Chương 7

PHÂN TÍCH HỒI QUY BỘI: VẤN ĐỀ VỀ ƯỚC LƯỢNG

Mô hình hai biến mà chúng ta đã nghiên cứu chi tiết trong những chương trước trên thực tế thường là không thỏa đáng. Chẳng hạn như, trong ví dụ của chúng ta về thu nhập-chi tiêu, chúng ta giả định ngầm rằng chỉ có thu nhập X ảnh hưởng đến chi tiêu Y. Nhưng lý thuyết kinh tế ít khi được đơn giản như vậy, bởi vì ngoài chi tiêu ra, một số những biến khác cũng có thể có ảnh hưởng đến chi tiêu tiêu dùng. Đơn cử một ví dụ dễ thấy là sự giàu có của người tiêu thụ. Một ví dụ khác, nhu cầu về một mặt hàng thường không chỉ phụ thuộc vào giá của nó mà thôi, mà còn phụ thuộc vào giá cả của những hàng hóa cạnh tranh hay bổ trợ khác, phụ thuộc vào thu nhập của người tiêu dùng, địa vị xã hội, v.v. Vì vậy, chúng ta cần phải mở rộng mô hình hồi qui hai biến đơn giản của chúng ta để xem xét đến những mô hình gồm có nhiều hơn hai biến. Việc đưa thêm nhiều biến vào dẫn tới việc thảo luận các mô hình hồi qui bội, tức những mô hình trong đó biến phụ thuộc, hay biến hồi qui phụ thuộc độc lập, Y phụ thuộc vào hai hay nhiều biến giải thích, hay biến hồi quy độc lập trở lên.

Mô hình hồi qui bội đơn giản nhất có thể có là hồi qui ba biến, với một biến độc lập và hai biến giải thích. Trong chương này và chương tiếp theo chúng ta sẽ nghiên cứu mô hình này, và trong Chương 9 chúng ta sẽ khái quát hóa để áp dụng nó vào những trường hợp nhiều hơn ba biến. Xuyên suốt tập sách, chúng ta quan tâm đến mô hình hồi qui tuyến tính bội, có nghĩa là, những mô hình tuyến tính theo thông số; chúng có thể là hoặc có thể không phải là tuyến tính theo các biến số.

7.1 MÔ HÌNH BA BIẾN : KÝ HIỆU VÀ CÁC GIẢ ĐỊNH

Khái quát hóa hàm hồi qui tổng thể (PFR) hai biến (2.4.2), chúng ta có thể viết PRF ba biến như sau:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \tag{7.1.1}$$

trong đó Y là biến phụ thuộc, X_2 và X_3 là các biến giải thích (hay biến hồi qui độc lập), u là số hạng nhiễu ngẫu nhiên, và i là quan sát thứ i; trong trường hợp dữ liệu là chuỗi thời gian, chỉ số dưới t sẽ biểu thị quan sát thứ t.[1]

Trong Phương trình (7.1.1) β_1 là số hạng tung độ gốc. Như thường lệ, nó cho biết ảnh hưởng trung bình của tất cả các biến bị loại ra khỏi mô hình đối với Y, mặc dù giải thích nó một cách máy móc là giá trị trung bình của Y khi X_2 và X_3 được lấy bằng zero. Hệ số β_2 và β_3 được gọi là **hệ số hồi qui riêng phần**, và ý nghĩa của nó sẽ được giải thích ở tiếp dưới.

Chúng ta tiếp tục hoạt động trong khuôn khổ mô hình hồi qui tuyến tính cổ điển (CRLM) được giới thiệu trong Chương 3. Đặc biệt, chúng ta giả định như sau:

[1] Để cho cân xứng về mặt ký hiệu, Pt. (7.1.1) cũng có thể được viết thành

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

với điều kiện là $X_{1i} = 1$ đối với mọi i.

Giá trị trung bình của u_i là 0 hay

$$E(u_i \mid X_{2i}, X_{3i}) = 0 \qquad \text{cho mỗi } i \tag{7.1.2}$$

Không có tương quan chuỗi, hay

$$\text{cov}(u_i, u_j) = 0 \qquad i \neq j \tag{7.1.3}$$

Phương sai có điều kiện không đổi, hay

$$\text{var}(u_i) = \sigma^2 \tag{7.1.4}$$

Tích sai giữa u_i và mỗi biến X có giá trị bằng 0 hay

$$\text{cov}(u_i, X_{2j}) = \text{cov}(u_i, X_{3j}) = 0 \tag{7.1.5}$$ [2]

Không có thiên lệch đặc trưng, hay

Mô hình được xác định đúng (7.1.6)

Không có cộng tuyến rõ ràng giữa các biến X, hay

Không có **quan hệ tuyến tính rõ ràng** giữa X_2 và X_3 (7.1.7)

Thêm vào đó, cũng như ở Chương 3, chúng ta giả định rằng mô hình hồi qui bội là *tuyến tính theo các thông số*, rằng các giá trị của biến hồi qui độc lập là được giữ cố định trong những lần lấy mẫu liên tiếp, và rằng có đủ sự biến đổi về các giá trị của các biến hồi qui độc lập.

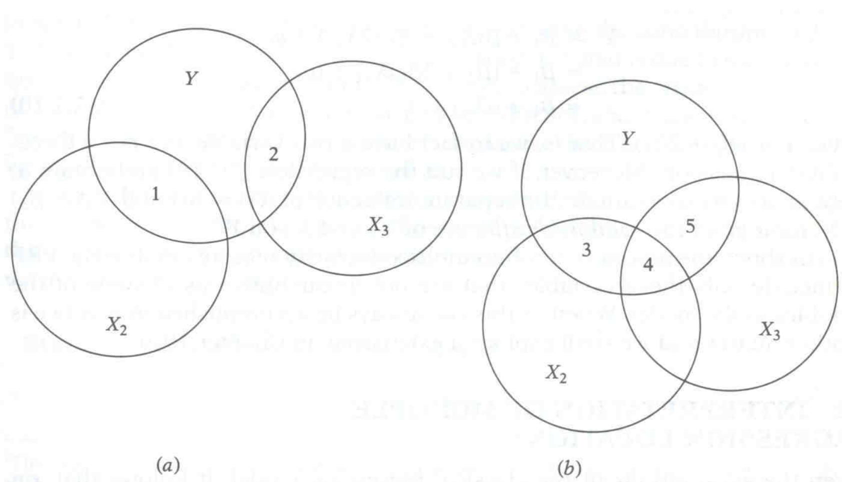
Cơ sở cho những giả định từ (7.1.2) cho đến (7.1.6) cũng tương tự như ta đã thảo luận trong Phần 3.2. Giả định (7.1.7), rằng không có quan hệ tuyến tính rõ ràng giữa X_2 và X_3, được gọi là giả định về sự *phi cộng tuyến*, hay **phi đa cộng tuyến** nếu có nhiều hơn một quan hệ tuyến tính rõ ràng có liên quan, là giả định mới và cần phải được giải thích.[3]

Nói một cách đơn giản, phi cộng tuyến có nghĩa là không có biến giải thích nào có thể được biểu thị dưới dạng tổ hợp tuyến tính với những biến giải thích còn lại. Ý nghĩa của điều này có thể thấy được từ biểu đồ Venn, hay Ballentine, đã được giới thiệu trong Chương 3. Trong hình này, vòng tròn Y tượng trưng cho sự biến đổi của biến phụ thuộc Y và các vòng tròn X_2 và X_3 lần lượt biểu thị cho sự biến đổi của biến hồi qui độc lập X_2 và X_3. Trong hình 7.1*a* vùng 1 biểu thị sự biến đổi của Y do X_2 giải thích (thông qua một hồi qui OLS) và vùng 2 biểu thị sự biến đổi của Y do X_3 giải thích. Trong hình 7.1*b*, vùng 3 và 4 biểu thị sự biến đổi của Y do X_2 giải thích và vùng 4 và 5 biểu thị sự của Y do X_3 giải thích. Nhưng bởi vì vùng 4 là vùng chung cho cả X_2 và X_3, một tiên nghiệm mà chúng ta không biết phần nào trong 4 thuộc về X_2 và phần nào thuộc về X_3. Vùng chung 4 tượng trưng cho trạng thái cộng tuyến. Giả định về tính không cộng tuyến đòi hỏi rằng không được có một sự trùng lặp nào giữa X_2 và X_3, có nghĩa là vùng chung 4 phải bằng không. Nói một cách một cách khác, điều kiện chúng ta cần là tương tự như tình huống được mô tả trong hình 7.1*a*.

---

[2] Giả định này tự động được thực hiện nếu X_2 và X_3 là không ngẫu nhiên và (7.1.2) là đúng.

[3] Trong mô hình hai biến chúng ta không cần giả định này. Tại sao?



HÌNH 7.1
Biểu đồ Ballentine, trình bày phi cộng tuyến (a) và cộng tuyến (b)

Phát biểu bằng thuật ngữ chuyên môn, không cộng tuyến có nghĩa là không hề tồn tại một tập hợp các số λ_2 và λ_3, không phải cả hai đều bằng không, sao cho

$$\lambda_2 X_{2i} + \lambda_3 X_{3i} = 0 \qquad (7.1.8)$$

Nếu tồn tại một quan hệ tuyến tính như vậy, khi đó X_2 và X_3 được coi là **cộng tuyến** hay **phụ thuộc tuyến tính**. Mặt khác, nếu (7.1.8) chỉ đúng khi $\lambda_2 = \lambda_3 = 0$, thì X_2 và X_3 được coi là *độc lập tuyến tính.*

Như vậy, nếu

$$X_{2i} = -4X_{3i} \qquad \text{hay } X_{2i} + 4X_{3i} = 0 \qquad (7.1.9)$$

hai biến này là phụ thuộc tuyến tính, và nếu cả hai đều được đưa vào trong một mô hình hồi qui chúng ta sẽ có cộng tuyến hoàn hảo hay một quan hệ tuyến tính rõ ràng giữa hai biến hồi qui độc lập.

Nhưng giả sử $X_{3i} = X_{2i}^2$. Điều này có vi phạm giả thiết không cộng tuyến hay không? Không, bởi vì quan hệ giữa hai biến ở đây là *không tuyến tính* và không hề phá vỡ yêu cầu là không được có *quan hệ tuyến tính rõ ràng* giữa hai biến hồi qui độc lập. Tuy nhiên, cần phải lưu ý rằng trong trường hợp này r^2 và r được tính theo quy ước sẽ cao, đặc biệt trong các mẫu của X_2 và X_3 có một ít các giá trị cực trị. Nhưng vấn đề này sẽ được nói tới nhiều hơn ở Chương 10.

Mặc dù chúng ta sẽ xem xét vấn đề đa cộng tuyến một cách chi tiết ở Chương 10, về mặt trực giác tính lôgíc của giả thiết phi đa cộng tuyến không phải là quá khó để không hiểu được. Giả sử rằng trong (7.1.1) *Y, X_2, và X_3* lần lượt biểu thị cho chi tiêu tiêu dùng, thu nhập và sự giàu có của người tiêu thụ. Khi quy định rằng chi tiêu tiêu dùng là có quan hệ tuyến tính với thu nhập và sự giàu có, lý thuyết kinh tế cho rằng sự giàu có và thu nhập có thể có một vài ảnh hưởng độc lập đối với tiêu dùng. Nếu không, không có lý do gì để đưa cả biến thu nhập và tiêu dùng vào trong mô hình. Trong trường hợp quá mức đặc biệt, nếu có quan hệ tuyến tính rõ ràng giữa thu nhập và sự giàu có, chúng ta chỉ có một biến độc lập, chớ không phải hai, và không có cách nào để đánh giá được từng ảnh hưởng *riêng* của thu nhập và sự giàu có đối với tiêu dùng. Để thấy được điều này một cách rõ ràng, cho $X_{3i} = 2X_{2i}$ trong hồi qui chi tiêu-thu nhập-sự giàu có. Khi đó hồi qui (7.1.1) trở thành

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + \beta_3(2X_{2i}) + u_i \\ &= \beta_1 + (\beta_2 + 2\beta_3)X_{2i} + u_i \\ &= \beta_1 + \alpha X_{2i} + u_i \end{aligned} \tag{7.1.10}$$

trong đó $\alpha = (\beta_2 + 2\beta_3)$. Có nghĩa là, sự thật chúng ta có một hồi qui hai biến thay vì hồi qui ba biến. Hơn nữa, nếu chúng ta thực hiện hồi qui (7.1.10) và thu được α, không có cách nào để ước lượng được các ảnh hưởng riêng biệt của X_2 (=β_2) và X_3 (=β_3) đối với Y, bởi vì α cho ta *ảnh hưởng kết hợp* của cả X_2 và X_3 trên Y.[4]

Tóm lại, giả định phi đa cộng tuyến đòi hỏi rằng trong hàm hồi qui tổng thể (PRF) chúng ta đưa vào chỉ những biến nào không phải là hàm tuyến tính của một số trong những biến trong mô hình. Liệu có thể luôn luôn đạt được điều này trên thực tế không lại là một vấn đề khác và chúng ta sẽ xem xét đến nó một cách bao quát trong Chương 10.

7.2 GIẢI THÍCH PHƯƠNG TRÌNH HỒI QUI BỘI

Với những giả định về mô hình hồi qui bội cổ điển, chúng ta suy ra, khi lấy kỳ vọng có điều kiện của Y ở cả hai vế của (7.1.1) chúng ta có

$$E(Y_i \mid X_{2i}, X_{3i}) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} \tag{7.2.1}$$

Diễn tả bằng ngôn ngữ, (7.2.1) cho biết **trung bình có điều kiện hay giá trị kỳ vọng của Y với điều kiện là đã biết các giá trị cố định hay đã cho của các biến X_2 và X_3.** Do đó, cũng tương tự như trong mô hình hai biến, phân tích hồi qui bội là phân tích hồi qui với điều kiện đã biết các giá trị cố định của các biến giải thích,và chúng ta thu được giá trị trung bình của Y hay trung bình tương ứng của Y đối với các giá trị cố định của các biến X.

7.3 Ý NGHĨA CỦA CÁC HỆ SỐ HỒI QUI RIÊNG PHẦN

Ý nghĩa của **hệ số hồi qui riêng phần** là như sau: β_2 đo lường sự thay đổi trong giá trị trung bình Y, $E(Y / X_2, X_3)$ khi X_2 thay đổi một đơn vị, *giữ X_3 không đổi*. Nói một cách khác, nó cho biết độ dốc của $E(Y / X_2, X_3)$ so với X_2, giữ X_3 không đổi.[5] Nói một cách khác, nó cho biết ảnh hưởng "trực tiếp" hay "ròng" của các thay đổi một đơn vị trong X_2 đối với giá trị trung bình của Y, loại trừ ảnh hưởng của X_3. Tương tự, β_3 đo lường thay đổi trong giá trị trung bình của Y khi X_3 thay đổi một đơn vị, giữ *X_2 không đổi*. Có nghĩa là, nó cho biết ảnh hưởng "trực tiếp" hay "ròng" của thay đổi một đơn vị trong X_3 đối với giá trị trung bình của Y, loại trừ ảnh hưởng của X_2.

Ý nghĩa chính xác của thuật ngữ *giữ không đổi* là gì?[6] Để hiểu được điều này, giả sử Y tượng trưng cho sản lượng và X_2 và X_3 tượng trưng cho lao động và vốn ở đầu vào. Giả sử thêm là cả X_2 và X_3 đều cần thiết đối với việc sản xuất Y và tỉ lệ chúng được sử

[4] Về mặt toán học, $\alpha = (\beta_2 + 2\beta_3)$ là một phương trình gồm có hai đại lượng chưa biết và không có cách độc nhất nào có thể ước tính được β_2 và β_3 từ α đã được ước lượng.

[5] Những bạn đọc có đầu óc về toán sẽ nhận thấy ngay là β_2 và β_3 là các đạo hàm riêng phần của $E(Y \mid X_2, X_3)$ tương ứng với X_2 và X_3.

[6] Các thuật ngữ *kiểm soát, giữ không đổi, lưu ý đến hay tính đến ảnh hưởng của,* và *hiệu chỉnh lại ảnh hưởng của* tất cả đều đồng nghĩa với nhau và sẽ được dùng thay thế lẫn nhau trong tài liệu này.

dụng để sản xuất Y là có thể thay đổi. Bây giờ, giả sử chúng ta tăng lao động ở đầu vào thêm một đơn vị, kết quả thu được là sản lượng gia tăng (tổng sản phẩm biên tế của lao động). Chúng ta có thể quy sự thay đổi sản lượng này chỉ là kết quả của lao động X_2 ở đầu vào mà thôi được không?[7] Nếu chúng ta làm như vậy, chúng ta đang *thổi phồng* sự đóng góp của X_2 đối với Y; X_2 có "công" đưa đến lượng thay đổi đó trong Y, mà đúng ra là nhờ vào sự gia tăng đồng thời của vốn ở đầu vào. Do đó, để đánh giá đóng góp "đúng" của X_2 đối với thay đổi trong Y (sản phẩm biên tế ròng của lao động), bằng cách nào đó chúng ta phải "kiểm soát" được ảnh hưởng của X_3. Tương tự như vậy, để đánh giá đóng góp "đúng" của X_3, chúng ta cũng phải kiểm soát ảnh hưởng của X_2.

Chúng ta tiến hành thực hiện quá trình kiểm soát này như thế nào? Nói một cách cụ thể, giả thiết rằng chúng ta muốn kiểm soát ảnh hưởng tuyến tính của vốn X_3 ở đầu vào khi đo ảnh hưởng của thay đổi một đơn vị của lao động X_2 ở đầu vào đối với sản lượng đầu ra. Để thực hiện điều này, chúng ta có thể tiến hành như sau:

Giai đoạn I: Hồi qui Y chỉ theo X_3 như sau:

$$Y_i = b_1 + b_{13}X_{3i} + u_{1i} \tag{7.3.1}$$

Phương trình (7.3.1) chỉ là hồi qui hai biến nếu không có ký hiệu mới, tự giải thích, trong đó u_i là số hạng phần dư (mẫu) (Lưu ý: Trong $b_{1\,3}$ chỉ số dưới 1 tượng trưng cho biến *Y*.)

Giai đoạn II: Hồi qui X_2 chỉ theo X_3 như sau:

$$X_{2i} = b_2 + b_{23}X_{3i} + u_{2i} \tag{7.3.2}$$

trong đó u_{2i} cũng là số hạng phần dư. Bây giờ

$$u_{1i} = Y_i - b_1 - b_{13}X_{3i} \tag{7.3.3}$$

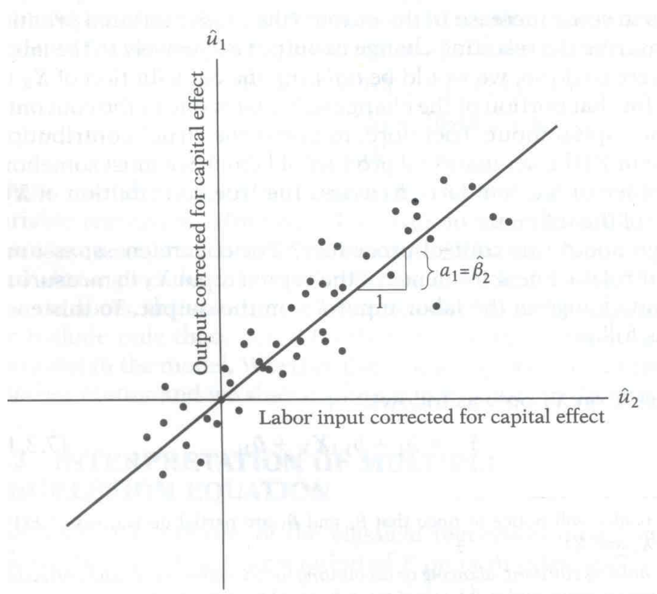
và

$$\begin{aligned} u_{2i} &= X_{2i} - b_2 - b_{23}X_{3i} \\ &= X_{2i} - X_{2I} \end{aligned} \tag{7.3.4}$$

trong đó Y_i và X_{2i} là những giá trị được ước lượng từ hồi qui (7.3.1) và (7.3.2).

