6.A.1). Bằng cách tiến hành như trong Phần 3.A.5 và sử dụng chú thích trong Phần 3.2, chúng ta lưu ý rằng hàm thích hợp không giới hạn chỉ đạt cực đại khi $\hat{\beta} = S_{xy}/S_{xx}$ và $\hat{\sigma}^2 = \sum \hat{u}_t^2/n = ESS/n$, trong đó ESS là tổng bình phương sai số. Giá trị cực đại tương ứng là

$$\hat{L} = \left(\frac{1}{\hat{\sigma}\sqrt{2\pi}} \right)^n \exp \left[- \sum \hat{u}_t^2 / (2\hat{\sigma}^2) \right] = \left(\frac{1}{\hat{\sigma}\sqrt{2\pi}} \right)^n e^{-n/2}$$

Theo *giả thuyết không* $\beta = 0$, mô hình trở thành $y_t = u_t$ và hàm thích hợp trở thành

$$L(\sigma^2) = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp \left[- \sum u_t^2 / (2\sigma^2) \right] = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp \left[- \sum y_t^2 / (2\sigma^2) \right]$$

Hàm này cực đại khi $\tilde{\sigma}^2 = \sum y_t^2/n = TSS/n$, trong đó TSS là tổng bình phương. Do vậy, giá trị cực đại theo *giả thuyết không* được cho bởi

$$\tilde{L} = \left(\frac{1}{\tilde{\sigma}\sqrt{2\pi}} \right)^n e^{-n/2}$$

Tỉ số thích hợp là $\lambda = \tilde{L}/\hat{L} = (\hat{\sigma}/\tilde{\sigma})^n = (\hat{\sigma}^2/\tilde{\sigma}^2)^{n/2}$. Trị thống kê kiểm định LR là

$$LR = -2 \ln \lambda = -n \ln (\hat{\sigma}^2 / \tilde{\sigma}^2) = -n \ln (ESS/TSS) = -n \ln (1 - R^2)$$

trong đó R^2 là R^2 chưa hiệu chỉnh của mô hình không giới hạn.

Đối với những mẫu lớn, LR có phân phối chi-square với bậc tự do là 1. Chúng ta sẽ bác bỏ giả thuyết không với $\beta = 0$ nếu $LR > K$, trong đó K là điểm trên χ^2 mà vùng bên phải của K là mức ý nghĩa.

6.A.2 KIỂM ĐỊNH WALD

Không giống như kiểm định LR, sử dụng hiệu số tung độ (xem hình 6.A.1), kiểm định Wald sử dụng phép đo bình phương hiệu số hoành độ. Đặc biệt, hiệu số hoành độ bình phương $(\beta - \beta_0)^2$, được gán trọng số bởi hàm dạng $I(\hat{\beta})$, được sử dụng:

$$W = (\beta - \beta_0)^2 I(\beta) \quad (6.A.4)$$

trong đó

$$I(\beta) = -E \left[\frac{\partial^2 \ln L}{\partial \beta^2} \right] \quad (6.A.5)$$

là giá trị kỳ vọng của đạo hàm bậc hai của hàm thích hợp-logarit theo β . Đó là một phép đo độ cong của hàm thích hợp-logarit. Hàm I được biết như là *ma trận thông tin*. Thủ tục tính toán đối với kiểm định này có thể được tiến hành bằng cách ước lượng mô hình giới hạn và mô hình không giới hạn, như đã được thực hiện trong Chương 4, và bằng cách xây dựng một trị thống kê F. Việc chứng minh có bài bản cho kiểm định này đòi hỏi về đại số tuyến tính. (xem Ramanathan, 1993, trang 273-275).

○ VÍ DỤ 6.A.2

Trong trường hợp hồi quy đơn, lưu ý rằng $\hat{\beta}$ tuân theo phân bố $N(\beta_0, \sigma^2/S_{xx})$. Vì thế, $z = (\hat{\beta} - \beta_0)/(\sigma/\sqrt{S_{xx}})$ tuân theo phân phối chuẩn, và vì thế z^2 là chi-square với bậc tự do bằng 1. Vì vậy, thống kê kiểm định Wald tương ứng với giả thuyết không $\beta = 0$ được cho bởi $W = \hat{\beta}^2 S_{xx} / \hat{\sigma}^2$. Từ phương trình (3.12) chúng ta có $\hat{\beta} S_{xx} = S_{xy}$. Chúng ta cũng đã tìm ra $\hat{\sigma}^2 = RSS/n$, tổng bình phương hồi quy, trong Phần 3.A.1. Sử dụng hai kết quả này, chúng ta có

$$W = \frac{\hat{\beta}(\hat{\beta} S_{xx})}{ESS/n} = \frac{nRSS}{ESS} = \frac{nR^2}{1 - R^2}$$

Như trong trường hợp kiểm định LR, hàm này có phân phối chi-square đối với mẫu lớn. Giả thuyết không sẽ bị bác bỏ nếu W vượt quá giá trị tới hạn K được rút ra trong Ví dụ 6.A.1.