Các phần dư u_{1i} và u_{2i} có ý nghĩa gì? Thuật ngữ u_{1i} tiêu biểu cho giá trị của Y_i sau khi loại bỏ ảnh hưởng (tuyến tính) của X_3 đối với nó, và tương tự u_{2i} biểu thị cho giá trị của X_{2i} sau khi đã loại bỏ ảnh hưởng (tuyến tính) của X_3 đối với nó. Vì vậy, có thể nói u_{1i} và u_{2i} là Y_i và X_{2i} "tinh khiết", có nghĩa là, đã được gạt bỏ ảnh hưởng (ô nhiễm) của X_3.

[7] Bởi vì trong sản xuất phải cần đến cả lao động và vốn, sự gia tăng này có thể dẫn đến sự gia tăng của vốn; lượng thay đổi của vốn sẽ phụ thuộc vào công nghệ sản xuất.



HÌNH 7.2
Đồ thị phân tán giữa sản lượng và lao động ở đầu vào được hiệu chỉnh cho ảnh hưởng tuyến tính của vốn

Giai đoạn III: Vì vậy, giờ đây nếu chúng ta tiếp tục lấy hồi qui của u_{1i} theo u_{2i} như sau,

$$u_{1i} = a_0 + a_1 u_{2i} + u_{3i} \quad (7.3.5)$$

trong đó u_{3i} cũng là số hạng phần dư mẫu. Vậy thì, a_{21} sẽ cho chúng ta một ước lượng của ảnh hưởng "thực" hay ròng của thay đổi một đơn vị trong X_2 đối với Y (có nghĩa là sản phẩm biên tế ròng của lao động) hay độ dốc thực của Y so với X_2, có nghĩa là, một ước lượng của β_2. Và sự thực là đúng như vậy, như chúng ta thấy ở phần Phụ lục 7A, Phần 7A.2. (Đồng thời xem bài tập 7.5.)

Về mặt hình học, chúng ta có Hình 7.2. Tuy nhiên, trên thực tế, không cần phải đi qua quá trình chậm chạp và tốn thời gian này, bởi vì a_1 có thể được ước lượng trực tiếp từ các công thức đã cho trong Phần 7.4 [xem phương trình (7.4.7)]. Quá trình ba giai đoạn đã phác thảo ở trên đơn thuần chỉ là một công cụ sư phạm để giúp bạn đọc tiếp thu được ý nghĩa của hệ số hồi qui riêng phần.

7.4 ƯỚC LƯỢNG BÌNH PHƯƠNG TỐI THIỂU THÔNG THƯỜNG (OLS) VÀ THÍCH HỢP TỐI ĐA (ML) CỦA CÁC HỆ SỐ HỒI QUI RIÊNG PHẦN

Để ước lượng các thông số của mô hình hồi qui ba biến (7.1.1), trước hết chúng ta xem xét phương pháp bình phương tối thiểu thông thường (OLS) đã giới thiệu trong Chương 3 và sau đó xem xét ngắn gọn phương pháp ước lượng thích hợp tối đa (ML) đã được bàn thảo trong Chương 4.

Các hàm ước lượng OLS

Để tìm các hàm ước lượng OLS, đầu tiên chúng ta viết hàm hồi qui mẫu (SRF) tương ứng với PRF của (7.1.1) như sau:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad (7.4.1)$$

trong đó u_i là số hạng phần dư, là số hạng tương ứng của mẫu với số hạng nhiễu ngẫu nhiên u_i.

Như đã lưu ý ở Chương 3, quá trình OLS bao gồm việc chọn các giá trị của các thông số chưa biết sao cho tổng các bình phương của phần dư (RSS) $\sum \hat{u}_i^2$ nhỏ nhất có thể được. Biểu diễn bằng ký hiệu toán học ta có,

$$\min \sum \hat{u}_i^2 = \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}\right)^2 \tag{7.4.2}$$

trong đó biểu thức thể hiện RSS có được bằng những phép tính đại số đơn giản từ (7.4.1).

Phương pháp đơn giản nhất để thu được các hàm ước lượng có khả năng sẽ tối thiểu hóa (7.4.2) là đạo hàm nó theo các đại lượng chưa biết, cho biểu thức thu được không, và giải các biểu thức này cùng một lúc. Như được trình bày ở Phụ lục 7A, Phần 7A.1, phương pháp này cho ta những *phương trình chuẩn* sau [so sánh với các phương trình (3.1.4) và (3.1.5)]:

$$\overline{Y} = \hat{\beta}_1 + \hat{\beta}_2 \overline{X}_2 + \hat{\beta}_3 \overline{X}_3 \tag{7.4.3}$$

$$\sum Y_i X_{2i} = \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i} X_{3i} \tag{7.4.4}$$

$$\sum Y_i X_{3i} = \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{2i} X_{3i} + \hat{\beta}_3 \sum X_{3i}^2 \tag{7.4.5}$$

Từ phương trình (7.4.3) chúng ta có thể thấy ngay lập tức rằng

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X}_2 - \hat{\beta}_3 \overline{X}_3 \tag{7.4.6}$$

chính là hàm ước lượng OLS của tung độ gốc tổng thể β_1.

Theo quy ước, gọi các mẫu tự viết thường (không viết dưới dạng chữ in) là biểu thị cho độ lệch so với các giá trị trung bình mẫu, chúng ta có thể rút ra được những công thức sau từ các phương trình chuẩn (7.4.3) và (7.4.5):

$$\hat{\beta}_2 = \frac{\sum y_i x_{2i}\left(\sum x_{3i}^2\right) - \left(\sum y_i x_{3i}\right)\left(\sum x_{2i} x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i} x_{3i}\right)^2} \tag{7.4.7}$$ [8]

$$\hat{\beta}_3 = \frac{\sum y_i x_{3i}\left(\sum x_{2i}^2\right) - \left(\sum y_i x_{2i}\right)\left(\sum x_{2i} x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i} x_{3i}\right)^2} \tag{7.4.8}$$

những phương trình này cho ta các hàm ước lượng OLS của các hệ số hồi qui riêng phần tương ứng β_2 và β_3 của tổng thể. Nhân đây lưu ý những điểm sau: (1) Các phương trình (7.4.7) và (7.4.8) bản chất là cân xứng bởi vì ta có thể thu được phương trình này từ phương trình kia bằng cách thay đổi vai trò của X_2 và X_3 cho nhau; (2) các mẫu số của hai phương trình này là giống nhau; và (3) trường hợp ba biến là sự mở rộng tự nhiên của trường hợp hai biến.**Các Phương Sai Và Sai Số Chuẩn Của Các Hàm Ước Lượng OLS**

[8]Hàm ước lượng này tương đương với a_1 trong (7.3.5), được trình bày ở Phụ lục 7A, Phần 7A.2.

Sau khi đã có được các hàm ước lượng OLS của các hệ số hồi qui riêng phần, chúng ta có thể tính được các phương sai và sai số chuẩn của các hàm ước lượng này bằng cách thức đã chỉ ra ở Phụ lục 3A.3. Tương tự như trong trường hợp hai biến, chúng ta cần có những sai số chuẩn vì hai mục đích chính: để thiết lập khoảng tin cậy và kiểm định các giả thiết thống kê. Các công thức có liên quan sẽ như sau:[9]

$$\operatorname{var}(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\bar{X}_2^2 \sum x_{3i}^2 + \bar{X}_3^2 \sum x_{2i}^2 - 2\bar{X}_2\bar{X}_3 \sum x_{2i}x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i}x_{3i}\right)^2}\right] \cdot \sigma^2 \qquad (7.4.9)$$

$$se(\hat{\beta}_1) = +\sqrt{\operatorname{var}(\hat{\beta}_1)} \qquad (7.4.10)$$

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sum x_{3i}^2}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i}x_{3i}\right)^2}\sigma^2 \qquad (7.4.11)$$

hoặc, một cách tương đương, ta có:

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2\left(1 - r_{23}^2\right)} \qquad (7.4.12)$$

trong đó $r_{2\,3}$ là hệ số tương quan giữa X_2 và X_3 của mẫu như đã được định nghĩa ở Chương 3.[10]

$$se(\hat{\beta}_2) = +\sqrt{\operatorname{var}(\hat{\beta}_2)} \qquad (7.4.13)$$

$$\operatorname{var}(\hat{\beta}_3) = \frac{\sum x_{2i}^2}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i}x_{3i}\right)^2}\sigma^2 \qquad (7.4.14)$$

hoặc, một cách tương đương, ta có:

$$\operatorname{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2\left(1 - r_{23}^2\right)} \qquad (7.4.15)$$

$$se(\hat{\beta}_3) = +\sqrt{\operatorname{var}(\hat{\beta}_3)} \qquad (7.4.16)$$

$$\operatorname{cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}\sigma^2}{\left(1 - r_{23}^2\right)\sqrt{x_{2i}^2}\sqrt{x_{3i}^2}} \qquad (7.4.17)$$

Trong tất cả các công thức này σ^2 là phương sai (phương sai có điều kiện không đổi) của các số hạng nhiễu tổng thể u_i.

Theo lập luận ở Phụ lục 3A.5, bạn đọc có thể chứng minh rằng một hàm ước lượng không thiên lệch của σ^2 là $\hat{\sigma}^2 = \dfrac{\sum \hat{u}_i^2}{n-3}$ (7.4.18)

[9] Việc chứng minh các công thức này dễ dàng hơn nếu sử dụng ký hiệu ma trận. Vì vậy, cách thức chứng minh được hoãn đến Chương 9.

[10] Sử dụng định nghĩa của r đã cho trong Chương 3, ta có

$$r_{23}^2 = \frac{\left(\sum x_{2i}x_{3i}\right)^2}{\sum x_{2i}^2 \sum x_{3i}^2}$$

Lưu ý rằng sự tương tự giữa hàm ước lượng σ^2 này và hàm ước lượng hai biến tương ứng với nó [σ^2 = $\left(\sum \hat{u}_1^2\right)/\left(n-2\right)$]. Các bậc tự do bây giờ là (n -3) bởi vì khi ước lượng $\left(\sum \hat{u}_1^2\right)$ trước hết chúng ta cần ước lượng β_1, β_2, và β_3, đã sử dụng 3 bậc tự do. (Lập luận này rất tổng quát. Như vậy, trong trường hợp bốn biến, bậc tự do sẽ là n - 4.)

Hàm ước lượng $\hat{\sigma}^2$ có thể được tính từ (7.4.18) một khi đã có sẵn các phần dư, nhưng cũng có thể có được nó một cách dễ dàng hơn bằng cách dùng mối quan hệ sau đây (xem Phụ lục 7A, Phần 7A.3 để biết bằng chứng):

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i} \tag{7.4.19}$$

đây chính là biểu thức trường hợp ba biến tương ứng với mối quan hệ trong phương trình (3.3.6).**Các đặc tính của Hàm ước lượng OLS**Các đặc tính của Hàm ước lượng OLS của mô hình hồi qui bội cũng tương tự với những đặc tính của mô hình hai biến. Cụ thể là: **1.** Đường (mặt phẳng) hồi qui ba biến đi ngang qua các trung bình $\bar{Y}, \bar{X}_2, \bar{X}_3$, đây là điều hiển nhiên chúng ta có thể thấy từ (7.4.3) [so sánh phương trình (3.1.7) của mô hình hai biến]. Đặc tính này nhìn chung thường được thỏa. Như vậy, mô hình hồi qui tuyến tính *k*- biến [một biến hồi qui phụ thuộc và (k-1) biến hồi qui độc lập]

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i \tag{7.4.20}$$

chúng ta có $\hat{\beta}_2$ (7.4.21)

2. Giá trị trung bình của Y_i (= Y_i) được ước lượng sẽ tương đương với giá trị trung bình của Y_i thực, điều này dễ chứng minh:

$$\begin{aligned}
\hat{Y}_i &= \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} \\
&= \left(\bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3\right) + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} \qquad \text{(Tại sao?)} \\
&= \left(\bar{Y} + \hat{\beta}_2 (X_{2i} - \bar{X}_2\right) + \hat{\beta}_3 (X_{3i} - \bar{X}_3) \\
&= \bar{Y} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i}
\end{aligned} \tag{7.4.22}$$

trong đó, các mẫu tự viết thường thông dụng dùng để biểu thị các giá trị của các biến khi chúng lệch khỏi các các giá trị trung bình tương ứng.

Lấy tổng của cả hai vế phương trình (7.4.22) theo các giá trị của mẫu và chia cho cỡ mẫu n ta có $\hat{Y} = \bar{Y}$ (Lưu ý: $\sum x_{2i} = \sum x_{3i} = 0$. Tại sao?) Lưu ý rằng với phương trình (7.4.22) ta có thể viết

$$\hat{y}_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} \tag{7.4.23}$$

trong đó $y_i = (Y_i - \bar{Y})$.

Do đó, SRF (7.4.1) có thể diễn tả dưới *dạng độ lệch* như sau

$$y_i = \hat{y}_i + \hat{u}_i = \hat{\beta}_2 x_{2i} = \hat{\beta}_3 x_{3i} + \hat{u}_i \tag{7.4.24}$$

3. Có thể chứng minh $\sum \hat{u}_i = \bar{\hat{u}} = 0$ từ phương trình (7.4.24). [Gợi ý: lấy tổng cả hai vế của (7.4.24) theo các giá trị của mẫu.]

4. Các phần dư u_i không tương quan với X_{2i} và X_{3i}, có nghĩa là, $\sum \hat{u}_i X_{2i} = \sum \hat{u}_i X_{3i} = 0$ (xem phụ lục 7A.1 để biết thêm minh chứng).

5. Các phần dư u_i không tương quan với Y_i, có nghĩa là, $\sum \hat{u}_i \hat{Y}_i = 0$. Tại sao? [Gợi ý: Nhân hai vế của (7.4.23) với u_i và lấy tổng theo các giá trị của mẫu.

6. Từ (7.4.12) và (7.4.15) ta thấy rõ ràng là $r_{2\,3}$, hệ số tương quan giữa X_2 và X_3, tăng dần về 1, các phương sai của β_2 và β_3 tăng theo các giá trị đã biết của σ^2 và $\sum x_{2i}^2$ hay $\sum x_{3i}^2$. Trong giới hạn, khi $r_{2\,3} = 1$ (tức cộng tuyến hoàn toàn), những phương sai này trở nên vô hạn. Ý nghĩa của điều này sẽ được tìm hiểu đầy đủ ở Chương 10, nhưng về mặt trực giác các bạn đọc có thể thấy rằng khi $r_{2\,3}$ tăng thì càng khó khăn hơn nếu muốn biết các giá trị thực của β_2 và β_3. [Chúng ta sẽ bàn thêm về điều này trong chương tới, nhưng xem lại phương trình (7.1.10).]

7. Từ (7.4.12) và (7.4.15) ta cũng thấy rõ ràng là đối với những giá trị của $r_{2\,3}$ và $\sum x_{2i}^2$ hay $\sum x_{3i}^2$, các phương sai của hàm ước lượng OLS sẽ tỉ lệ trực tiếp so với σ^2, có nghĩa là, chúng tăng khi σ^2 tăng. Tương tự, đối với những giá trị đã biết của σ^2 và $r_{2\,3}$ phương sai của β_2 tỉ lệ nghịch với $\sum x_{2i}^2$, có nghĩa là biến động trong các giá trị mẫu của X_2 càng lớn thì phương sai của β_2 càng nhỏ, và do đó có thể ước lượng được β_2 một cách chính xác hơn. Điều tương tự cũng đúng với phương sai của β_3.

8. Với những giả thiết của mô hình hồi qui tuyến tính cổ điển, mà ta đã trình bày cặn kẽ ở Phần 7.1, chúng ta có thể chứng minh rằng các hàm ước lượng OLS của hệ số hồi qui riêng phần không những là tuyến tính và không thiên lệch mà còn có phương sai nhỏ nhất trong nhóm các hàm ước lượng không thiên lệch tuyến tính. Nói tóm lại, *chúng là BLUE:* nói một cách khác, chúng thỏa định lý Gauss-Markov. (Chứng cớ tương tự với trường hợp hai biến đã được chứng minh ở Phụ lục 3A, Phần 3A.6 và sẽ được trình bày một cách súc tích hơn ở Chương 9 bằng cách sử dụng các ký hiệu ma trận.)

Hàm Ước Lượng Thích Hợp Tối Đa Chúng tôi đã lưu ý trong Chương 4, theo các giả thiết cho rằng u_i, số hạng nhiễu tổng thể, có phân phối chuẩn với trung bình là không và phương sai σ^2 là hằng số, các hàm ước lượng thích hợp tối đa (ML) và hàm ước lượng OLS của hệ số hồi qui của mô hình hai biến là giống nhau. Điều này mở rộng cho cả các mô hình với số lượng biến là bất kỳ. (Xem minh chứng ở Phần Phụ lục 7A, Phần 7A.4.) Tuy nhiên, điều này không đúng với hàm ước lượng của σ^2. Có thể cho thấy là hàm ước lượng ML của σ^2 là $\sum \hat{u}_i^2 / n$ bất kể đến số lượng biến trong mô hình, trong khi đó hàm ước lượng OLS của σ^2 là $\sum \hat{u}_i^2 / (n-2)$ trong trường hợp hai biến, $\sum \hat{u}_i^2 / (n-3)$ trong trường hợp ba biến, và $\sum \hat{u}_i^2 / (n-k)$ trong trường hợp mô hình có k biến (7.4.20). Nói tóm lại, hàm ước lượng OLS của σ^2 có tính đến số bậc tự do, trong khi hàm ước lượng ML thì không. Dĩ nhiên, nếu n là một số rất lớn, hàm ước lượng ML và OLS của σ^2 sẽ có khuynh hướng tiến gần nhau hơn. (Tại sao?)

7.5 HỆ SỐ XÁC ĐỊNH BỘI CỦA R^2 VÀ HỆ SỐ TƯƠNG QUAN BỘI R Trong trường hợp hai biến chúng ta đã thấy rằng r^2 được định nghĩa trong (3.5.5) là số đo độ thích hợp của phương trình hồi qui; nghĩa là, nó cho biết tỉ lệ hay phần trăm của toàn bộ biến động trong biến phụ thuộc Y được giải thích bởi biến giải thích (đơn) X. Ký hiệu r^2 này có thể được dễ dàng mở rộng ra cho các mô hình hồi qui có chứa nhiều hơn hai biến. Như vậy, trong mô hình ba biến chúng ta muốn biết tỉ lệ biến đổi trong Y được giải thích một cách liên kết bởi các biến X_2 và X_3. Đại lượng cho ta thông tin này được gọi là **hệ số xác định bội** và được ký hiệu là R^2; về mặt khái niệm nó cũng giống như r^2.

Để suy ra R^2, chúng ta có thể thực hiện giống như các phép tính của r^2 trong phần 3.5. Nhớ lại rằng

$$\begin{aligned} Y_i &= \hat\beta_1 + \hat\beta_2 X_{2i} + \hat\beta_3 X_{3i} + \hat u_i \\ &= \hat y_i + \hat u_i \end{aligned} \tag{7.5.1}$$

trong đó Y_i là giá trị ước lượng của Y_i từ đường hồi qui thích hợp và là hàm ước lượng của $E(Y_i | X_{2i}, X_{3i})$ đúng. Khi chuyển thành các mẫu tự viết thường để biểu thị độ lệch so với giá trị trung bình, phương trình (7.5.1) có thể được viết lại thành

$$\begin{aligned} y_i &= \hat\beta_2 x_{2i} + \hat\beta_3 x_{3i} + \hat u_i \\ &= \hat Y_i + \hat u_i \end{aligned} \tag{7.5.2 }$$

Bình phương cả hai vế của phương trình (7.5.1) và lấy tổng theo các giá trị của mẫu, chúng ta có

$$\begin{aligned} \sum y_i^2 &= \sum \hat y_i^2 + \sum \hat u_i^2 + 2\sum \hat y_i \hat u_i \\ &= \sum \hat y_i^2 + \sum \hat u_i^2 \qquad \text{(Tại sao?)} \end{aligned} \tag{7.5.3}$$

Bằng ngôn ngữ, phương trình (7.5.3) phát biểu rằng tổng của các bình phương toàn phần (TSS) bằng tổng bình phương giải thích (ESS) + tổng bình phương phần dư (RSS). Giờ đây, ta thay thế $\sum \hat u_i^2$ trong phương trình (7.4.19), ta có

$\sum y_i^2 = \sum \hat y_i^2 + \sum \hat y_i^2 - \hat\beta_2 \sum \hat y_i x_{2i} - \hat\beta_3 \sum y_i x_{3i}$ xắp xếp lại phương trình này; ta có

$$ESS = \sum \hat y_i^2 = \hat\beta_2 \sum y_i x_{2i} - \hat\beta_3 \sum y_i x_{3i} \tag{7.5.4}$$

Bây giờ, theo định nghĩa

$$\begin{aligned} R^2 &= \frac{ESS}{TSS} \\ &= \frac{\hat\beta_2 \sum y_i x_{2i} + \hat\beta_3 \sum y_i x_{3i}}{\sum y_i^2} \end{aligned} \tag{7.5.5}$$
[11]

[so sánh phương trình (7.5.5) với (3.5.6).]Bởi vì các đại lượng trong (7.5.5) thường được tính toán một cách quen thuộc, R^2 có thể được tính một cách dễ dàng. Lưu ý rằng R^2, giống như r^2, nằm trong khoảng 0 đến 1. Nếu nó bằng 1, đường hồi qui thích hợp giải thích 100 phần trăm cho sự biến đổi của Y. Mặt khác, nếu nó bằng 0, mô hình không giải thích bất cứ một biến đổi nào của Y. Tuy nhiên, R^2 thường nằm giữa hai giá trị cực đại này. Độ thích hợp của mô hình được cho là "tốt hơn" nếu R^2 tiến càng gần đến 1. Nhớ lại rằng, trong trường hợp hai biến chúng ta đã định nghĩa đại lượng r là hệ số tương quan và biểu thị rằng nó là số đo mức độ quan hệ (tuyến tính) giữa hai biến. Tương tự với r, trong mô hình ba biến hay nhiều hơn là hệ số **tương quan bội**, được ký hiệu là R, và nó là số đo của độ quan hệ giữa Y và tất cả các biến giải thích một cách liên kết. Mặc dù r có thể là âm hay dương, R luôn được coi là dương. Tuy nhiên, trên thực tế, tầm quan trọng của R rất nhỏ. Đại lượng có nhiều ý nghĩa hơn là R^2. Trước khi tiếp tục đi xa hơn, chúng ta hãy thiết lập mối quan hệ sau đây giữa R^2 và phương sai của hệ số hồi qui riêng phần trong mô hình hồi qui bội *k*-biến được thể hiện qua phương trình (7.4.20):

[11] Lưu ý rằng R^2 cũng có thể được tính như sau: $R^2 = 1 - \sum \hat u_i^2 / \sum y_i^2$. Tại sao?

$$\operatorname{var}(\hat{\beta}_j) = \frac{\sigma^2}{\sum x_j^2}\left(\frac{1}{1-R_j^2}\right) \quad (7.5.6)$$

trong đó $\hat{\beta}_j$ là hệ số hồi qui riêng phần của biến hồi qui độc lập X_j và R_j^2 là R^2 trong hồi qui của X_j trên (k - 2) biến hồi qui độc lập còn lại. [Lưu ý: Có (k-1) biến hồi qui độc lập trong mô hình hồi qui k-biến.] Mặc dù sự hữu dụng của phương trình (7.5.6) sẽ được thấy rõ trong Chương 10 về đa cộng tuyến, hãy quan sát rằng phương trình này chỉ đơn giản là sự mở rộng của công thức đã cho trong (7.4.12) và (7.4.15) cho mô hình hồi qui ba-biến, một biến hồi qui phụ thuộc và hai biến hồi qui độc lập.

7.6 VÍ DỤ: 7.1: ĐƯỜNG CONG PHILLIS BỔ SUNG KỲ VỌNG CỦA NƯỚC MỸ, 1970-1982. Bằng cách minh họa các ý tưởng đã được giới thiệu trong chương này cho tới bây giờ, hãy xem xét mô hình sau đây:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \quad (7.6.1)$$

trong đó Y_t = mức lạm phát thực (%) vào thời điểm t, X_{2t} = tỉ lệ thất nghiệp tại thời điểm t, và X_{3t} = mức lạm phát tiên đoán hay kỳ vọng (%) tại thời điểm t. Mô hình này được biết với tên gọi là *đường cong Phillis bổ sung kỳ vọng*.[12] Theo lý thuyết kinh tế vĩ mô β_2 được kỳ vọng là số âm (tại sao?) và β_3 được kỳ vọng là số dương (các bạn có thấy được cơ sở lý luận hay không?); sự thật là theo lý thuyết chúng ta sẽ có β_3 =1. Để kiểm định mô hình này, chúng ta thu thập dữ liệu trong bảng 7.1. Dựa trên những dữ liệu này, phương pháp OLS đưa đến những kết quả sau.[13]

$$\begin{aligned} \hat{Y}_t &= 7.1933 \;-\; 1.3925X_{2t} \;+\; 1.4700X_{3t} \\ &\quad (1.5948) \quad (0.3050) \quad\quad (0.1758) \\ &\quad R^2 = 0.8766 \end{aligned} \quad (7.6.2)$$

BẢNG 7.1
Tỉ lệ lạm phát thực Y (%), tỉ lệ thất nghiệp X_2 (%) và tỉ lệ lạm phát kỳ vọng X_3(%); Mỹ, 1970-1982

Năm	**Y***	**X_2**	**X_3**
1970	5.92	4.9	4.78
1971	4.30	5.9	3.84
1972	3.30	5.6	3.13
1973	6.23	4.9	3.44
1974	10.97	5.6	6.84
1975	9.14	8.5	9.47
1976	5.77	7.7	6.51
1977	6.45	7.1	5.92
1978	7.60	6.1	6.08
1979	11.47	5.8	8.09
1980	13.46	7.1	10.01
1981	10.24	7.6	10.81
1982	5.99	9.7	8.00

[12] Muốn đọc thêm về vấn đề này, xem Rudiger Dornbush và Stanley Fischer, *Kinh tế Vĩ mô*, McGraw-Hill, An bản lần 3, New York, 1984, trang 425.
[13] Tôi mang ơn Alan Gilbert vì đã thu thập những dữ liệu này.

Nguồn: Dữ liệu về Y và X_2 được thu thập từ các bài báo khác nhau của Business Statistics (Thống kê Kinh doanh) 1982, Bộ thương mại Mỹ, Văn phòng phân tích Kinh tế; dữ liệu X_3 được lấy từ Sự kiện Kinh tế (Economic Review), Federal Reserve Bank of Richmond, các số phát hành khác nhau.
* Thay đổi phần trăm trong Chỉ số Giá cả Người tiêu dùng

trong đó các số trong ngoặc là những sai số chuẩn ước lượng. Cách giải thích hồi qui này là như sau: Trong giai đoạn mẫu, nếu cả hai X_2 và X_3 được cố định bằng 0, mức lạm phát thực trung bình sẽ bằng khoảng 7.19%. Nhưng như chúng ta đã lưu ý trong nhiều lần, cách giải thích tung độ gốc này đơn thuần là máy móc. Thông thường nó không có một ý nghĩa gì về mặt kinh tế hay thực tế. Hệ số hồi qui riêng phần - 1.3925 có nghĩa là bằng cách giữ cho X_3 (mức lạm phát kỳ vọng) là không đổi, mức lạm phát thực trung bình tăng (giảm) vào khoảng 1.4% đối với sự giảm (tăng) của mỗi đơn vị (ở đây là đơn vị phần trăm) của tỉ lệ thất nghiệp trong giai đoạn 1970-1982. Tương tự, bằng cách giữ cho tỉ lệ thất nghiệp không đổi, giá trị hệ số 1.4700 cho thấy rằng trong cùng giai đoạn mức lạm phát thực trung bình tăng khoảng 1.47% đối với mỗi gia tăng điểm phần trăm của mức lạm phát tiên đoán hay kỳ vọng. Giá trị R^2 0.88 có nghĩa là hai biến giải thích gộp lại giải thích cho khoảng 88% sự biến đổi của mức lạm phát thực, một mức năng lực giải thích khá cao bởi vì R^2 cao nhất chỉ có thể bằng 1.

Đứng về mặt kỳ vọng tiên liệu, cả hai biến giải thích đều có các dấu hiệu kỳ vọng. Hệ số của biến lạm phát kỳ vọng về mặt thống kê có bằng 1 không? Chúng ta sẽ trả lời câu hỏi này trong Chương8.

7.7 HỒI QUI ĐƠN TRONG BỐI CẢNH HỒI QUI BỘI: GIỚI THIỆU KHÁI NIỆM THIÊN LỆCH ĐẶC TRƯNG[14]

Giả định (7.1.6) về mô hình hồi qui tuyến tính cổ điển cho rằng mô hình hồi qui áp dụng trong phân tích là được xác định đúng, có nghĩa là không có sai số hay thiên lệch đặc trưng (xem Chương 3 đọc các câu nhận xét giới thiệu). Mặc dù đề tài về phân tích đặc trưng sẽ được bàn luận một cách tường tận chi tiết hơn trong Chương 13, ví dụ minh họa đã cho trong phần trước là một cơ hội để giúp các bạn đọc hiểu được tầm quan trọng của giả định (7.1.6) nhưng đồng thời còn làm sáng tỏ thêm về ý nghĩa của hệ số hồi qui riêng phần và là phần giới thiệu tương đối bài bản cho đề tài thiên lệch đặc trưng.Giả định rằng (7.1.6) là mô hình "thực" giải thích hành vi của mức lạm phát thực trên khía cạnh mức thất nghiệp và mức lạm phát kỳ vọng. Nhưng giả sử có người nhất mực cho rằng mô hình hồi qui hai biến sau là thích hợp (đường cong Phillips gốc):

$$Y_t = b_1 + b_{12} X_{2t} + \hat{u}_{1t} \tag{7.7.1}$$

trong đó Y_t = Y_t = mức lạm phát thực (%) vào thời điểm t, X_{2t} = tỉ lệ thất nghiệp tại thời điểm t, và u_t = phần dư. Hệ số độ dốc, $b_{1\,2}$, cho biết ảnh hưởng thay đổi một đơn vị của tỉ lệ thất nghiệp đối với mức lạm phát thực trung bình.Bởi vì (7.6.1) là mô hình "đúng", (7.7.1) tạo nên một sai số đặc trưng; ở đây sai số chính là ở chỗ *loại bỏ* biến X_3, mức lạm phát kỳ vọng, ra khỏi mô hình. Chúng ta biết rằng β_2 của hồi qui bội (7.6.1) là hàm ước lượng không thiên lệch của β_2 đúng, có nghĩa là, $E(\beta_2) = \beta_2$. (Tại sao?) Liệu chỉ có $b_{1\,2}$, hệ số hồi qui đơn trong hồi qui của Y theo X_2 thôi, cũng cho ta một hàm ước lượng không

[14] Phần này chịu ảnh hưởng của Ronald J. Wonnacott và Thomas H. Wonnacott, *Kinh tế lượng,* An bản lần 2, John Wiley, New York, 1979, trang 95-98.

thiên lệch của β_2? Có nghĩa, liệu $E(b_{12}) = \beta_2$? (Nếu trường hợp này đúng là vậy, thì $b_{12} = \beta_2$). Xét ví dụ của chúng ta, hệ số của biến tỉ lệ thất nghiệp trong (7.7.1) có cung cấp cho ta một ước lượng không thiên lệch về ảnh hưởng đúng của nó đối với mức lạm phát thực không, biết rằng chúng ta đã loại bỏ X_3, mức lạm phát kỳ vọng, ra khỏi phân tích này? *Tổng quát* câu trả lời là b_{12} sẽ không phải là một hàm ước lượng không thiên lệch của β_2. Đồng thời, var(b_{12}) có thể là một hàm ước lượng thiên lệch của var(β_2). Sự thật là, chúng ta có thể chứng minh rằng (xem Phụ lục 7A, Phần 7A.5)

$$b_{12} = \beta_2 + \beta_3 b_{32} + \text{số hạng sai số} \qquad (7.7.2)$$

trong đó b_{23} là hệ số độ dốc của hồi qui của X_3 theo X_2, tức nghĩa là [15]

$$X_{3t} = b_2 + b_{32} X_{2t} + \hat{u}_{2t} \qquad (7.7.3)$$

trong đó u_2 là số hạng phần dư. Lưu ý rằng (7.7.3) đơn giản chỉ là hồi qui của biến bị loại bỏ X_3 theo X_2. Từ (7.7.2) có thể chứng minh một cách dễ dàng rằng

$$E(b_{12}) = \beta_2 + \beta_3 b_{32} \qquad (7.7.4)$$

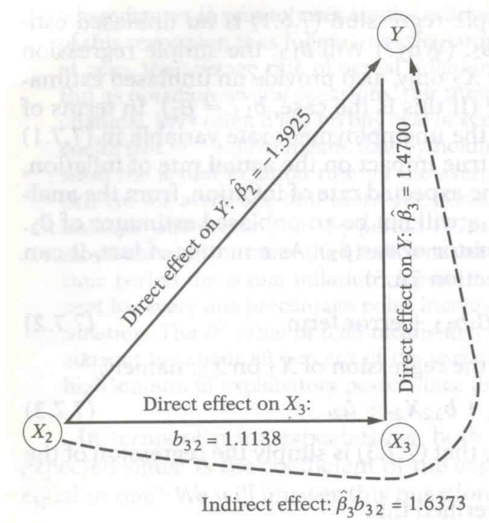
(Lưu ý: Đối với một mẫu đã cho $[b_{32} = (\sum x_{3i}x_{2i}) / \sum x_{2i}^2]$ là một hằng số đã biết.) Như phương trình (7.7.4) cho thấy, chừng nào $\beta_3 b_{32}$ không bằng 0, b_{12} sẽ là một hàm ước lượng thiên lệch của β_2. Nếu $\beta_3 b_{32}$ là số dương, tính một cách trung bình, b_{12} sẽ ước lượng quá cao β_2 (tại sao?), có nghĩa là b_{12} là thiên lệch về bên trên và nếu nó là số âm, tính một cách trung bình, b_{12} sẽ ước lượng quá thấp β_2 (tại sao?), có nghĩa là nó bị thiên lệch về bên dưới.

Tất cả những điều này thật sự có ý nghĩa gì? Như phương trình (7.7.2) cho thấy, hệ số hồi qui đơn b_{12} không chỉ là số đo của ảnh hưởng "trực tiếp" hay "ròng" của X_2 trên Y (tức giữ cho ảnh hưởng của X_3 không đổi) mà còn là số đo của ảnh hưởng gián tiếp hay kích thích trên Y thông qua ảnh hưởng của nó đối với biến bị loại bỏ X_3. Nói tóm lại, b_{12} là số đo của ảnh hưởng "toàn bộ" (trực tiếp lẫn gián tiếp) của X_2 trên Y, trong khi đó β_2 chỉ là số đo của ảnh hưởng trực tiếp hay ròng của X_2 đối với Y, bởi vì ảnh hưởng của X_3 là không đổi khi chúng ta ước lượng hồi qui bội (7.6.2), như chúng ta đã làm trong (7.6.2). Diễn đạt bằng ngôn ngữ chúng ta có

Ảnh hưởng gộp của X_2 đối với Y(=b_{12}) = ảnh hưởng trực tiếp trên X_2 đối với Y(=β_2) + ảnh hưởng gián tiếp của X_2 đối với Y(=$\beta_3 b_{32}$) (7.7.5)

Xét trong ví dụ của chúng ta, ảnh hưởng gộp của thay đổi một đơn vị trong tỉ lệ thất nghiệp đối với mức lạm phát thực bằng với ảnh hưởng trực tiếp của nó (tức, giữ cho ảnh hưởng của mức lạm phát kỳ vọng không đổi) cộng với ảnh hưởng gián tiếp là kết quả của nó (tức mức thất nghiệp) gây ra đối với mức lạm phát kỳ vọng (= b_{32}), mà bản thân nó có một số ảnh hưởng trực tiếp (= β_3) đối với mức lạm phát thực. Tất cả những điều này có thể được thấy rõ ràng hơn qua hình 7.3; những con số trình bày ở hình này là lấy từ ví dụ minh họa sắp được giải thích ở dưới.

[15] Điều này có vi phạm giả định "phi đa cộng tuyến không? Câu trả lời nằm ở Chú thích 6.



HÌNH 7.3
Các ảnh hưởng trực tiếp và gián tiếp của X_2 lên Y.

Chúng ta đã nói đủ về lý thuyết. Giờ hãy quay lại ví dụ đường cong Phillips để minh họa. Dùng dữ liệu đã cho trong bảng 7.1, chúng ta tính (7.7.1) như sau

$$\begin{aligned} Y_t &= 6.1272 + 0.2448X_{2t} \\ &\quad\ (4.2853) \quad (0.6304) \\ t &= (1.2498) \quad (0.3885) \qquad r^2 = 0.0135 \end{aligned} \tag{7.7.6}$$

Đặc điểm nổi bật của phương trình này là $b_{12} = 0.2448$ không chỉ là số dương (một đường cong Phillips dốc dương?) mà còn khác zero không đáng kể về mặt thống kê. Nhưng từ (7.6.2) chúng ta quan sát thấy $\hat\beta_2 = -1.3925$ không những có dấu tiên nghiệm đúng, như chúng tôi sẽ trình bày ở Chương 8), mà còn khác zero rất lớn. Tại sao như vậy? Câu trả lời nằm trong số hạng ảnh hưởng gián tiếp, hay yếu tố thiên lệch; $\hat\beta_3 b_{32}$, đã cho trong (7.7.4). Từ (7.6.2) chúng ta biết rằng $\hat\beta_3 = 1.4700$. Để tính được b_{23}, chúng ta tiến hành hồi qui (7.7.3), thu được các kết quả sau:

$$\begin{aligned} X_{3t} &= \ \ 0.7252 \quad + \quad 1.1138X_{2t} \\ &\quad\ (2.7267) \qquad\quad (0.4011) \\ t &= (-0.2659) \qquad (2.7769) \qquad r^2 = 0.4120 \end{aligned} \tag{7.7.7}$$

Như phương trình này cho thấy, $b_{23} = 1.1138$ có nghĩa là khi X_2 gia tăng thêm một đơn vị, tính trung bình X_3 sẽ tăng thêm 1.11 đơn vị.[16] Nhưng nếu X_3 tăng thêm bằng như vậy đơn vị, ảnh hưởng của nó trên Y sẽ là (1.4700)(1.1138) = $\hat\beta_3 b_{23} = 1.6373$. Như vậy, từ (7.7.2) cuối cùng chúng ta có

$$\begin{aligned} \hat\beta_2 + \hat\beta_3 b_{32} &= -1.3925 + 1.6373 \\ &= 0.2248 \\ &= b_{12} \text{ [xem Pt. (7.7.6)]} \end{aligned}$$

Ý nghĩa của cuộc thảo luận trong phần này là đơn giản là như sau: Nếu phải cần đến một hồi qui ba biến; đừng tiến hành chạy một hồi qui hai biến hay hồi qui đơn. Hay nói một

[16] Nhưng chẳng phải chúng ta, vì giả định phi đa cộng tuyến, phải loại trừ việc đưa các biến hồi qui độc lập có tương quan vào trong mô hình của chúng ta hay sao? Toàn bộ câu trả lời sẽ được đưa ra trong Chương 10. Ở đây chỉ lưu ý là giả định phi đa cộng tuyến gắn với hàm hồi qui tổng thể chớ không phải với hàm hồi qui mẫu; trong một mẫu đã biết chúng ta không thể kiểm soát các biến X có liên quan như thế nào ngoại trừ tiến hành những thí nghiệm có kiểm soát, điều này không phải là một viễn cảnh thú vị gì trong hầu hết các ngành khoa học xã hội.

cách tổng quát hơn, nếu các bạn chọn một mô hình hồi qui nhất định làm mô hình "đúng", đừng sửa đổi nó bằng cách bỏ bớt một biến hay nhiều hơn ra khỏi mô hình. Nếu các bạn bỏ quên nguyên tắc này, bạn sẽ thu được những ước lượng thiên lệch của các thông số. Không những vậy, bạn rất có thể sẽ ước lượng thấp phương sai đúng (σ^2) và như vậy ước lượng thấp cả sai số chuẩn của các hệ số hồi qui. Mặc dù chúng tôi sẽ chứng minh điều này một cách bài bản ở Chương 13, các bạn có thể thấy sơ qua điều này bằng cách so sánh các kết quả của hồi qui (7.6.2) và (7.7.6): Sai số chuẩn β_2 nhỏ hơn nhiều (liên hệ với hệ số của nó) ở (7.6.2) so với β_2 (liên hệ với hệ số của nó) ở (7.7.6). Do đó, các khoảng tin cậy và kiểm định giả thiết dựa trên mô hình (đúng) (7.6.2) có nhiều khả năng đáng tin cậy hơn so với các các khoảng tin cậy và giả thiết kiểm nghiệm dựa trên những mô hình được xác định sai (7.7.6).