6.A.3 KIỂM ĐỊNH NHÂN TỬ LAGRANGE

Kiểm định LM trong Chương 2 dựa trên kỹ thuật nhân tử Lagrange để tối ưu hóa các ràng buộc. Mô hình giới hạn có được bằng cách áp đặt điều kiện β bằng với β_0 . Điều này gợi ý rằng chúng ta tối đa hóa logarit của hàm thích hợp theo β và σ^2 , với ràng buộc $\beta = \beta_0$. Như chúng ta đã thấy

ở Phần 2.A.2, điều này tương đương với cực đại hóa $\ln L(\beta) - \mu(\beta - \beta_0)$, trong đó μ là nhân tử Lagrange. Điều kiện đạo hàm bậc nhất cho việc cực đại hóa là

$$\frac{\partial \ln L}{\partial \beta} = \mu$$

Nếu giả thuyết không $\beta = \beta_0$ là đúng, ước lượng thích hợp cực đại giới hạn sẽ gần với giá trị ước lượng không giới hạn. Chúng ta cũng lưu ý rằng nếu nhân tử Lagrange, μ , là 0, thì phương trình sẽ cho giá trị ước lượng thích hợp cực đại. Do đó, nhân tử Lagrange có thể được diễn giải như là “giá mờ” của ràng buộc $\beta = \beta_0$. Nếu giá cao, ràng buộc nên bị bác bỏ vì không nhất quán với số liệu. Điều này chính là động cơ dẫn đến kiểm định LM. Kiểm định LM dựa trên đạo hàm riêng phần $(\beta \ln L)/\partial \beta$, được biết đến như *hàm giá trị điểm* và được mô tả bởi S. Engle (1982), có được từ thống kê kiểm định cho mô hình hồi quy bội và cho thấy rằng kiểm định có thể được thực hiện bằng cách chạy hồi quy phụ trên các phần dư ước lượng của mô hình giới hạn (cũng có thể xem Ramanathan, 1993, trang 276-277). Các bước thực hiện được trình bày trong Phần 6.14. Thống kê kiểm định LM có dạng

$$LM = S^2(\beta_0) I(\beta_0)^{-1} \quad (6.A.6)$$

Trong hình 6.A.1, hàm giá trị điểm, đạo hàm riêng phần của hàm thích hợp-logarit, chính là độ dốc của của đồ thị tại β_0 . Giả thuyết ngược lại tương ứng với $S(\beta) = 0$: có nghĩa là độ dốc gần tới 0. Vì thế, kiểm định Wald dựa trên hiệu số hoành độ giữa $\hat{\beta}$ và β_0 trong đồ thị, kiểm định LR dựa trên hiệu số tung độ, và kiểm định LM dựa trên độ dốc của đường cong β_0 . Mỗi kiểm định là phép đo hợp lý về khoảng cách giữa giả thuyết không và giả thuyết ngược lại. Một cách độc lập nhau, Engle (1982) và Buse (1982) đã chỉ ra rằng khi hàm thích hợp-logarit là hàm bậc hai (như phương trình ở Phần 6.A.2), thì tất cả ba thủ tục kiểm định này đều cho kết quả như nhau. Đối với một mô hình tuyến tính tổng quát, có sự bất cân xứng về ràng buộc giữa ba tiêu chuẩn kiểm định. Điều này được thể hiện như sau:

$$W \geq LR \geq LM$$

Điều đó có nghĩa là bất cứ khi nào kiểm định LM bác bỏ giả thuyết không với các hệ số zero, thì các kiểm định khác cũng vậy. Tương tự, bất cứ khi nào kiểm định Wald không bác bỏ giả thuyết không thì các kiểm định khác cũng vậy. Nói một cách máy móc, kiểm định LR thì rườm rà nhất, trừ khi chuyển đổi sang kiểm định t , F , hay kiểm định χ^2 . Hai cách kiểm định khác đơn giản hơn, như đã thể hiện trong tài liệu.

○ VÍ DỤ 6.A.3

Trong trường hợp hồi quy đơn, hàm giá trị điểm được cho bởi

$$S = \frac{\partial \ln L}{\partial \beta} = \frac{\sum (y_t - \beta x_t) x_t}{\sigma^2} = \frac{\sum x_t u_t}{\sigma^2}$$

và phương sai của nó là $\sum x_t^2 / \sigma^2 = S_{xx} / \sigma^2$. Do đó,

$$z^2 = \frac{S^2}{\text{Var}(S)} = \frac{(\sum x_t u_t)^2}{\sigma^2 S_{xx}} \sim \chi_1^2$$

Do đó, trị thống kê kiểm định LM được cho bởi

$$LM = \frac{(\sum x_t \hat{u}_t)^2}{\tilde{\sigma}^2 S_{xx}}$$

Hãy xem xét hồi quy phụ $u_t = \gamma x_t + v_t$. Làm theo các bước giống như ví dụ 6.A.1, dễ dàng có được các phương trình sau:

$$\hat{\gamma} = \frac{\sum x_t \hat{u}_t}{S_{xx}}, \quad RSS_{aux} = \hat{\gamma} \sum x_t \hat{u}_t, \quad \tilde{\sigma}^2 = \frac{\sum \hat{u}_t^2}{n}$$

Thay thế các biểu thức này vào trị thống kê kiểm định LM, chúng ta có

$$LM = n[RSS_{aux} / (\sum \hat{u}_t^2)] = n[RSS_{aux} / TSS_{aux}] = nR_{aux}^2$$

Điều này tạo ra kết quả đã được cho trong tài liệu rằng trị thống kê kiểm định LM bằng số quan sát nhân với R^2 chưa hiệu chỉnh của hồi quy phụ. Mặc dù chứng minh này chỉ cho trường hợp hồi quy đơn được xem xét ở đây, nhưng nó cũng đúng cho mô hình hồi quy bội tổng quát. Muốn biết thêm chi tiết, hãy xem Ramanathan (1993), trang 276-278.