7.8 R^2 VÀ R^2 CÓ HIỆU CHỈNHMột đặc tính quan trọng của R^2 đó là nó là một hàm không giảm của số lượng các biến giải thích hay biến hồi qui độc lập có trong mô hình; khi số lượng các biến hồi qui độc lập gia tăng, R^2 hầu như luôn luôn sẽ tăng theo và không bao giờ giảm. Phát biểu một cách khác, thêm một biến X sẽ không làm giảm R^2. Để thấy được điều này, hãy nhớ lại định nghĩa của hệ số xác định:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} \tag{7.8.1}$$

Bây giờ $\sum y_i^2$ là độc lập với số lượng các biến X trong mô hình bởi vì nó chỉ đơn giản là $\sum (Y_i - \bar{Y})^2$. Tuy nhiên RSS, $\sum \hat{u}_i^2$ phụ thuộc vào số lượng các biến độc lập trong mô hình. Bằng trực giác, ta thấy rõ là khi số lượng các biến X gia tăng, $\sum \hat{u}_i^2$ có khuynh hướng giảm (ít nhất thì nó cũng sẽ không tăng); như vậy, R^2 đã được định nghĩa trong (7.8.1) sẽ gia tăng. Vì lý do này, trong khi so sánh hai mô hình hồi qui với cùng *biến phụ thuộc* nhưng có số biến X khác nhau, các bạn cần phải cẩn thận trong việc chọn lựa mô hình với R^2 cao nhất.
Để so sánh hai số hạng R^2, ta cần phải tính đến số lượng biến X có trong mô hình. Có thể thực hiện điều này được một cách dễ dàng nếu chúng ta xem xét một hệ số xác định thay thế khác, là như sau: $R^2 = 1 - \frac{\sum \hat{u}_i^2 / (n-k)}{\sum y_i^2 / (n-1)}$ (7.8.2)

trong đó k = số lượng các thông số trong mô hình *bao gồm cả số hạng tung độ gốc*. (Trong hồi qui ba biến, k = 3. Tại sao?) Như vậy R^2 đã được định nghĩa được gọi là **R^2 có hiệu chỉnh**, được ký hiệu là $\bar{R}^2$. Thuật ngữ có hiệu chỉnh có nghĩa là hiệu chỉnh theo bậc tự do tương quan với tổng các binh phương trong (7.8.1): $\sum \hat{u}_i^2$ có n - k bậc tự do trong một mô hình có k thông số, bao gồm cả số hạng tung độ gốc, và $\sum y_i^2$ co n - 1 bậc tự do. (Tại sao?) Đối với trường hợp ba biến, chúng ta biết rằng $\sum \hat{u}_i^2$ có n - 3 bậc tự do.

Phương trình (7.8.2) cũng có thể được viết thành

$$\bar{R}^2 = 1 - \frac{\hat{\sigma}^2}{S_Y^2} \tag{7.8.3}$$

trong đó $\hat{\sigma}^2$ là phương sai phần dư, một hàm ước lượng không thiên lệch của σ^2 đúng, và S_Y^2 là phương sai của mẫu của Y. Dễ dàng thấy rằng $\overline{R}^2$ và R^2 là có liên quan với nhau, bởi vì nếu thay thế (7.8.1) vào (7.8.2), chúng ta thu được

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k} \tag{7.8.4}$$

Từ (7.8.4) ta thấy ngay lập tức, rõ ràng là (1) đối với $k > 1$, $\overline{R}^2 < R^2$ điều này có nghĩa là khi số lượng các biến X tăng, R^2 có hiệu chỉnh tăng ít hơn R^2 không hiệu chỉnh; và (2) $\overline{R}^2$ có thể là âm, mặc dù R^2 dĩ nhiên là không âm.[17] Trong trường hợp $\overline{R}^2$ là âm khi đem áp dụng, giá trị của nó được coi là bằng không. (Các bạn đọc nên kiểm chứng rằng đối với ví dụ minh họa đã cho ở trên $\overline{R}^2$ bằng 0.8519, nhỏ hơn giá trị R^2 0.8766.)

Trên thực tiễn, ta nên dùng R^2 nào? Như Theil lưu ý:

. . . dùng $\overline{R}^2$ tốt hơn R^2 bởi vì R^2 có khuynh hướng cho ra một bức tranh quá lạc quan về độ thích hợp của hồi qui, đặc biệt khi số lượng các biến giải thích là không quá nhỏ so với số lượng các lần quan sát.[18]

Nhưng quan điểm của ông Theil không hoàn toàn được mọi người tán đồng, bởi vì ông không đưa ra một chứng minh lý thuyết chung nào cho sự "ưu việt" hơn của $\overline{R}^2$. Ví dụ như, tác giả Goldberger lập luận rằng R^2 sau đây, gọi là **R^2 sửa đổi**, cũng hoàn toàn tốt như vậy:[19]

$$\textbf{R}^2\ \textbf{sửa đổi} = (1 - k/n)R^2 \tag{7.8.5}$$

Lời khuyên của ông ta là cứ trình bày R^2, *n* và *k* và để độc giả quyết định hiệu chỉnh R^2 như thế nào bằng cách lưu ý đến *n* và *k*.

Mặc dù vậy, chính R^2 có hiệu chỉnh như trong (7.4.8) là được trình bày trong hầu hết các phần mềm thống kê cùng với R^2 quy ước. Chúng tôi khuyên các bạn đọc nên xem $\overline{R}^2$ như là một số thống kê tổng hợp khác.

Bên cạnh R^2 và R^2 có hiệu chỉnh là đại lượng đo độ thích hợp, những tiêu chí khác thường được sử dụng để đánh giá sự thỏa đáng của một mô hình hồi qui. Hai trong số những tiêu chí này là **tiêu chí Thông tin của Akaike** và **tiêu chí Tiên đoán của Amemiya,** chúng được sử dụng để lựa chọn giữa các mô hình cạnh tranh với nhau. Chúng ta sẽ thảo luận những tiêu chí này khi xem đến vấn đề chọn lựa mô hình một cách chi tiết hơn trong một chương sau (xem Chương 14).

So Sánh Hai Giá Trị R^2

---

[17] Tuy nhiên, lưu ý rằng nếu $R^2 = 1$, $\overline{R}^2 = R^2 = 1$. Khi $R^2 = 0$, $\overline{R}^2 = (1-k)(n-k)$, trong trường hợp đó $\overline{R}^2$ có thể là âm nếu k > 1.

[18] Henri Theil, *Introduction to Econometrics* (Giới thiệu Kinh tế lượng), Prentice-Hall, Englewood Cliffs, N. J., 1978, trang 135.

[19] Arthur S. Goldberger, *Khóa học Kinh tế lượng*, Havard U. Press, Cambridge, Massachsetts,1991, trang 178. Về quan điểm phê bình hơn về R^2 xem S. Cameron, "Tại sao R bình phương có hiệu chỉnh được trình bày?", *Journal of Quantitative Economics (Tạp chí về Kinh tế Định lượng)*, tập 9, số 1, tháng 1, 1993, tr. 183-186. Ông lập luận rằng "Nó [R^2] KHÔNG phải là một trị thống kê kiểm định và dường như kh6ng có một sự bào chữa nào về mặt trực giác để sử dụng nó như một trị thống kê mô tả. Cuối cùng, chúng ta cần hiểu rõ rằng nó không phải là một công cụ hữu dụng để ngăn ngừa sự khai thác dữ liệu" (trang 186).

Điều quan trọng phải lưu ý khi so sánh hai mô hình trên nền tảng hệ số xác định, dù có hiệu chỉnh hay không, đó là *cỡ mẫu n và biến phụ thuộc của hai mô hình phải giống nhau*; các biến giải thích có thể có bất cứ dạng gì. Như vậy, đối với các mô hình

$$\ln Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \tag{7.8.6}$$

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_i \tag{7.8.7}$$

các số hạng R^2 đã được tính không thể mang so sánh được. Lý do là như sau: Theo định nghĩa, R^2 là số đo tỉ lệ biến thiên trong biến phụ thuộc do (các) biến giải thích giải thích. Như vậy, trong (7.8.6) R^2 đo tỉ lệ *biến thiên* trong ln Y do X_2 và X_3 giải thích, trong khi đó trong (7.8.7) R^2 đo tỉ lệ *biến thiên* trong Y, và hai số đo này không giống nhau: như đã lưu ý ở Chương 6, thay đổi trong ln Y dẫn tới một thay đổi tương đối hay tỉ lệ trong Y, trong khi đó thay đổi trong Y dẫn tới một thay đổi tuyệt đối. Do đó, $\text{var}Y_i/\text{var}Y_i$ không tương đương với $\text{var}(\ln Y_i)/\text{var}(\ln Y_i)$, có nghĩa là, hai hệ số xác định không giống nhau.[20]

Nếu chúng ta xem lại hàm nhu cầu cà phê (3.7.1), với đặc trưng tuyến tính, và (6.4.5), có đặc trưng tuyến tính logarit, do đó hai số hạng r^2 0.6628 và 0.7448, không thể so sánh trực tiếp với nhau được.[21] Như vậy làm cách nào chúng ta có thể so sánh được các số hạng R^2 của các mô hình như (3.7.1) và (6.4.5)? Chúng ta sẽ trình bày điều này bằng ví dụ về nhu cầu cà phê của chúng ta.

Ví Dụ 7.2 : Xem Xét Lại Hàm Nhu Cầu Cà Phê

Để so sánh các giá trị R^2 tính được từ hai mô hình trong đó các biến phụ thuộc là không giống nhau, như trong mô hình (3.7.1) và (6.4.5), chúng ta tiến hành như sau:

1. Lấy $\ln\hat{Y}_t$ từ mô hình (6.4.5), lấy các giá trị đối logarit của chúng, và rồi tính R^2 giữa đối logarit của $\ln\hat{Y}_t$ và Y_t theo phương thức đã nêu ở phương trình (3.5.14). Giá trị R^2 này có thể so sánh được giá trị R^2 của mô hình (3.7.1).

2. Một cách khác, chúng ta lấy Y_t từ (3.7.1), chuyển chúng thành $(\ln Y_t)$, và sau cùng tính R^2 giữa $(\ln Y_t)$ và $\ln(Y_t)$ theo phương trình (3.5.14). Giá trị R^2 này có thể so sánh được với giá trị R^2 thu được từ (6.4.5).

Giả sử trước hết chúng ta quyết định so sánh giá trị R^2 của mô hình tuyến tính (3.7.1) với giá trị R^2 của mô hình logarit kép (6.4.5). Từ Y được ước lượng từ (3.7.1) đầu tiên chúng ta lấy $(\ln Y_t)$, sau đó lấy logarit của Y_t thực, rồi tính r^2 giữa hai tập hợp giá trị

[20] Từ định nghĩa của R^2 , chúng ta biết rằng

$$1 - R^2 = \frac{RSS}{TSS} = \frac{\sum \hat{u}_i^2}{\sum (Y_i - \bar{Y})^2}$$

đối với mô hình tuyến tính, và

$$1 - R^2 = \frac{\sum \hat{u}_i^2}{\sum (\ln Y_i - \ln \bar{Y})^2}$$

đối với mô hình log. Bởi vì các mẫu số ở vế bên phải của những biểu thức này là khác nhau, ta không thể so sánh trực tiếp hai số hạng R^2.

[21] Đối với đặc trưng tuyến tính, RSS = 0.1491 (tổng bình phương phần dư của tiêu thụ cà phê), và đối với đặc trưng log tuyến tính, RSS = 0.0226 (tổng bình phương phần dư của logarít của tiêu thụ cà phê). Những phần dư này có các bậc độ lớn khác nhau và vì vậy không thể so sánh một cách trực tiếp.

này theo phương trình (3.5.14). Dùng dữ liệu cho Bảng 7.2, bạn đọc có thể kiểm chứng rằng giá trị R^2 được tính như vậy là 0.7318, có thể so sánh trực tiếp với giá trị r^2 của mô hình tuyến tính-logarit (6.4.5), tức 0.7448, mặc dù giá trị R^2 thu được từ mô hình tuyến tính-logarít có cao hơn một ít.

Mặt khác nếu chúng ta muốn so sánh giá trị R^2 của mô hình tuyến tính-logarít với R^2 thu được từ mô hình tuyến tính , chúng ta tính $\ln\hat{Y}_t$ từ (6.4.5), thu được giá trị đối logarit của chúng, và cuối cùng tính R^2 giữa những giá trị đối logarit này và các giá trị thực của Y bằng cách dùng công thức (3.5.14). Các bạn đọc có thể kiểm tra lại dữ liệu đã cho trong bảng 7.2 rằng R^2 giá trị này là 0.7187, cao hơn một ít so với R^2 0.6628 thu được từ mô hình tuyến tính (3.7.1).

Dùng một trong hai phương pháp này, chúng ta thấy rằng mô hình tuyến tính-logarit cho ta độ thích tốt hơn.

Bảng 7.2
Dữ liệu thô để so sánh hai giá trị R^2

Năm	Y_t (1)	$\hat{Y}_t$ (2)	$\ln\hat{Y}_t$ (3)	Đối loga của $\ln\hat{Y}_t$ (4)	$\ln Y_t$ (5)	$\ln(\hat{Y}_t)$ (6)
1970	2.57	2.321887	0.843555	2.324616	0.943906	0.842380
1971	2.50	2.336272	0.853611	2.348111	0.916291	0.848557
1972	2.35	2.345863	0.860544	2.364447	0.854415	0.852653
1973	2.30	2.341068	0.857054	2.356209	0.832909	0.850607
1974	2.25	2.326682	0.846863	2.332318	0.810930	0.844443
1975	2.20	2.331477	0.850214	2.340149	0.788457	0.846502
1976	2.11	2.173233	0.757943	2.133882	0.746688	0.776216
1977	1.94	1.823176	0.627279	1.872508	0.662688	0.600580
1978	1.97	2.024579	0.694089	2.001884	0.678034	0.705362
1979	2.06	2.115689	0.731282	2.077742	0.722706	0.749381
1980	2.02	2.130075	0.737688	2.091096	0.703098	0.756157

Lưu ý: Cột (1): Các giá trị Y thực từ Bảng 3.4
Cột (2): Các giá trị Y ước lượng từ mô hình tuyến tính (3.7.1)
Cột (3): Các giá trị Y ước lượng từ mô hình log kép (6.4.5)
Cột (4): Đối logarít của các giá trị ở cột (3)
Cột (5): Các giá trị logarít của Y ở cột (1)
Cột (6): Các giá trị logarít của Y_t ở cột (2)

"Trò chơi" của Tối đa hóa $\bar{R}^2$

Để kết thúc phần này, chúng tôi có một lời cảnh giác: Đôi khi các nhà nghiên cứu chơi trò chơi tối đa hóa $\bar{R}^2$, có nghĩa là, chọn mô hình nào cho $\bar{R}^2$ cao nhất. Nhưng điều này có thể nguy hiểm, bởi vì trong phân tích hồi qui mục đích của chúng ta không phải là thu được một $\bar{R}^2$ cao cho mỗi mô hình mà đúng hơn là thu được những ước lượng đáng tin cậy của các hệ số hồi qui tổng thể thực và rút ra những suy diễn thống kê về chúng. Trong khi phân tích thực nghiệm vẫn thường thu được $\bar{R}^2$ rất cao nhưng thấy rằng một số những hệ số hồi qui hoặc là không có ý nghĩa về mặt thống kê hoặc có dấu trái ngược lại với những kỳ vọng tiên nghiệm. Do đó, nhà nghiên cứu nên quan tâm hơn đến sự liên hệ về mặt lý thuyết hay logíc của các biến giải thích đối với biến phụ thuộc và ý nghĩa thống kê của chúng. Nếu trong quá trình này chúng ta thu được một $\bar{R}^2$ cao, điều này là hoàn toàn tốt; mặt khác nếu $\bar{R}^2$ mà thấp, điều này không nhất thiết có nghĩa là mô hình bị kém.[22]

Sự thật là, Goldberger đã chỉ trích rất mạnh mẽ vai trò của R^2. Ông ta đã phát biểu:

> Đứng trên quan điểm của chúng tôi, R^2 có một vai trò rất khiêm tốn trong phân tích hồi qui, nó là một đại lượng đo của độ thích hợp của hồi qui tuyến tính bình phương tối thiểu (LS) mẫu trong một tập hợp các dữ liệu. Không có yếu tố nào trong mô hình hồi quy cổ điển CR [CLRM] đòi hỏi rằng R^2 phải cao. Như vậy, một R^2 cao không phải là bằng chứng có lợi cho mô hình và một R^2 thấp không phải là bằng chứng bất lợi cho nó.
>
> Sự thật, điều quan trọng nhất về R^2 là nó không hề quan trọng trong mô hình CR. Mô hình CR chú ý đến các thông số trong một tổng thể, không chú ý đến độ thích hợp của mẫu . . . . . Nếu một người cứ khăng khăng đòi hỏi một đại lượng về thành công của dự đoán (hay đúng hơn là thất bại) thì σ^2 là đủ: nói cho cùng, thông số σ^2 là sai số dự đoán bình phương kỳ vọng, nó là kết quả của nếu CEF [PRF] của tổng thể được dùng làm biến dự đoán. Nói một cách khác, bình phương sai số chuẩn của dự đoán . . . . . với các giá trị liên quan của x [biến hồi qui độc lập] có thể cung cấp cho ta biết rất nhiều điều.[23]

7.9 CÁC HỆ SỐ TƯƠNG QUAN RIÊNG PHẦN

Giải Thích Các Hệ Số Tương Quan Riêng phần Và Đơn

Trong Chương 3 chúng ta đã giới thiệu hệ số tương quan r là một đại lượng đo mức độ quan hệ tuyến tính giữa hai biến. Đối với mô hình hồi qui ba biến chúng ta có thể tính ba hệ số tương quan: $r_{1\,2}$ (tương quan giữa Y và X_2), $r_{1\,3}$ (hệ số tương quan giữa Y và X_3) và $r_{2\,3}$ (hệ số tương quan giữa X_2 và X_3); lưu ý rằng vì mục đích đơn giản hóa ký hiệu, chúng ta dùng ký hiệu 1 ở dưới để biểu thị cho Y. Những hệ số tương quan này được gọi là **hệ số tương quan đơn hay gộp**, hay **hệ số tương quan bậc zero**. Những hệ số này có thể được tính bằng định nghĩa của hệ số tương quan đã cho trong (3.5.13).

Nhưng bây giờ ta hãy xem xét câu hỏi sau: $r_{1\,2}$ có thật sự là đại lượng đo mức độ quan hệ (tuyến tính) "đúng" giữa Y và X_2 không khi một biến thứ ba X_3 có thể có quan hệ với cả hai? Câu hỏi này cũng tương tự như câu hỏi sau: Giả sử mô hình hồi qui đúng là (7.7.1) nhưng chúng ta bỏ biến X_3 ra khỏi mô hình và đơn thuần chỉ lấy hồi qui Y theo X_2, thu được hệ số độ dốc gọi là b_{12}. Hệ số này có tương đương với hệ số đúng β_2 nếu mô hình (7.7.1) được ước lượng lúc ban đầu? Câu trả lời có thể được thấy rõ ràng từ thảo luận của chúng ta ở Phần 7.7. Nhìn chung, r_{12} ít có khả năng phản ánh được mức độ tương quan đúng giữa Y và X_2 khi có sự hiện diện của X_3. Sự thật là, nó có thể đưa ra một cảm tưởng sai lầm về bản chất của quan hệ giữa Y và X_2, như chúng ta sẽ thấy ngay

[22] Một số tác giả muốn giảm nhẹ việc sử dụng R^2 như là một đại lượng đo độ thích hợp cũng như việc sử dụng nó để so sánh hai hay nhiều hơn các giá trị R^2. Xem *Interpreting and Using Regression (Giải thích và Sử dụng Hồi qui)* của Christopher H. Achen, Sage Publication, Beverly Hills, Calif., 1982, tr.58-67 và "R^2 và Phép biến đổi của các Biến Hồi qui" của C. Granger và P. Newbold, *Journal of Econometrics*, tập 4, 1976, tr.205-210. Nhân tiện đây, thực tiễn của việc chọn một mô hình dựa trên R^2 cao nhất, một kiểu khai thác dữ liệu, giới thiệu một điều gọi là **thiên lệch tiền kiểm định**, điều này có thể phá hỏng một số tính chất của hàm ước lượng OLS của mô hình hồi qui tuyến tính cổ điển. Về chủ đề này, người đọc có thể tham khảo George G. Judge, Carter R. Hill, William E. Griffiths, Helmut Lukepohl và Tsoung-Chao Lee, *Introduction to the Theory and Practice of Econometrics (Nhập Môn về Lý thuyết và Thực tiễn của Kinh tế lượng)*, John Wiley, New York, 1982,Chương 21.

[23] Arther S. Goldberger, đã đề cập, trang 177-178

dưới đây. Do đó, điều chúng ta cần là một hệ số tương quan độc lập không chịu bất kỳ ảnh hưởng của X_3 lên X_2 và Y. Một hệ số tương quan như vậy có thể tính được và được gọi một cách thích hợp là **hệ số tương quan riêng phần**. Về mặt khái niệm, nó cũng tương tự như hệ số hồi qui riêng phần. Chúng ta định nghĩa

$r_{1\,2.\,3}$ = hệ số tương quan riêng phần giữa Y và X_2, giữ X_3 không đổi
$r_{1\,3.\,2}$ = hệ số tương quan riêng phần giữa Y và X_3, giữ X_2 không đổi
$r_{2\,3.\,1}$ = hệ số tương quan riêng phần giữa X_2 và X_3, giữ Y không đổi

Một cách để tính các hệ số tương quan riêng phần ở trên là như sau: Nhớ lại quá trình ba giai đoạn chúng ta đã thảo luận trong Phần 7.3. Trong giai đoạn III chúng ta lấy hồi qui u_{1i} theo u_{2i}, chính là Y_i và X_{2i} tinh khiết, có nghĩa là, gạt bỏ ảnh hưởng tuyến tính của X_3. Do đó, nếu chúng ta bây giờ tính hệ số đơn tương quan giữa u_{1i} theo u_{2i}, chúng ta sẽ thu được $r_{1\,2.\,3}$ bởi vì biến X_3 giờ đây là không đổi. Bằng ký hiệu ta có

$$\begin{aligned} r_{\hat{u}_1\hat{u}_2} &= r_{12.3} \\ &= \frac{\sum(\hat{u}_{1i}-\bar{\hat{u}}_1)(u_{2i}-\bar{\hat{u}}_2)}{\sqrt{\sum(\hat{u}_{1i}-\operatorname{var}\hat{u}_1)^2(\hat{u}_{2i}-\bar{\hat{u}}_2)^2}} \\ &= \frac{\sum\hat{u}_{1i}\hat{u}_{2i}}{\sqrt{\sum(\hat{u}_{1i}^2\sum\hat{u}_{2i}^2)}} \end{aligned} \tag{7.9.1}$$

trong đó chúng ta áp dụng tính chất là $\bar{\hat{u}}_1 = \bar{\hat{u}}_2 = 0$. (Tại sao?)

Từ phần thảo luận ở trước rõ ràng là tương quan riêng phần giữa Y và X_2 giữ X_3 không đổi chính là hệ số tương quan đơn (hay bậc-zero) giữa các phần dư từ hồi qui của Y trên X_3 và hồi qui của X_2 trên X_3. Số hạng $r_{1\,3.\,2}$ và $r_{2\,3.\,1}$ phải được giải thích tương tự như nhau.

Trên thực tế, chúng ta không cần phải đi qua lại quá trình ba giai đoạn để tính toán các tương quan riêng phần bởi vì có thể dễ dàng thu được chúng từ các hệ số tương quan đơn hay bậc-zero như sau (để biết cách chứng minh, xem các bài tập):[24]

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{\left(1-r_{13}^2\right)\left(1-r_{23}^2\right)}} \tag{7.9.2}$$

$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{\left(1-r_{12}^2\right)\left(1-r_{23}^2\right)}} \tag{7.9.3}$$

$$r_{23.1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{\left(1-r_{12}^2\right)\left(1-r_{13}^2\right)}} \tag{7.9.4}$$

Các tương quan riêng phần đã cho trong các phương trình (7.9.2) cho đến (7.9.4) được gọi là **các hệ số tương quan bậc nhất**. Với từ *bậc* chúng tôi muốn nói rằng số lượng các chỉ số thứ hai ở dưới. Như vậy, $r_{1\,2.\,3\,4}$ sẽ là hệ số tương quan bậc hai, $r_{1\,2.\,3\,4\,5}$ sẽ là hệ số tương quan bậc ba v.v. Như đã lưu ý ở trước, $r_{1\,2}$, $r_{1\,3}$ v.v. được gọi là các *tương quan bậc*

[24] Hầu hết các chương trình điện toán cho phân tích hồi qui đa biến thường tính các hệ số tương quan đơn; vì vậy, các hệ số tương quan riêng phần có thể được tính bằng các chương trình có sẵn.

zero hay đơn. Cách giải thích $r_{12.34}$ là nó cho ta hệ số tương quan giữa Y và X_2, giữ X_3 và X_4 không đổi.

Giải thích Các hệ số tương quan
Riêng phần và Đơn

Trong trường hợp hai biến, ýnghĩa của r đơn rất đơn giản: Nó là đại lượng đo mức độ quan hệ (tuyến tính) (và không phải là quan hệ nhân quả) giữa biến độc lập Y và biến giải thích đơn X. Nhưng một khi chúng ta vượt ra khỏi trường hợp hai biến, chúng ta cần phải chú ý cẩn thận đến cách giải thích hệ số tương quan đơn. Ví dụ, từ (7.9.2) chúng ta quan sát được những điều sau:

1. Ngay cả nếu $r_{12} = 0$, $r_{12.3}$ sẽ không bằng 0 trừ khi r_{13} hay r_{23} hoặc cả hai đều bằng 0.
2. Nếu $r_{12} = 0$, r_{13} và r_{23} không bằng 0 và có cùng dấu, $r_{12.3}$ sẽ mang dấu âm, trong khi đó nếu chúng mang dấu ngược nhau, nó sẽ là số dương. Ví dụ sau sẽ làm sáng tỏ điểm này. Gọi Y = sản lượng vụ mùa , X_2 = lượng mưa, và X_3 = nhiệt độ. Giả định rằng r_{12} = 0, có nghĩa là, không có quan hệ giữa sản lượng vụ mùa và lượng mưa. Giả định thêm rằng r_{13} là dương và r_{23} âm. Khi đó, như (7.9.2) cho thấy, $r_{12.3}$ sẽ là dương; có nghĩa là, giữ cho nhiệt độ không đổi, có quan hệ đồng biến giữa sản lượng vụ mùa và lượng mưa. Tuy nhiên, kết quả dường như là nghịch lý này không có gì là ngạc nhiên. Bởi vì nhiệt độ X_3 ảnh hưởng cả sản lượng vụ mùa Y và lượng mưa X_2, để biết được quan hệ ròng giữa sản lượng vụ mùa và lượng mưa, chúng ta cần loại bỏ ảnh hưởng của biến nhiệt độ "phiền toái". Ví dụ này cho thấy một người có thể bị sai lầm như thế nào bởi hệ số tương quan đơn.
3. Số hạng $r_{12\,3}$ và r_{12} (và những so sánh tương tự) không cần phải có cùng dấu.
4. Trong trường hợp hai biến chúng ta đã thấy rằng r^2 nằm giữa 0 và 1. Đặc tính này vẫn đúng đối với bình phương các hệ số tương quan riêng phần. Sử dụng đặc tính này bạn đọc sẽ chứng minh được rằng từ (7.9.2)chúng ta có thể thu được biểu thức sau:

$$0 \le r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23} \le 1 \tag{7.9.5}$$

biểu thức này cho ta quan hệ lẫn nhau giữa ba hệ số tương quan bậc-zero. Cũng có thể suy ra được những biểu thức tương tự từ các phương trình (7.9.3) và (7.9.4).
5. Giả sử rằng $r_{13} = r_{23} = 0$. Điều này có nghĩa là r_{12} cũng sẽ bằng zerô không? Có thể thấy rõ câu trả lời từ (7.9.5). Việc Y và X_3, và X_2 và X_3 không tương quan không có nghĩa là Y và X_2 không tương quan.

Nhân đây, lưu ý rằng $r^2_{12.3}$ có thể được gọi là **hệ số xác định riêng phần** và có thể được giải thích là tỉ lệ của biến thiên trong Y không phải do ảnh hưởng của biến X_3 gây ra, mà được giải thích bằng cách do đưa X_2 vào mô hình (xem bài tập 7.6). Về mặt khái niệm nó giống như R^2.

Trước khi chúng ta tiếp tục, lưu ý những quan hệ sau giữa các R^2, các hệ số tương quan đơn, và các hệ số tương quan riêng phần:

$$R^2 = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2} \tag{7.9.6}$$

$$R^2 = r_{12}^2 + (1 - r_{12}^2)r_{13.2}^2 \tag{7.9.7}$$

$$R^2 = r_{13}^2 + (1 - r_{13}^2) r_{12.3}^2 \quad (7.9.8)$$

Để kết luận phần này, hãy xem xét điểm sau: Chúng ta đã phát biểu trước đây rằng R^2 sẽ không giảm nếu có thêm một biến giải thích được đưa vào trong mô hình, có thể thấy điều này một cách rõ ràng từ (7.9.7). Phương trình này phát biểu rằng tỉ lệ biến thiên trong Y do X_2 và X_3 liên kết giải thích là tổng của hai phần: phần chỉ do X_2 giải thích (= r^2_{12}) và phần không do X_2 giải thích (=1- r^2_{12}) nhân với tỉ lệ được giải thích bởi X_3 sau khi giữ cho ảnh hưởng của X_2 không đổi. Bây giờ $R^2 > r^2_{12}$ miễn là $r^2_{12.3} > 0$. Tình huống xấu nhất, $r^2_{12.3}$ sẽ bằng không, trong trường hợp đó $R^2 = r^2_{12}$.

7.10 VÍ DỤ 7.10: HÀM SẢN XUẤT COBB-DOUGLAS : NÓI THÊM VỀ DẠNG HÀM SỐ

Trong Phần 6.4 chúng ta đã cho thấy bằng những phép biến đổi thích hợp chúng ta có thể biến đổi các quan hệ phi tuyến tính thành các quan hệ tuyến tính để chúng ta có thể hoạt động trong khuôn khổ các mô hình hồi qui tuyến tính cổ điển. Các phép biến đổi khác nhau đã được thảo luận đến trong bối cảnh trường hợp hai biến có thể dễ dàng được mở rộng qua cho trường hợp các mô hình hồi qui bội. Trong phần này, chúng ta chứng minh các phép biến đổi này bằng cách mở rộng đa biến của mô hình tuyến tính-logarít hai biến; những trường hợp khác có thể được tìm thấy trong các bài tập và qua các ví dụ minh họa được thảo luận đến trong suốt phần còn lại của tập sách này. Ví dụ cụ thể mà chúng ta thảo luận là **hàm sản xuất Cobb-Douglas** nổi tiếng trong lý thuyết về sản xuất.

Hàm sản xuất Cobb-Douglas , ở dạng ngẫu nhiên của nó, có thể được biểu diễn như sau

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{ui} \quad (7.10.1)$$

trong đó Y = sản lượng

X_2 = nhập lượng lao động

X_3 = nhập lượng vốn

u = số hạng nhiễu ngẫu nhiên

e = cơ số của logarít tự nhiên

Từ phương trình (7.10.1) rõ ràng là quan hệ giữa sản lượng và hai yếu tố nhập lượng là không tuyến tính. Tuy nhiên, nếu chúng ta biến đổi logarít mô hình này, chúng ta được

$$\ln Y_i = \ln\beta_1 + \beta_2 \ln X_{3i} + u_i \;= \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \quad (7.10.2)$$

trong đó $\beta_0 = \ln\beta_1$.

Viết như vậy, mô hình là tuyến tính theo thông số β_0, β_2 và β_3 và do đó là một mô hình hồi qui tuyến tính. Tuy nhiên lưu ý rằng nó không tuyến tính theo các biến Y và X nhưng tuyến tính theo logarít của các biến này. Tóm lại, (7.10.2) là một *logarít-logarít, logarit kép*, hay mô hình *tuyến tính-logarít*, là mô hình hồi qui bội tương ứng với mô hình tuyến tính logarít hai biến (6.4.3).

Các đặc tính của hàm sản xuất Cobb-Douglas được biết rất rộng rãi:

1. β_2 là độ co dãn (riêng phần) của sản lượng so với nhập lượng lao động, có nghĩa là, nó đo thay đổi phần trăm trong sản lượng ứng với, ví dụ như, thay đổi 1 phần trăm trong nhập lượng lao động, giữ cho nhập lượng vốn không đổi (xem bài tập 7.10).

2. Tương tự như vậy, β_3 là độ co dãn (riêng phần) của sản lượng so với nhập lượng vốn, giữ cho nhập lượng lao động không đổi.

3. Tổng của ($\beta_2 + \beta_3$) cho ta thông tin về *sinh lơi theo quy mô*, có nghĩa là, sự đáp ứng của sản lượng trước một tỉ lệ thay đổi của nhập lượng. Nếu tổng này bằng 1, thì không có *sinh lợi cố định theo quy mô*, có nghĩa là, tăng gấp hai lần nhập lượng sẽ làm tăng sản lượng lên gấp hai lần, tăng gấp ba lần nhập lượng sẽ làm tăng sản lượng lên gấp ba lần, và cứ vậy. Nếu tổng này nhỏ hơn 1, tức có hiện tượng *sinh lơi giảm dần theo quy mô* -tăng gấp hai lần nhập lượng sẽ không làm sản lượng tăng lên gấp hai. Cuối cùng, nếu tổng này lớn hơn 1, có hiện tượng *sinh lơi tăng dần theo quy mô* - tăng gấp hai lần nhập lượng sẽ làm sản lượng tăng lên hơn gấp hai.

Trước khi tiếp tục, lưu ý rằng bất cứ khi nào các bạn có mô hình hồi qui tuyến tính-logarit với một số lượng biến bất kỳ, hệ số của mỗi biến X là số đo độ co dãn (riêng phần) của biến phụ thuộc Y so với biến X đó. Như vậy, nếu bạn có một mô hình tuyến tính-logarit với k-biến:

$$\ln Y_i = \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \ldots\ldots + \beta_k \ln X_{ki} + u_i \qquad (7.10.3)$$

mỗi hệ số hồi qui (riêng phần), β_2 cho đến β_k, là độ co dãn (riêng phần) của Y so với các biến X_2 cho đến X_k.[25]

Để minh họa hàm sản xuất Cobb-Douglas, chúng ta thu thập các dữ liệu trình bày trong Bảng 7.3; những dữ liệu này là của khu vực nông nghiệp của Đài Loan trong giai đoạn 1958-1972.

BẢNG 7.3
Tổng sản lượng thực, ngày lao động và nhập lượng vốn trong khu vực nông nghiệp của Đài Loan, 1958-1972

Năm	Tổng sản lượng thực (triệu NT $)*, Y	Ngày lao động (triệu ngày), X_2	Nhập lượng vốn thực (triệu NT $), X_3
1958	16,607.7	275.5	17,803.7
1959	17,511.3	274.4	18,096.8
1960	20,171.2	269.7	18,271.8
1961	20,932.9	267.0	19,167.3
1962	20,406.0	267.8	19,647.6
1963	20,831.6	275.0	20,803.5
1964	24,806.3	283.0	22,076.6
1965	26,465.8	300.7	23,445.2
1966	27,403.0	307.5	24,939.0
1967	28,628.7	303.7	26,713.7
1968	29,904.5	304.7	29,957.8
1969	27,508.2	298.6	31,585.9
1970	29,035.5	295.5	33,474.5
1971	29,281.5	299.0	34,821.8

[25] Để xem xét điều này, lấy vi phân riêng phần (7.10.3) theo logarít của mỗi biến X. Vì vậy, $\partial \ln Y / \partial \ln X_2 = (\partial Y / \partial X_2)(X_2 / Y) = \beta_2$, giá trị này theo định nghĩa, là độ co dãn của Y theo X_2, và $\partial \ln Y / \partial \ln X_3 = (\partial Y / \partial X_3)(X_3 / Y) = \beta_3$, theo định nghĩa là độ co dãn của Y theo X_3, và v.v.

1972	31,535.8	288.1	41,794.3

Nguồn: Thomas Pei-Fan Chen. "Economic growth and Structural change in Taiwan—1952-1972, A Production Function Approach," luận văn tiến sĩ không xuất bản, Khoa Kinh tế, Graduate Center; City University of New York, June 1976, Bảng II.
* Đôla mới của Đài Loan

Giả định rằng mô hình (7.10.2) thỏa mãn được các giả định của mô hình hồi qui tuyến tính cổ điển,[26] chúng ta lấy hồi qui sau bằng phương pháp OLS (xem Phụ lục 7A, Phần 7A.7 để biết kết quả in ra từ máy tính):

$$\begin{aligned}
\ln Y_i &= -3.3348 \quad + \quad 1.4988 \ln X_{2i} \quad + \quad 0.4899 \ln X_{3i} \\
&\quad\ (2.4495) \qquad\quad (0.5398) \qquad\qquad (0.1020) \\
t &= (-1.3629) \qquad (2.7765) \qquad\qquad (4.8005) \\
&\qquad\qquad R^2 = 0.8890 \qquad df = 12 \\
&\qquad\qquad R^2 = 0.8705
\end{aligned} \tag{7.10.4}$$

Từ phương trình (7.10.4) chúng ta thấy rằng trong khu vực nông nghiệp của Đài Loan trong giai đoạn 1958-1972 độ co dãn sản lượng của lao động và vốn là bằng 1.4988 và 0.4899. Nói một cách khác, trong giai đoạn được xem xét này, giữ nhập lượng vốn không đổi, gia tăng 1 phần trăm trong lao động dẫn đến trung bình vào khoảng 1.5 phần trăm gia tăng trong sản lượng. Tương tự như vậy, giữ nhập lượng lao động không đổi, gia tăng 1 phần trăm trong vốn dẫn đến trung bình vào khoảng 0.5 phần trăm gia tăng trong sản lượng. Cộng hai độ co dãn sản lượng lại, chúng ta được 1.9887, đây là giá trị của thông số sinh lợi theo quy mô. Rõ ràng là, trong giai đoạn này, khu vực nông nghiệp của Đài Loan có đặc điểm là sinh lợi theo quy mô tăng dần.[27]

Từ quan điểm thống kê thuần túy, đường hồi qui ước lượng thích hợp với các dữ liệu rất tốt. Giá trị R^2 0.8890 có nghĩa là vào khoảng 89% sự biến thiên trong (logarit của) sản lượng là do (logarit của) lao động và vốn. Trong Chương 8, chúng ta sẽ thấy bằng cách nào những sai số chuẩn ước lượng này có thể được sử dụng để kiểm định giả thiết về các giá trị "đúng" của các thông số của hàm sản xuất Cobb-Douglas cho nền kinh tế Đài Loan (xem bài tập 8.15).

7.11 CÁC MÔ HÌNH HỒI QUI ĐA THỨC

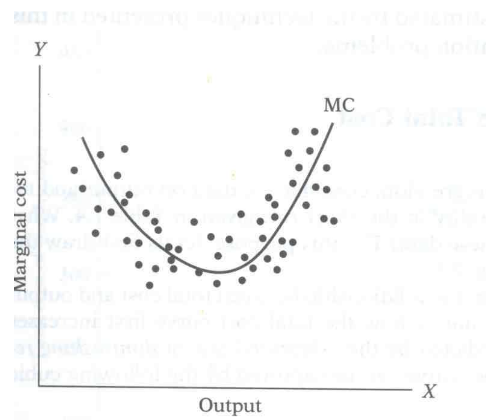
Chúng ta kết thúc chương này bằng cách xem xét một nhóm các mô hình hồi qui bội, những **mô hình hồi qui đa thức**, được sử dụng rộng rãi trong các cuộc nghiên cứu kinh tế lượng có liên quan đến hàm sản xuất và chi phí. Khi giới thiệu những mô hình này, chúng tôi sẽ mở rộng thêm phạm vi của các mô hình này để có thể dễ dàng áp dụng những mô hình hồi qui tuyến tính cổ điển vào chúng.

Để xác định các ý tưởng, xem hình 7.4, hình 7.4 này cho thấy quan hệ giữa chi phí sản xuất biên tế ngắn hạn (MC) (gọi là biến Y) của một loại hàng hóa và mức sản lượng

[26] Lưu ý rằng trong hàm sản xuất Cobb-Douglas (7.10.1) chúng ta đã giới thiệu số hạng sai số ngẫu nhiên một cách đặc biệt là biến đổi nó thành dạng logarít rồi đưa về dạng tuyến tính thường dùng. Về vấn đề này, xem phần 6.8.

[27] Chúng ta bỏ qua câu hỏi về sự thích hợp của mô hình theo quan điểm lý thuyết cũng như câu hỏi về việc liệu người ta có thể đo lường được sinh lợi theo quy mô từ chuỗi dữ liệu theo thời gian hay không.

của nó (gọi là biến X). Đường cong MC vẽ cho thấy trong hình, tức đường cong hình chữ U theo sách giáo khoa, cho thấy quan hệ giữa MC và sản lượng là không tuyến tính. Nếu chúng ta phải định lượng mối quan hệ này từ các điểm rời rạc, làm cách nào chúng ta có thể thực hiện được? Nói một cách khác, loại mô hình kinh tế lượng nào thể hiện được bản chất giảm dần lúc ban đầu và sau đó tăng dần của chi phí biên tế?



HÌNH 7.4
Đường cong chi phí biên tế có dạng hình chữ U
Về mặt hình học, đường MC được mô tả trong hình 7.4 là một *parabol*. Về mặt toán học, một parabol được biểu thị bởi phương trình sau:

$$Y = \beta_0 + \beta_1 X + \beta_3 X^2 \qquad (7.11.1.)$$

được gọi là một hàm *bậc hai*, hay một cách tổng quát hơn, một *đa thức bậc hai* theo biến X-số mũ cao nhất của X biểu thị cho bậc đa thức (nếu cộng thêm X^3 vào trong hàm ở trên, nó sẽ là một đa thức bậc ba, v.v.)

Dạng ngẫu nhiên của (7.11.1) có thể được viết như sau

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i \qquad (7.11.2)$$

được gọi là một hồi qui *đa thức bậc hai*.

Hồi qui đa thức bậc k tổng quát có thể được viết thành

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \ldots\ldots \beta_k X_i^k + u_i \qquad (7.11.3)$$

Lưu ý rằng trong những loại hồi qui đa thức này chỉ có một biến giải thích ở bên vế phải nhưng nó xuất hiện với những lũy thừa khác nhau, như vậy khiến cho chúng trở thành những mô hình hồi qui bội. Nhân đây, lưu ý rằng nếu X_i được giả thiết là cố định hoặc không ngẫu nhiên, các số hạng lũy thừa của X_i cũng trở thành cố định hoặc không ngẫu nhiên.

Những mô hình này có gây ra vấn đề khó khăn đặc biệt nào về ước lượng không? Bởi vì đa thức bậc hai (7.11.2) hay đa thức bậc k (7.11.3) là tuyến tính theo các thông số β, chúng có thể được ước lượng bằng các OLS thông thường hay phương pháp ML. Nhưng còn về vấn đề cộng tuyến thì sao? Chẳng lẽ các X khác nhau này không có tương quan cao sao, bởi vì tất cả chúng đều là lũy thừa của X? Có, nhưng nhớ rằng các số hạng như X^2, X^3, X^4, v.v. đều là hàm không tuyến tính của X và vì vậy, nói một cách chặt chẽ, chúng không vi phạm giả định về phi đa cộng tuyến.[28] Tóm lại, các mô hình hồi qui đa thức có thể được ước lượng bằng các kỹ thuật đã được trình bày trong chương này và không gây ra một vấn đề mới nào về ước lượng.

Ví dụ 7.4: Ước lượng Hàm Tổng Chi phí

Để minh họa một ví dụ về hồi qui đa thức, hãy xem xét các dữ liệu trong Bảng 7.4 về sản lượng và tổng chi phí sản xuất ngắn hạn của một loại sản phẩm. Loại mô hình hồi qui nào sẽ thích hợp với các dữ liệu? Để thực hiện mục đích này, trước hết chúng ta hãy vẽ một đồ thị phân tán, như trong hình 7.5.

Bảng 7.4
Tổng chi phí (Y) và sản lượng (X)

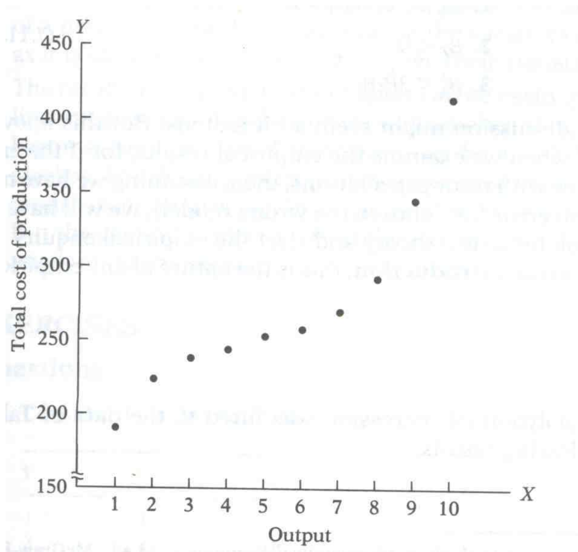
Sản lượng	Tổng chi phí
1	193
2	226
3	240
4	244
5	257
6	260
7	274
8	297
9	350
10	420

Từ hình này rõ ràng là quan hệ giữa tổng chi phí và sản lượng tương tự như một đường cong hình chữ S kéo dài; lưu ý đường tổng chi phí lúc đầu gia tăng từ từ và sau đó tăng nhanh, như quy luật nổi tiếng về *sinh lợi giảm dần* đã dự báo. Dạng hình chữ S này của đường tổng chi phí có thể được thể hiện bởi *đa thức bậc ba* sau:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i \qquad (7.11.4)$$

trong đó Y = tổng chi phí và X = sản lượng.

[28] Chúng ta sẽ xem vấn đề này một lần nữa ở Chương 10. Ở Chương 10 chúng ta sẽ bàn thảo một cách kỹ lưỡng tất cả các vấn đề về đa cộng tuyến.

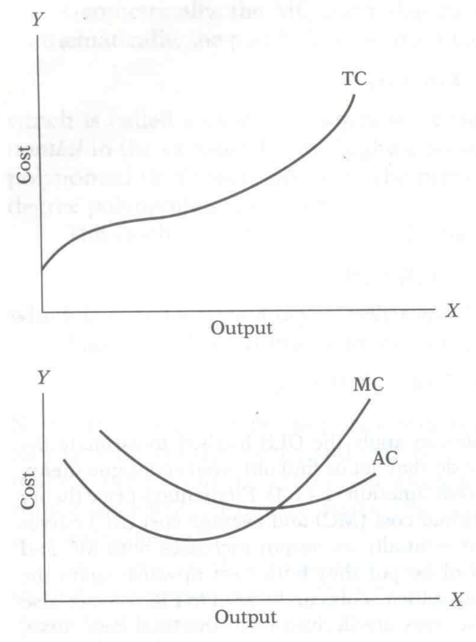


HÌNH 7.5
Đường tổng chi phí

Với các dữ liệu đã cho trong Bảng 7.4, chúng ta có thể áp dụng phương pháp OLS để ước lượng các thông số của (7.11.4). Nhưng trước khi tiến hành, chúng ta hãy tìm xem lý thuyết kinh tế nói gì về hàm chi phí bậc ba ngắn hạn (7.11.4) này. Lý thuyết về giá cơ sở cho thấy rằng các đường chi phí sản xuất biên tế ngắn hạn (MC) và chi phí sản xuất trung bình (AC) có dạng điển hình hình chữ U—ban đầu, khi sản lượng gia tăng cả MC và AC giảm, nhưng ở một mức sản lượng nhất định chúng sẽ quay lên trên, một lần nữa đây là kết quả của quy luật sinh lợi giảm dần. Có thể thấy điều này trong hình 7.6 (đồng thời xem hình 7.4). Và bởi vì các đường MC và AC là được suy ra từ đường tổng chi phí, bản chất hình chữ U của những đường này đặt một số hạn chế lên các thông số của đường tổng chi phí (7.11.4). Sự thật là có thể cho thấy rằng các các thông số của (7.11.4) phải thỏa mãn những hạn chế sau nếu chúng ta muốn quan sát các đường chi phí trung bình và chi phí biên tế ngắn hạn có dạng điển hình hình chữ U;[29]

1. β_0, β_1 và $\beta_3 > 0$
2. $\beta_2 < 0$
3. $\beta_2^2 < 3\beta_1\beta_3$ (7.11.5)

[29] Xem Alpha C.Chiang, *Fundamental Methods of Mathematical Economics (Các Phương pháp Căn bản của Kinh tế học Toán học)*, ấn bản lần 3, McGraw-Hill, New York, 1984, tr.250-252.



Hình 7.6
Các hàm chi phí ngắn hạn

Tất cả những bàn luận lý thuyết này có vẻ như hơi tẻ nhạt. Nhưng kiến thức này là vô cùng hữu dụng khi chúng ta xem xét các kết quả thực nghiệm, bởi vì nếu các kết quả thực nghiệm không phù hợp với các kỳ vọng tiên nghiệm, khi đó giả định rằng chúng ta chưa hề phạm một sai số đặc trưng nào (ví dụ như chọn mô hình sai), chúng ta sẽ phải sửa đổi lý thuyết của chúng ta hoặc phải đi tìm một lý thuyết mới và bắt đầu lại công cuộc truy tìm thực nghiệm lại từ đầu. Nhưng những chúng ta đã lưu ý trong **Phần Giới Thiệu**, đây là bản chất của bất kỳ một cuộc điều tra thực nghiệm nào.

Các Kết quả Thực nghiệm

Khi hồi qui đa thức bậc ba thích hợp với các dữ liệu của Bảng 7.4, chúng ta thu được những kết quả sau:

$$\begin{aligned} Y_i &= 141.7667 + 63.4776X_i - 12.9615X_i^2 + 0.9396X_i^3 \\ &\quad\ (6.3753) \qquad (4.7787) \qquad (0.9857) \qquad (0.0591) \\ &\quad\ R^2 = 0.9983 \end{aligned} \tag{7.11.6}$$

(Lưu ý: Các con số trong ngoặc là các sai số chuẩn ước lượng.) Mặc dù chúng ta sẽ xem xét ý nghĩa thống kê của những kết quả này trong chương sau, các bạn đọc có thể kiểm chứng rằng chúng tuân theo đúng với những kỳ vọng lý thuyết đã liệt kê trong (7.11.5). Chúng tôi để vấn đề này lại cho người đọc coi như là một bài tập để các bạn giải thích hồi qui (7.11.6).

7.12 TÓM TẮT VÀ KẾT KUẬN

1. Chương này giới thiệu mô hình hồi qui tuyến tính bội đơn giản nhất có thể có, tức mô hình hồi qui ba biến. Hiểu rằng thuật ngữ *tuyến tính* có nghĩa là tuyến tính theo thông số và không nhất thiết theo các biến.
2. Mặc dù mô hình hồi qui ba biến trên nhiều phương diện là sự mở rộng của mô hình hai biến, nhưng vẫn có nhiều khái niệm mới có liên quan,ví dụ như *hệ số hồi qui riêng phần, hệ số tương quan riêng phần, hệ số tương quan bội, R^2 có hiệu chỉnh và không hiệu chỉnh (đối với các bậc tự do), đa cộng tuyến,* và *thiên lệch đặc trưng.*
3. Chương 7 này còn xem xét dạng hàm số của mô hình hồi qui bội, ví dụ như *hàm sản xuất Cobb-Douglas* và *mô hình hồi qui đa thức*.
4. Mặc dù R^2 và R^2 có hiệu chỉnh là những đại lượng tổng quát cho biết một mô hình đã được chọn lựa thích hợp như thế nào với một tập hợp dữ liệu đã cho, chúng ta không nên nhấn mạnh quá mức tầm quan trọng của chúng. Điều quan trọng là các kỳ vọng lý thuyết nền tảng về mô hình theo các dấu tiên nghiệm của các hệ số của các biến tham gia vào trong mô hình và, như sẽ được trình bày trong chương kế tiếp, ý nghĩa thống kê của chúng.
5. Các kết quả đã trình bày trong chương này có thể dễ dàng được khái quát hóa cho một mô hình hồi qui tuyến tính bội với một lượng biến hồi qui độc lập bất kỳ. Nhưng các phép toán đại số trở nên vô cùng buồn tẻ. Có thể tránh khỏi sự buồn tẻ này bằng cách dùng đến đại số ma trận. Đối với các bạn đọc có quan tâm, việc mở rộng sang mô hình hồi qui k-biến dùng đại số ma trận sẽ được trình bày ở Chương 9, là chương tùy ý chọn lựa. Nhưng các bạn đọc nói chung có thể đọc các chương còn lại của tập sách này mà không cần phải biết gì nhiều về đại số ma trận.

BÀI TẬP

Câu hỏi

7.1. Xem xét các dữ liệu sau:

Y	X_2	X_3
1	1	2
3	2	1
8	3	-3

Dựa trên những dữ liệu này, ước lượng các hồi qui sau:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + u_{1i} \qquad (1)$$
$$Y_i = \lambda_1 + \lambda_2 X_{3i} + u_{2i} \qquad (2)$$
$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_{1i} \qquad (3)$$

Lưu ý: Chỉ ước lượng các hệ số, không ước lượng các sai số chuẩn.

a) $\alpha_2 = \beta_2$? Tại sao bằng hoặc không bằng?

b) $\lambda_3 = \beta_3$? Tại sao bằng hoặc không bằng?

Các Anh / Chị rút ra được những kết luận quan trọng gì từ bài tập này?

7.2. Từ những dữ liệu sau hãy ước lượng các hệ số hồi qui riêng phần, các sai số chuẩn của chúng và các giá trị R^2 có hiệu chỉnh và không hiệu chỉnh:

$$\bar{Y} = 367.693 \qquad \bar{X}_2 = 402.760 \qquad \bar{X}_3 = 8.0$$
$$\sum(Y_i - \bar{Y})^2 = 66042.269 \qquad \sum(X_{2i} - \bar{X}_2)^2 = 84855.096$$
$$\sum(X_{3i} - \bar{X}_3)^2 = 280.000 \qquad \sum(Y_i - \bar{Y})(X_{2i} - \bar{X}_2) = 74778.346$$
$$\sum(Y_i - \bar{Y})(X_{3i} - \bar{X}_3) = 4250.900 \qquad \sum(X_{2i} - \bar{X}_2)(X_{3i} - \bar{X}_3) = 4796.000$$

n = 15

7.3. Chứng minh rằng (7.9.1) và (7.9.2) là tương đương.

7.4. Chứng minh rằng $a_0 = 0$ trong phương trình (7.3.5). (Gợi ý: nhớ lại rằng tổng các phần dư bình phương tối thiểu bằng không, giả định rằng tung độ gốc có hiện diện trong mô hình.)

7.5. Chứng minh rằng (7.4.7) còn có thể được biểu diễn như sau

$$\hat{\beta}_2 = \frac{\sum y_i(x_{2i} - b_{23}x_{3i})}{\sum (x_{2i} - b_{23}x_{3i})^2}$$

= biến thiên đồng thời ròng (của x_3) giữa y và x_2

biến thiên ròng (của x_3) trong x_2

trong đó b_{23} là hệ số độ dốc trong hồi qui của X_2 theo X_3 như trong (7.3.2). (Gợi ý: Nhớ lại rằng $b_{23} = \sum x_{2i}x_{3i} / \sum x_{3i}^2$.)

7.6. Chứng minh rằng $r_{12.3}^2 = (R^2 - r_{13}^2)/(1 - r_{13}^2)$ và diễn giải phương trình này.

7.7. Nếu quan hệ $\alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 = 0$ là đúng đối với tất cả các giá trị của X_1, X_2 và X_3, tìm các giá trị của ba hệ số tương quan riêng phần này.

7.8. Liệu có thể có được những kết quả sau từ một tập hợp dữ liệu không?

a) $r_{23} = 0.9$, $r_{13} = -0.2$, $r_{12} = 0.8$

b) $r_{12} = 0.6$, $r_{23} = -0.9$, $r_{31} = -0.5$

c) $r_{21} = 0.01$, $r_{13} = 0.06$, $r_{23} = -0.7$

[*] **7.9.** Nếu $Z = aX + bY$ và $W = cX - dY$, và nếu hệ số tương quan giữa X và Y là r nhưng Z và Y không tương quan, chứng minh rằng $\sigma_z \sigma_w = (a^2 + b^2)\sigma_x \sigma_y (1 - r^2)^{1/2}$, trong đó $\sigma_z, \sigma_w, \sigma_x$ và σ_y là độ lệch chuẩn của bốn biến và trong đó *a, b, c* và *d* là các hằng số.

7.10. Chứng minh rằng β_2 và β_3 trong (7.10.2) thật sự cho biết độ co dãn theo sản lượng của lao động và vốn. (Có thể trả lời câu hỏi này được mà không dùng tới toán cao cấp; chỉ cần nhớ định nghĩa của hệ số co dãn và nhớ rằng thay đổi trong logarít của một biến là một thay đổi tương đối, giả định rằng những thay đổi là khá nhỏ.)

7.11. Nếu $X_3 = a_1 X_1 + a_2 X_2$, trong đó a_1 và a_2 những hằng số, chứng minh rằng giá trị bằng số của ba tương quan riêng phần này là bằng 1, $r_{13.2}$ mang dấu của a_1, và $r_{23.1}$ mang dấu của a_2, và $r_{13.2}$ trái dấu với a_1/a_2.

7.12. Một cách tổng quát, $R^2 \neq r_{12}^2 + r_{13}^2$, nhưng điều này chỉ đúng khi nếu $r_{23} = 0$. Cho nhận xét và chỉ ra ý nghĩa của phát hiện này. [Gợi ý: Xem phương trình (7.9.6).]

[*] Tùy ý lựa chọn

7.13. Trong điều kiện nào thì $\beta_2 = b_{1\,2}$ trong đó $b_{1\,2}$ là hệ số độ dốc trong hồi qui của Y theo X_2 chỉ như trong trường hợp được trình bày ở (7.7.1)?

7.14. Xem xét những mô hình sau.[#]

Mô hình A: $Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + u_{1t}$

Mô hình B: $(Y_t - X_{2t}) = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_{2t}$

a) Các ước lượng OLS của α_1 và β_1 có giống nhau không? Tại sao?
b) Các ước lượng OLS của α_3 và β_3 có giống nhau không? Tại sao?
c) Giữa α và β có quan hệ gì?
d) Các bạn có thể so sánh các số hạng R^2 của hai mô hình không? Tại sao có hoặc tại sao không?

7.15. Giả sử các bạn ước lượng hàm sản xuất[©]

$$Y_i = \alpha_1 + \alpha_2 X_i + u_{1i}$$

và hàm tiết kiệm

$$Z_i = \beta_1 + \beta_2 X_i + u_{2i}$$

trong đó Y = tiêu dùng, Z = tiết kiệm, X = thu nhập, và X = Y + Z, có nghĩa là thu nhập bằng với tiêu dùng cộng với tiết kiệm.

a) Mối quan hệ giữa α_2 và β_2, nếu có, là mối quan hệ gì? Trình bày các tính toán của các bạn.
b) Tổng các bình phương phần dư, RSS, của hai mô hình có bằng nhau không? Giải thích.
c) Các bạn có thể so sánh các số hạng R^2 của hai mô hình hay không? Tại sao có hoặc tại sao không?

7.16. Giả sử các bạn biểu diễn mô hình Cobb-Douglas đã cho trong (7.10.1) như sau:

$$Y_i = \beta_1 X_{2i}^{\beta} X_{3i}^{\beta} \hat{u}_i$$

Nếu các bạn biến đổi logarít của mô hình này, các bạn sẽ có lnu_i là số hạng nhiễu bên vế bên phải.

a) Các bạn phải đặt ra những giả định thống kê nào về ln u_i để có thể áp dụng mô hình hồi qui tuyến tính chuẩn cổ điển(CNLRM)? Làm cách nào để có thể kiểm định điều này với các dữ liệu cho trong Bảng 7.3?
b) Các giả định giống nhau có áp dụng được cho u_i không? Tại sao có hoặc tại sao không?

7.17. *Hồi qui qua mô hình gốc tọa độ.* Xem xét hồi qui sau đây qua gốc tọa độ:

$$Y_i = \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i$$

a) Làm cách nào chúng ta ước lượng được các đại lượng chưa biết?
b) Trong mô hình này $\sum \hat{u}_i$ có bằng zero không? Tại sao có hoặc tại sao không?
c) Trong mô hình này $\sum \hat{u}_i X_{2i} = \sum \hat{u}_i X_{3i} = 0$?
d) Khi nào thì chúng ta sẽ sử dụng một mô hình như vậy?
e) Các bạn có thể khái quát hóa các kết quả của mình cho mô hình k-biến không?

[#] Có sửa đổi từ Wojciech W. Charemza và Derek F. Deadman, *Econometric Practice (Thực hành Kinh tế lượng):* Từ lập mô hình tổng quát đến cụ thể, Đồng tích phân và Tự hồi quy vectơ, Edward Elgar, Brookfield, Vermont, 1992, tr.18.

[©] Có sửa đổi từ Peter Kennedy, *A Guide to Econometrics, (Hướng dẫn Kinh tế lượng),* ấn bản lần 3, The MIT Press, Cambridge, Massachusetts, 1992,tr.308, Bài tập #9.

(Gợi ý: Làm theo như phần thảo luận về trường hợp hai-biến trong chương 6.)

Bài tập.

7.18. Bảng dưới đây cho ta dữ liệu về tổng sản lượng thực, nhập lượng lao động và nhập lượng vốn thực trong khu vực sản xuất công nghiệp của Đài Loan.

Năm	**Tổng sản lượng thực (triệu NT $)*, Y**	**Ngày lao động (tính trên 1000 người), X_2**	**Nhập lượng vốn thực (triệu NT $), X_3**
1958	8911.4	281.5	120,753
1959	10,873.2	284.4	122,242
1960	11,132.5	289.0	125,263
1961	12,086.5	375.8	128,539
1962	12,767.5	375.2	131,427
1963	16,347.1	402.5	134,267
1964	19,542.7	478.0	139,038
1965	21,075.9	553.4	146,450
1966	23,052.0	616.7	153,714
1967	26,128.2	695.7	164,783
1968	29,563.7	790.3	176,864
1969	33,376.6	816.0	188,146
1970	38,354.3	848.4	205,841
1971	46,868.3	873.1	221,748
1972	54,308.0	999.2	239,715

*Đôla Đài Loan mới

Nguồn: Thomas Pei-Fan Chen. "Economic growth and Structural change in Taiwan (Tăng trưởng Kinh tế và Thay đổi Cấu trúc ở Đài Loan)—1952-1972, A Production Function Approach,(Một phương pháp hàm sản xuất)", luận văn tiến sĩ không xuất bản, Khoa Kinh tế, Graduate Center; City University of New York, June 1976, Bảng II.

a) Ap dụng các mô hình sau để làm thích hợp với các dữ liệu ở trên:

$$Y_t = \beta_1 + \hat{\beta}_2 X_{2t} + \hat{\beta}_3 X_{3t} + u_t$$
$$\ln Y_t = \alpha_1 + \alpha_2 \ln X_{2t} + \alpha_3 X_{3t} + u_t$$

b) Mô hình nào cho ta độ thích hợp tốt hơn và tại sao?

c) Đối với mô hình hồi qui tuyến tính - logarít, α_2 và α_3 cho biết các độ co dãn tương ứng theo sản lượng của lao động và vốn. Làm cách nào để tính được các độ co dãn tương tự cho mô hình tuyến tính?

d) Làm cách nào để so sánh các giá trị R^2 của hai mô hình? (Trình bày các tính toán của các bạn.)

e) Các kết quả của khu vực sản xuất công nghiệp khác với các kết quả của khu vực nông nghiệp đã cho trong Bảng 7.3 như thế nào?

f) Những giả thiết nào đã được giả định về số hạng nhiễu trong mô hình tuyến tính - logarít? Làm cách nào để kiểm định những giả thiết này?

7.19. Xem lại các dữ liệu của Vương quốc Anh về thay đổi phần trăm của lương và tỉ lệ thất nghiệp đã cho trong bảng 6.4. Sử dụng những dữ liệu này, xem xét xem các dạng biến đổi sau của đường cong Phillips có độ thích với hợp các dữ liệu của Vương quốc Anh không:

$$Y_i = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t}^2 + u_t$$

trong đó Y = thay đổi phần trăm hàng năm trong tỉ lệ tiền lương và X = tỉ lệ thất nghiệp.

a) Giải thích các kết quả của bạn.
b) Cơ sở để đưa bình phương tỉ lệ thất nghiệp vào mô hình này là gì? Một tiên nghiệm, bạn có kỳ vọng rằng β_3 sẽ là âm hay dương?
c) Mô hình này có thật sự phải là một mô hình hồi qui bội không bởi vì chỉ có một biến giải thích, tức tỉ lệ thất nghiệp, có mặt trong mô hình?
d) Các kết quả của bạn so sánh với các kết quả thu được trong (6.6.2) và trong bài tập 6.20 ra sao?
e) Bạn có thể so sánh các số hạng R^2 của các mô hình khác nhau không? Tại sao?
f) Bạn sẽ chọn mô hình nào: mô hình bậc hai ở đây, mô hình nghịch đảo trong (6.6.2) hay mô hình tuyến tính trong bài tập 6.20? Bạn sử dụng những tiêu chuẩn nào?

7.20. *Nhu cầu về hoa hồng.*[*] Bảng dưới đây cho dữ liệu hàng quý về những biến sau:
Y = số lượng hoa hồng bán được, theo tá
X_2 = giá bán sỉ trung bình của hoa hồng , $/tá
X_3 = giá bán sỉ trung bình của hoa cẩm chướng, $/tá
X_4 = thu nhập khả dụng trung bình hàng tuần của gia đình, $/tuần
X_5 = biến xu hướng có các giá trị 1, 2 và tiếp tục, trong giai đoạn 1971-III cho đến 1975-II tại khu vực trung tâm Detroit.

Năm và quý	**Y**	**X_2**	**X_3**	**X_4**	**X_5**
1971 - III	11,484	2.26	3.49	158.11	1
-IV	9,348	2.54	2.85	173.36	2
1972 - I	8,429	3.07	4.06	165.26	3
- II	10,079	2.91	3.64	172.92	4
- III	9,240	2.73	3.21	178.46	5
- IV	8,862	2.77	3.66	198.62	6
1973 - I	6,216	3.59	3.76	186.28	7
- II	8,253	3.23	3.49	188.98	8
- III	8,038	2.60	3.13	180.49	9
- IV	7,476	2.89	3.20	183.33	10
1974 - I	5,911	3.77	3.65	181.87	11
- II	7,950	3.64	3.60	185.00	12
- III	6,134	2.82	2.94	184.00	13
- IV	5,868	2.96	3.12	188.20	14
1975 - I	3,160	4.24	3.58	175.67	15
- II	5,872	3.69	3.53	188.00	16

Bạn được yêu cầu xem xét các hàm nhu cầu sau:

[*] Tôi biết ơn Joe Walsh vì đã thu thập những dữ liệu này từ một nhà buôn sỉ lớn ở khu vực trung tâm Detroit và tiếp đó đã xử lý những dữ liệu này.

$$Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \alpha_4 X_{4t} + \alpha_5 X_{5t} + u_t$$
$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 X_{3t} + \beta_4 \ln X_{4t} + \beta_5 X_{5t} + u_t$$

a) Ước lượng các thông số của mô hình tuyến tính và giải thích các kết quả.
b) Ước lượng các thông số của mô hình tuyến tính - logarít và giải thích các kết quả.
c) β_2, β_3 và β_4 lần lượt cho biết *giá riêng*, *giá chéo*, và *độ co dãn theo thu nhập* của cầu. Dấu tiên nghiệm của các hệ số là gì ? Các kết quả có giống như các kỳ vọng tiên nghiệm không?
d) Làm cách nào để tính giá riêng, giá chéo, và độ co dãn thu nhập cho mô hình tuyến tính này?
e) Dựa trên các phân tích của bạn, bạn sẽ chọn mô hình nào và cho biết tại sao?

7.21. *Hoạt động khoan thăm dò dầu*. Các giếng khoan thăm dò dầu là các giếng được khoan nhằm tìm và sản xuất dầu lửa và/hay khí đốt trong một khu vực được nâng cấp hay nhằm mở rộng phạm vi của một nguồn dự trữ khí đốt hay dầu lửa đã biết: Bảng sau cho ta các dữ liệu về những biến này:*

* Tôi mang ơn Raymond Savino vì đã thu thập và xử lý những dữ liệu này.

Ngàn giếng khoan thăm dò dầu, (Y)	**Giá mỗi thùng, hằng số \$, (X_2)**	**Sản lượng quốc nội (triệu thùng mỗi ngày), (X_3)**	**GNP, hằng số Tỉ \$, (X_4)**	**Thời điểm (X_5)**
8.01	4.89	5.52	487.67	1948 = 1
9.06	4.83	5.05	490.59	1949 = 2
10.31	4.68	5.41	533.55	1950 = 3
11.76	4.42	6.16	576.57	1951 = 4
12.43	4.36	6.26	598.62	1952 = 5
13.31	4.55	6.34	621.77	1953 = 6
13.10	4.66	6.81	613.67	1954 = 7
14.94	4.54	7.15	654.80	1955 = 8
16.17	4.44	7.17	668.84	1956 = 9
14.71	4.75	6.71	681.02	1957 = 10
13.20	4.56	7.05	679.53	1958 = 11
13.19	4.29	7.04	720.53	1959 = 12
11.70	4.19	7.18	736.86	1960 = 13
10.99	4.17	7.33	755.34	1961 = 14
10.80	4.11	7.54	799.15	1962 = 15
10.66	4.04	7.61	830.70	1963 = 16
10.75	3.96	7.80	874.29	1964 = 17
9.47	3.85	8.30	925.86	1965 = 18
10.31	3.75	8.81	980.98	1966 = 19
8.88	3.69	8.66	1,007.72	1967 = 20
8.88	3.56	8.78	1,051.83	1968 = 21
9.70	3.56	9.18	1,078.76	1969 = 22
7.69	3.48	9.03	1,075.31	1970 = 23
6.92	3.53	9.00	1,107.48	1971 = 24
7.54	3.39	8.78	1,171.10	1972 = 25
7.47	3.68	8.38	1,234.97	1973 = 26
8.63	5.92	8.01	1,217.81	1974 = 27
9.21	6.03	7.78	1,202.36	1975 = 28
9.23	6.12	7.88	1,271.01	1976 = 29
9.96	6.05	7.88	1,332.67	1977 = 30
10.78	5.89	8.67	1,385.10	1978 = 31

Nguồn: Bộ Thông tin Năng lượng, Báo cáo 1978cho Quốc hội.

Y = số lượng các giếng khoan thăm dò
X_2 = chi phí trên mỗi nguồn khoan trong giai đoạn trước
(tính bằng giá đôla không đổi, 1972 = 100)
X_3 = sản lượng quốc nội
X_4 = GNP đôla không đổi (1972 = 100)
X_5 = biến xu hướng, 1984 = 1, 1949 = 2, ... , 1978 = 31

Xem xem mô hình sau có thích hợp với các dữ liệu không:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 \ln X_{3t} + \beta_4 X_{4t} + u_t$$

a) Bạn có thể đưa ra một cơ sở hợp lý tiên nghiệm cho mô hình này không?

b) Giả thiết rằng mô hình này chấp nhận được, hãy ước lượng các thông số của mô hình và các sai số chuẩn của chúng, và tính R^2 và $\bar{R}^2$.
c) Nhận xét các kết quả thu được theo những kỳ vọng tiên nghiệm.
d) Bạn sẽ có đề nghị đặc trưng nào khác không để giải thích hoạt động khoan thăm dò dầu? Tại sao?

7.22. *Chi tiêu ngân sách quốc phòng Mỹ,1962-1981.* Để giải thích ngân sách quốc phòng Mỹ, người ta yêu cầu bạn xem xét mô hình sau:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 \ln X_{3t} + \beta_4 X_{4t} + u_t$$

trong đó Y_t = Chi tiêu ngân sách quốc phòng của năm *t*, $/tỉ

X_{2t} = GNP của năm *t*, $/tỉ

X_{3t} = doanh số bán hàng quân sự Mỹ/viện trợ trong năm *t*, $/tỉ

X_{4t} = doanh số công nghiệp hàng không/viện trợ trong năm *t*, $/tỉ

X_{5t} = các xung đột quân sự phải sử dụng trên 100, 000 lính. Biến này có giá trị bằng 1 khi phải huy động 100, 000 lính hay nhiều hơn nhưng bằng 0 khi số lượng lính dưới 100,000.

Để kiểm định mô hình này, các bạn có các dữ liệu sau:

a) Ước lượng các thông số của mô hình này và các sai số chuẩn và tính R^2, R^2 có hiệu chỉnh và $\bar{R}^2$.
b) Nhận xét về các kết quả, lưu ý đến bất kỳ kỳ vọng tiên nghiệm nào của bạn về quan hệ giữa Y và các biến X khác nhau.
c) Bạn có thể muốn đưa thêm (những) biến nào khác vào trong mô hình và cho biết tại sao?

Năm	Chi tiêu ngân sách quốc phòng, Y	GNP, X_2	Doanh số bán hàng quân sự Mỹ / viện trợ, X_3	Doanh số công nghiệp hàng không, X_4	Các cuộc xung đột 100,000 +, X_5
1962	51.1	560.3	0.6	16.0	0
1963	52.3	590.5	0.9	16.4	0
1964	53.6	632.4	1.1	16.7	0
1965	49.6	684.9	1.4	17.0	1
1966	56.8	749.9	1.6	20.2	1
1967	70.1	793.9	1.0	23.4	1
1968	80.5	865.0	0.8	25.6	1
1969	81.2	931.4	1.5	24.6	1
1970	80.3	992.7	1.0	24.8	1
1971	77.7	1,077.6	1.5	21.7	1
1972	78.3	1,185.9	2.95	21.5	1
1973	74.5	1,326.4	4.8	24.3	0
1974	77.8	1,434.2	10.3	26.8	0
1975	85.6	1,549.2	16.0	29.5	0
1976	89.4	1,718.0	14.7	30.4	0
1977	97.5	1,918.3	8.3	33.3	0
1978	105.2	2,163.9	11.0	38.0	0
1979	117.7	2,417.8	13.0	46.2	0
1980	135.9	2,633.1	15.3	57.6	0
1981	162.1	2,937.7	18.0	68.9	0

Nguồn: Các dữ liệu được Albert Lucchino thu thập từ các ấn bản khác nhau của chính phủ.

7.23. *Nhu cầu về thịt gà ở Mỹ, 1960-1982.* Để nghiên cứu về tiêu thụ thịt gà bình quân đầu người ở Mỹ, bạn có được những dữ liệu sau:

trong đó Y = tiêu thụ thịt gà bình quân đầu người, cân Anh

X_2 = thu nhập khả dụng thực bình quân đầu người, đôla

X_3 = giá bán lẻ thực thịt gà tính trên mỗi cân Anh, xu

X_4 = giá bán lẻ thực thịt heo tính trên mỗi cân Anh, xu

X_5 = giá bán lẻ thực thịt bò tính trên mỗi cân Anh, xu

X_6 = giá thực đa hợp của các sản phẩm thay thế cho thịt gà tính trên mỗi cân Anh, xu, đây là trung bình có trọng số của các giá bán lẻ thực trên một pao của thịt heo và thịt bò, các trọng lượng là mức tiêu thụ thịt bò và thịt heo tương đối tính trong tổng tiêu thụ thịt heo và thịt bò.

Bây giờ hãy xem xét các hàm nhu cầu sau đây:

$$\ln Y_t = \alpha_1 + \alpha_2 \ln X_{2t} + \alpha_3 \ln X_{3t} + u_t \quad (1)$$

$$\ln Y_t = \gamma_1 + \gamma_2 \ln X_{2t} + \gamma_3 \ln X_{3t} + \gamma_4 \ln X_{4t} + u_t \quad (2)$$

$$\ln Y_t = \lambda_1 + \lambda_2 \ln X_{2t} + \lambda_3 \ln X_{3t} + \lambda_4 \ln X_{5t} + u_t \quad (3)$$

$$\ln Y_t = \theta_1 + \theta_2 \ln X_{2t} + \theta_3 \ln X_{3t} + \theta_4 \ln X_{4t} + \theta_5 \ln X_{5t} + u_t \quad (4)$$

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_4 \ln X_{6t} + u_t \quad (5)$$

Năm	**Y**	**X_2**	**X_3**	**X_4**	**X_5**	**X_6**
1960	27.8	397.5	42.2	50.7	78.3	65.8
1961	29.9	413.3	38.1	52.0	79.2	66.9
1962	29.8	439.2	40.3	54.0	79.2	67.8
1963	30.8	459.7	39.5	55.3	79.2	69.6
1964	31.2	492.9	37.3	54.7	77.4	68.7
1965	33.3	528.6	38.1	63.7	80.2	73.6
1966	35.6	560.3	39.3	69.8	80.4	76.3
1967	36.4	624.6	37.8	65.9	83.9	77.2
1968	36.7	666.4	38.4	64.5	85.5	78.1
1969	38.4	717.8	40.1	70.0	93.7	84.7
1970	40.4	768.2	38.6	73.2	106.1	93.3
1971	40.3	843.3	39.8	67.8	104.8	89.7
1972	41.8	911.6	39.7	79.1	114.0	100.7
1973	40.4	931.1	52.1	95.4	124.1	113.5
1974	40.7	1,021.5	48.9	94.2	127.6	115.3
1975	40.1	1,165.9	58.3	123.5	142.9	136.7
1976	42.7	1,349.6	57.9	129.9	143.6	139.2
1977	44.1	1,449.4	56.5	117.6	139.2	132.0
1978	46.7	1,575.5	63.7	130.9	165.5	132.1
1979	50.6	1,759.1	61.6	129.8	203.3	154.4
1980	50.1	1,994.2	58.9	128.0	219.6	174.9
1981	51.7	2,258.1	66.4	141.0	221.6	180.8
1982	52.9	2,478.7	70.4	168.2	232.6	189.4

Nguồn: Dữ liệu về Y được lấy rút ra từ Citibase và dữ liệu về X_2 cho đến X_6 là từ Bộ nông nghiệp Mỹ. Tôi mang ơn Robert J.Fisher vì đã thu thập những dữ liệu này và những phân tích thống kê.

Từ lý thuyết kinh tế vi mô ta biết được rằng nhu cầu về một hàng hóa thường phụ thuộc vào thu nhập thực của người tiêu dùng, giá cả thực của hàng hóa, và giá cả thực của các hàng hóa cạnh tranh hay hỗ trợ. Chú ý những điều cân nhắc này và trả lời các câu hỏi sau

a) Các bạn sẽ chọn hàm nhu cầu nào trong số những hàm cho ở trên, và cho biết tại sao?
b) Các bạn giải thích các hệ số của $\ln X_{2t}$ và $\ln X_{3t}$ trong những mô hình này như thế nào?
c) Các đặc trưng trong (2) và (4) khác nhau ở chỗ nào?
d) Nếu chọn đặc trưng (4) các bạn có dự đoán được trước những vấn đề gì? (Gợi ý: Giá của cả thịt heo và thịt bò tính theo giá thịt gà.)
e) Bởi vì đặc trưng (5) bao gồm giá đa hợp của thịt heo và thịt bò, các bạn có nghĩ rằng hàm nhu cầu (5) tốt hơn (4) không? Tại sao?

f) Thịt heo và/hay thịt bò có phải là các sản phẩm cạnh tranh hay hỗ trợ của thịt gà không? Làm sao các bạn biết?
g) Giả thiết rằng hàm (5) là hàm nhu cầu "đúng". Hãy ước lượng các thông số của mô hình này, tính các sai số chuẩn của chúng, R^2, $\bar{R}^2$ và R^2 có hiệu chỉnh. Giải thích các kết quả này.
h) Bây giờ giả sử bạn tiến hành mô hình (2) "không đúng". Hãy đánh giá các hậu quả của việc xác định đặc trưng sai này bằng cách xem xét các giá trị γ_2 và γ_3 so với β_2 và β_3 tương ứng. (Gợi ý: Chú ý đến phần thảo luận trong Phần 7.7).

7.24. Trong một cuộc nghiên cứu về sự luân chuyển công nhân trên thị trường lao động, James F. Ragan, Jr., thu được những kết quả sau cho nền kinh tế Mỹ trong giai đoạn 1950-I cho đến 1979-IV.* (Các con số trong ngoặc là các trị thống kê *t* ước lượng.)

$$\begin{aligned}\ln Y_t = \underset{(4.28)}{4.47} - \underset{(-5.31)}{0.34} \ln X_{2t} + \underset{(3.46)}{1.22} \ln X_{3t} + \underset{(3.10)}{1.22} \ln X_{4t}\\ + \underset{(1.10)}{0.80} \ln X_{5t} - \underset{(-3.09)}{0.0054} \ln X_{6t} \qquad \bar{R}^2 = 0.5370\end{aligned}$$

Lưu ý: Chúng ta sẽ thảo luận các trị thống kê *t* trong chương tới.

Trong đó Y = tỉ lệ nghỉ việc trong khu vực sản xuất, được định nghĩa là số lượng người tự rời bỏ công việc tính trên 100 công nhân
X_2 = một biến công cụ hay thay thế cho tỉ lệ thất nghiệp của nam đã trưởng thành
X_3 = phần trăm công nhân trẻ hơn 25 tuổi
$X_4 = N_{t-1}/N_{t-4}$ = tỉ lệ thâu dụng nhân công trong quý (t-1) so với quý (t-4)
X_5 = phần trăm nữ công nhân
X_6 = xu hướng theo thời gian (1950-I = 1)

a) Giải thích các kết quả ở trên.
b) Quan hệ nghịch biến quan sát được giữa các logarít của Y và X_2 có là một tiên nghiệm đúng đắn không?
c) Tại sao hệ số của ln X_3 dương?
d) Bởi vì hệ số xu hướng là số âm, tỉ lệ nghỉ việc giảm dài hạn là bằng bao nhiêu và tại sao lại có sự suy giảm này?
e) $\bar{R}^2$ có "quá" thấp không?
f) Bạn có thể ước lượng các sai số chuẩn của các hệ số hồi qui từ các dữ liệu đã cho không? Tại sao có hoặc tại sao không?

7.25. Xem xét hàm nhu cầu đơn cho tiền tệ sau:

$$M_t = \beta_0 Y_t^{\beta_1} r_t^{\beta_2} e^{u_t}$$

trong đó M_t = tổng cán cân tiền mặt thực tại thời điểm *t*
Y_t = tổng thu nhập quốc gia thực tại thời điểm *t*
r_t = mức lãi suất dài hạn

* *Nguồn:* Xem bài viết của Ragan, "Tỉ lệ thay đổi công nhân trên thị trường lao động: Nghiên cứu về Tỉ lệ thay và tiết kiệm thải công nhân," Báo *Economic Review*, Ngân hàng Dự trữ Liên Bang của thành phố Kansas, tháng Năm 1981, tr.13-22.

a) Với các dữ liệu sau, ước lượng độ co dãn của tổng cán cân tiền mặt thực so với tổng thu nhập thực và mức lãi suất dài hạn.
b) Nếu thay vì áp dụng hàm nhu cầu ở trên, các bạn phải áp dụng mô hình $(M/Y)_t = \alpha\beta_t^{\beta}$, các bạn sẽ giải thích các kết quả như thế nào? Trình bày các phép tính cần thiết.

Lưu ý: Để chuyển đổi các đại lượng danh nghĩa thành các đại lượng thực, chia các đại lượng danh nghĩa cho hệ số giảm phát giá ẩn.

Năm	Tiền danh nghĩa, crores or rupees	Thu nhập ròng danh nghĩa, trên 100 crores of rupees	Hệ số giảm phát giá ẩn	Mức lãi suất dài hạn, %
1948 - 1949	1,898.69	86.5	100.00	3.03
1949 - 1950	1,880.29	90.1	102.15	3.07
1950 - 1951	1,979.49	95.3	107.68	3.15
1951 - 1952	1,803.79	99.7	109.56	3.41
1952 - 1953	1,764.71	98.2	103.81	3.66
1953 - 1954	1,793.97	104.8	104.49	3.64
1954 - 1955	1,920.63	96.1	93.48	3.70
1955 - 1956	2,216.95	99.8	95.23	3.74
1956 - 1957	2,341.89	113.1	102.82	3.99
1957 - 1958	2,413.16	113.9	104.59	4.18
1958 - 1959	2,526.02	126.9	108.15	4.13
1959 - 1960	2,720.22	129.5	109.19	4.05
1960 - 1961	2,868.61	141.4	111.19	4.06
1961 - 1962	3,045.82	148.0	113.32	4.16
1962 - 1963	3,309.98	154.0	115.70	4.49
1963 - 1964	3,752.12	172.1	123.19	4.66
1964 - 1965	4,080.06	200.1	132.96	4.80

PHỤ LỤC 7A

7A.1 NGUỒN GỐC CỦA CÁC HÀM ƯỚC LƯỢNG CHO TRONG CÁC PHƯƠNG TRÌNH (7.4.3) CHO ĐẾN (7.4.5)

Lấy đạo hàm riêng phần của phương trình

$$\sum \hat{u}_i^2 = \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}\right)^2 \tag{7.4.2}$$

theo 3 biến chưa biết và cho phương trình thu được về không, ta có

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_1} = 2\sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}\right)(-1) = 0$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} = 2\sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}\right)(-X_{2i}) = 0$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_3} = 2\sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}\right)(-X_{3i}) = 0$$

Rút gọn những phương trình này lại, chúng ta thu được các phương trình (7.4.3) cho đến (7.4.5).

Nhân tiện, lưu ý rằng ba phương trình trên cũng có thể được viết thành

$$\sum \hat{u}_i = 0$$

$$\sum \hat{u}_i X_{2i} = 0 \quad \text{(Tại sao?)}$$

$$\sum \hat{u}_i X_{3i} = 0$$

cho thấy rằng các đặc tính của bình phương tối thiểu là thích hợp, tức tổng các phần dư bằng không, và chúng không tương quan với các biến giải thích X_2 và X_3.

Nhân đây, lưu ý rằng để có được các hàm ước lượng OLS của mô hình hồi qui tuyến tính k-biến (7.4.20) chúng ta tiến hành tương tự. Như vậy, đầu tiên chúng ta viết

$$\sum \hat{u}_i^2 = \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki}\right)^2$$

Lấy đạo hàm riêng phần của biểu thức này theo mỗi biến trong *k* biến chưa biết, cho các phương trình thu được bằng không, và chuyển vế các số hạng, chúng ta thu được *k* phương trình chuẩn sau theo *k* biến chưa biết:

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_{2i} + \hat{\beta}_3 \sum X_{3i} + \cdots + \hat{\beta}_k \sum X_{ki}$$

$$\sum Y_i X_{2i} = \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i} X_{3i} + \cdots + \hat{\beta}_k \sum X_{2i} X_{ki}$$

$$\sum Y_i X_{3i} = \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{2i} X_{3i} + \hat{\beta}_3 \sum X_{3i}^2 + \cdots + \hat{\beta}_k \sum X_{2i} X_{ki}$$

. . . .

$$\sum Y_i X_{ki} = \hat{\beta}_1 \sum X_{ki} + \hat{\beta}_2 \sum X_{2i} X_{ki} + \hat{\beta}_3 \sum X_{3i} X_{ki} + \cdots + \hat{\beta}_k \sum X_{ki}^2$$

Hay, chuyển sang các mẫu tự viết thường, những phương trình này có thể được biểu diễn như sau

$$\sum y_i x_{2i} = \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum x_{2i} x_{3i} + \cdots + \hat{\beta}_k \sum x_{2i} x_{ki}$$

$$\sum y_i x_{3i} = \hat{\beta}_2 \sum x_{2i} x_{3i} + \hat{\beta}_3 \sum x_{3i}^2 + \cdots + \hat{\beta}_k \sum x_{3i} x_{ki}$$

. . . .

$$\sum y_i x_{ki} = \hat{\beta}_2 \sum x_{2i} x_{ki} + \hat{\beta}_3 \sum x_{3i} x_{ki} + \cdots + \hat{\beta}_k \sum x_{ki}^2$$

Cần lưu ý thêm nữa là mô hình k-biến cũng còn thỏa mãn những phương trình sau:

$$\sum \hat{u}_i = 0$$
$$\sum \hat{u}_i X_{2i} = \sum \hat{u}_i X_{3i} = \cdots + \sum \hat{u}_i X_{ki} = 0$$

7.2A ĐỒNG ĐẲNG THỨC GIỮA a_1 CỦA (7.3.5) VÀ β_2 CỦA(7.4.7)

Các hàm ước lượng OLS của a_1 là

$$a_1 = \frac{\sum (\hat{u}_{1i} - \bar{\hat{u}}_1)(\hat{u}_{2i} - \bar{\hat{u}}_2)}{\sum (\hat{u}_{2i} - \bar{\hat{u}}_2)^2}$$
$$= \frac{\sum \hat{u}_{1i}\hat{u}_{2i}}{\sum \hat{u}_{2i}^2} \quad \text{bởi vì } \bar{\hat{u}}_1 = \bar{\hat{u}}_2 = 0 \qquad \text{(Tại sao?)}$$

Bởi vì $\bar{\hat{u}}_1 = \bar{\hat{u}}_2 = 0$, phương trình (7.3.1) và (7.3.2) có thể được viết thành

$$y_i = b_{13} x_{3i} + \hat{u}_{1i}$$
$$x_{2i} = b_{23} x_{3i} + \hat{u}_{2i}$$

trong đó các mẫu tự viết thường, như thường lệ, biểu thị cho các độ lệch so với các giá trị trung bình.

Thay $\hat{u}_{1i}$ và $\hat{u}_{2i}$ trong phương trình trên vào phương trình của a_1, ta có

$$a_1 = \frac{\sum (y_i - b_{13} x_{13})(x_{2i} - b_{23} x_{3i})}{\sum \left(x_{2i} - b_{23} x_{3i}\right)^2}$$
$$= \frac{\sum y_i x_{2i} - b_{23} \sum y_i x_{3i} - b_{13} \sum x_{2i} x_{3i} + b_{13} b_{23} \sum x_{3i}^2}{\sum x_{2i}^2 + b_{23}^2 \sum x_{3i}^2 - 2b_{23} \sum x_{2i} x_{3i}}$$

Lưu ý rằng $b_{23} = \sum x_{2i} x_{3i} / \sum x_{3i}^2$ và $b_{13} = \sum y_i x_{3i} / \sum x_{3i}^2$, bạn đọc có thể dễ dàng kiểm chứng được rằng a_1 ở trên thật sự được rút gọn thành β_2 cho trong (7.4.7).

7A.3 NGUỒN GỐC CỦA PHƯƠNG TRÌNH 7.4.19

Nhớ lại rằng

$$\hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}$$

còn có thể được viết thành

$$\hat{u}_i = y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i}$$

trong đó các mẫu tự viết thường, như thường lệ, biểu thị cho các độ lệch so với các giá trị trung bình.

Bây giờ thì

$$\begin{aligned}\sum \hat{u}_i^2 &= \sum (\hat{u}_i \hat{u}_i) \\ &= \sum \hat{u}_i (y_i - \hat{\beta} x_{2i} - \hat{\beta}_3 x_{3i}) \\ &= \sum \hat{u}_i y_i\end{aligned}$$

trong đó chúng ta sử dụng tính chất là $\sum \hat{u}_i x_{2i} = \sum \hat{u}_i x_{3i} = 0$. (Tại sao?) Đồng thời

$$\sum \hat{u}_i y_i = \sum y_i \hat{u}_i = \sum y_i (y_i - \hat{\beta} x_{2i} - \hat{\beta}_3 x_{3i})$$

có nghĩa là,

$$\sum \hat{u}_i^2 = \sum y_i - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i}$$

đây chính là kết quả chúng ta mong muốn.

7.4A ƯỚC LƯỢNG THÍCH HỢP TỐI ĐA CỦA PHƯƠNG TRÌNH HỒI QUI BỘI

Mở rộng các ý tưởng đã giới thiệu trong Chương 4, Phụ lục 4A, chúng ta có thể viết hàm thích hợp-logarít cho mô hình hồi qui tuyến tính k-biến (7.4.29) thành

$$\ln L = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi - \frac{1}{2}\sum \frac{(Y_i - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki})^2}{\sigma^2}$$

Lấy đạo hàm riêng phần của hàm này theo β_1, β_2, ...β_k và σ^2, chúng ta thu được các phương trình (K+1) sau

$$\frac{\partial \ln L}{\partial \beta_1} = -\frac{1}{\sigma^2}\sum (Y_i - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki})(-1) \quad (1)$$

$$\frac{\partial \ln L}{\partial \beta_2} = -\frac{1}{\sigma^2}\sum (Y_i - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki})(-X_{2i}) \quad (2)$$

....

$$\frac{\partial \ln L}{\partial \beta_k} = -\frac{1}{\sigma^2}\sum (Y_i - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki})(-X_{ki}) \quad (K)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum (Y_i - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki})^2 \quad (K+1)$$

Cho các phương trình này bằng không (điều kiện bậc nhất cho tối ưu hóa) và đặt $\tilde{\beta}_1, \tilde{\beta}_2, \ldots.. \tilde{\beta}_k$ và $\tilde{\sigma}^2$ tượng trưng cho các hàm ước lượng ML, sau những phép biến đổi đại số đơn giản chúng ta thu được

$$\sum Y_i = n\tilde{\beta}_1 + \tilde{\beta}_2 \sum X_{2i} + \cdots + \tilde{\beta}_k \sum X_{ki}$$

$$\sum Y_i X_{2i} = \tilde{\beta}_1 \sum X_{2i} + \tilde{\beta}_2 \sum X_{2i}^2 + \cdots + \tilde{\beta}_k \sum X_{2i} X_{ki}$$

....

$$\sum Y_i X_{ki} = \tilde{\beta}_1 \sum X_{ki} + \tilde{\beta}_2 \sum X_{2i} X_{ki} + \cdots + \tilde{\beta}_k \sum X_{ki}^2$$

đây chính là những phương trình chuẩn của lý thuyết bình phương tối thiểu thông thường, như bạn có thể thấy từ Phụ lục 7A, Phần 7A.1. Do đó, các hàm ước lượng ML, các $\tilde{\beta}$, tương tự như các hàm ước lượng OLS, các $\hat{\beta}$, đã cho trước đây. Nhưng như đã lưu ý trong Chương 4, Phụ lục 4A sự bằng nhau này không phải là ngẫu nhiên.

Thay các hàm ước lượng ML (=OLS) vào phương trình thứ (K+1) vừa có ở trên, chúng ta thu được sau khi đã rút gọn, hàm ước lượng ML của σ^2 sau

$$\begin{aligned}\tilde{\sigma}^2 &= \frac{1}{n}\sum (Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_{2i} - \ldots - \tilde{\beta}_k X_{ki})^2 \\ &= \frac{1}{n}\sum \hat{u}_i^2\end{aligned}$$

Như đã được lưu ý trong bài đọc, hàm ước lượng này khác với hàm ước lượng OLS $\hat{\sigma}^2 = \sum \hat{u}_i^2 / (n-k)$. Và bởi vì hàm ước lượng OLS là một ước lượng không thiên lệch của σ^2, kết luận này có nghĩa là hàm ước lượng ML $\tilde{\sigma}^2$ là một hàm ước lượng thiên lệch. Nhưng như chúng ta có thể dễ dàng kiểm chứng một cách gần đúng, $\tilde{\sigma}^2$ cũng là không thiên lệch.

7A.5 CHỨNG MINH RẰNG $E(b_{12}) = \beta_2 + \beta_3 b_{32}$ (PHƯƠNG TRÌNH 7.7.4)

Dưới dạng độ lệch mô hình hồi qui tổng thể ba biến có thể được viết thành

$$y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + (u_i - \bar{u}) \tag{1}$$

Đầu tiên nhân với x_2 và sau đó nhân với x_3, các phương trình chuẩn thông thường trở thành

$$\sum y_i x_{2i} = \beta_2 \sum x_{2i}^2 + \beta_3 \sum x_{2i} x_{3i} + \sum x_{2i}(u_i - \bar{u}) \tag{2}$$

$$\sum y_i x_{3i} = \beta_2 \sum x_{2i} x_{3i} + \beta_3 \sum x_{3i}^2 + \sum x_{3i}(u_i - \bar{u}) \tag{3}$$

Chia (2) cho $\sum x_{2i}^2$ ở hai vế, chúng ta được

$$\frac{\sum y_i x_{2i}}{\sum x_{2i}^2} = \beta_2 + \beta_3 \frac{\sum x_{2i} x_{3i}}{\sum x_{2i}^2} + \frac{\sum x_{2i}(u_i - \bar{u})}{\sum x_{2i}^2} \tag{4}$$

Bây giờ nhớ rằng

$$b_{12} = \frac{\sum y_i x_{2i}}{\sum x_{2i}^2}$$

$$b_{32} = \frac{\sum x_{2i} x_{3i}}{\sum x_{2i}^2}$$

Phương trình (4) có thể được viết thành

$$b_{12} = \beta_2 + \beta_3 b_{32} + \frac{\sum x_{2i}(u_i - \bar{u})}{\sum x_{2i}^2} \tag{5}$$

Lấy giá trị kỳ vọng của (5) ở hai vế, cuối cùng ta được

$$E(b_{12}) = \beta_2 + \beta_3 b_{32} \tag{6}$$

trong đó chúng ta áp dụng những đặc tính (a) đối với một mẫu đã cho, b_{23} là một đại lượng cố định đã biết, (b) β_2 và β_3 là hằng số, và (c) u_i không tương quan với X_{2i} (cũng như với X_{3i}).

Không những b_{12} thiên lệch, mà phương sai của nó cũng có nhiều khả năng là thiên lệch. Có thể chứng minh điều này như sau. Theo định nghĩa,

$$\text{var}(b_{12}) = E[b_{12} - E(b_{12})]^2 \tag{7}$$

Thay (5) và (6) vào (7) và rút gọn, chúng ta có thể chứng minh rằng

$$\text{var}(b_{12}) = \frac{\sigma^2}{\sum x_{2i}^2} \tag{8}$$

trong khi đó từ (7.4.12) chúng ta biết rằng

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)} \tag{7.4.12}$$

Rõ ràng là, (8) và (7.4.12) là không giống nhau. Lưu ý một khám phá thú vị rằng mặc dù b_{12} là một hàm ước lượng thiên lệch, phương sai của nó có thể nhỏ hơn phương sai của $\hat{\beta}_2$ nếu σ^2 *ước lượng* trong hai mô hình không khác nhau quá nhiều. Trong trường hợp đó, phương sai của b_{12} có thể nhỏ hơn nhiều so với phương sai của $\hat{\beta}_2$ nếu r_{23}^2 lớn (tại sao?). Dĩ nhiên, rất có thể là ước lượng của σ^2 trong mô hình có đặc trưng sai có thể lớn hơn ước lượng của σ^2 thu được mô hình có đặc trưng đúng, trong trường hợp đó phương sai của b_{12} có thể không nhất thiết phải nhỏ hơn phương sai của $\hat{\beta}_2$.