XÂY DỰNG MÔ HÌNH HỒI QUI TRONG PHÂN TÍCH KINH TẾ

Lê Dân*

1. Đặt vấn đề

Theo xu hướng chung trong nghiên cứu kinh tế, phân tích định lượng ngày càng được sử dụng phổ biến. Trong đó, mô hình hồi qui được các nhà phân tích sử dụng nhiều nhất vì kết quả đem lại từ phân tích hồi qui rất đa dạng mà những phương pháp khác rất khó thực hiện được. Tuy nhiên, từ việc xây dựng mô hình đến việc lựa chọn mô hình và giải thích kết quả thường rất khó. Thông thường, các nhà phân tích chú trọng nhiều đến việc ước lượng các tham số của mô hình mà bỏ quên việc đánh giá mô hình được lựa chọn có tốt hay không. Điều này dẫn đến việc giải thích kết quả, đánh giá và hoạch định chính sách không đáng tin cậy có thể gây ra hậu quả kinh tế nghiêm trọng. Chính vì vậy, cần phải lựa chọn phương pháp ước lượng, lựa chọn tiêu chuẩn đánh giá mô hình, giải thích được ý nghĩa các tham số của mô hình.

2. Những sai lầm gặp phải khi xây dựng mô hình và tiêu chuẩn lựa chọn

Mô hình tốt hay xấu ảnh hưởng rất lớn đến chất lượng phân tích nên cần coi trọng vấn đề này. Chính vì vậy, chúng ta cần quan tâm nhiều đến việc lựa chọn mô hình. Khi xây dựng mô hình cần tránh những sai lầm có thể gặp sau [3]:

1. Bỏ sót một số biến thích đáng,
2. Bao gồm một số biến không cần thiết,
3. Chọn dạng mô hình sai,
4. Sai số đo lường,
5. Chỉ định yếu tố ngẫu nhiên không đúng,
6. Vi phạm một số giả thiết của mô hình hồi qui tuyến tính cổ điển.

Trong trường hợp gặp phải các loại sai lầm này thì các ước lượng có thể bị chệch, phương sai lớn và việc tìm khoảng tin cậy và kiểm định giả thuyết không đáng tin cậy.

Khi lựa chọn mô hình cần xem xét nhiều tiêu chuẩn khác nhau. Theo Hendry và Richard, một mô hình được chọn trong phân tích phải đảm bảo những tiêu chuẩn cơ bản sau [2], [3]:

- Vững về lý thuyết: có nghĩa là mô hình phải có ý nghĩa kinh tế. Ví dụ, nếu theo giả thuyết của Milton Friedman về thu nhập thường xuyên, hệ số chặn trong mô hình hồi qui của tiêu dùng thường xuyên theo thu nhập thường xuyên hy vọng bằng 0; khi xây dựng các hàm sản xuất thì không nên chọn hàm tuyến tính mà phải chọn hàm sản xuất Cobb-Douglas [1].

- Phù hợp với dữ liệu: mô hình phải có sai số mô hình nhỏ nhất. Để đảm bảo tiêu chuẩn này, khi chọn mô hình có thể sử dụng những tiêu chuẩn sau:

Hệ số xác định điều chỉnh $\overline{R^2}$: Vì khi tăng số biến giải thích trong mô hình làm tăng hệ số xác định nên Henry Theil đã xây dựng hệ số xác định điều chỉnh, theo công thức:

$$\overline{R^2} = 1 - \frac{RSS/(n-k)}{TSS/(n-1)} = 1 - (1-R^2)\frac{n-1}{n-k}$$

* Khoa Thống kê-Tin học, trường ĐH Kinh tế Đà Nẵng

Với biểu thức này, chúng ta thấy rằng $\overline{R^2} \leq R^2$. Không như R^2, hệ số xác định điều chỉnh $\overline{R^2}$ không tăng theo số biến có trong mô hình. Chính vì vậy, hệ số xác định điều chỉnh thường được dùng hơn hệ số xác định.

Tiêu chuẩn thông tin Akaike AIC (Akaike Information Criterion)

Nhằm xem việc thêm biến giải thích vào trong mô hình chúng ta sẽ xem xét tiêu chuẩn ACI được tính như sau:

$$AIC = e^{2k/n}\frac{\sum \hat{u}_i^2}{n} = e^{2k/n}\frac{RSS}{n}$$

Với k là số biến giải thích (bao gồm cả số hạng chặn) và n là số quan sát. Chúng ta có thể biểu diễn như sau:

$$LnAIC = \frac{2k}{n} + Ln(\frac{RSS}{n})$$

Với LnACI là log theo cơ số tự nhiên của ACI. Trong số sách và một số phần mềm xác định ACI theo dạng log. Với biểu thức đã cho, AIC chịu sự nghiêm ngặt hơn so với hệ số xác định khi thêm biến vào trong mô hình. Trong khi so sánh hai hay nhiều mô hình, mô hình nào có AIC nhỏ sẽ được chọn. Một tiện lợi của AIC là rất hữu ích không chỉ trọng nội mẫu mà còn thực hiện dự đoán ngoài mẫu. Nó còn thường được sử dụng để xác định độ trễ trong mô hình tự hồi quy AR(p).

Tiêu chuẩn thông tin Schwarz (Schwarz Information Criterion SIC)

Tương tự như ACI, tiêu chuẩn SIC được xác định như sau:

$$SIC = n^{k/n}\frac{\sum \hat{u}_i^2}{n} = n^{k/n}\frac{RSS}{n}$$

hoặc theo dạng log

$$LnSIC = \frac{k}{n}Ln(n) + Ln(\frac{RSS}{n})$$

SIC chịu sự nghiêm ngặt hơn AIC. Như AIC, giá trị của SIC nhỏ, mô hình tốt hơn.

- Ngoài ra, nếu sử dụng phương pháp bình phương bé nhất thì mô hình được lựa chọn phải thoả mãn những giả thiết: không tự tương quan giữa các phần dư, không có đa cộng tuyến, tồn tại phương sai đồng nhất [3] [5].

3. Những mô hình thường gặp trong phân tích kinh tế và cách giải thích ý nghĩa các tham số của mô hình [1] [6]

- Trong phân tích thực tiễn có rất nhiều mô hình khác nhau tuỳ thuộc mối quan hệ giữa biến phụ thuộc và những biến giải thích. Tính đa dạng của mô hình rất khó mô tả hết, chính vì vậy bài viết chỉ quan tâm một số mô hình thường dùng.

a) Mô hình hồi quy tuyến tính

Dạng kỳ vọng: $E(Yi) = \beta_1 + \beta_2 X_{2i} + ... + \beta_k X_{ki}$ $\forall_i = \overline{1,n}$ (1)

Trong đó: Y là biến phụ thuộc vào X_j là các biến giải thích; β_1 gọi là hệ số chặn và β_j ($\forall j = \overline{2,k}$) là các hệ số góc hay còn gọi các hệ số hồi qui riêng và u_i là các sai số ngẫu nhiên có kỳ vọng bằng 0 phương sai hữu hạn.

Xét mô hình (1), chúng ta thực hiện đạo hàm riêng và có kết quả là:

$$\beta_j = \frac{\partial E(Y_i)}{\partial X_{ji}}$$

Trong kinh tế, chúng ta có thể tính xấp xỉ như sau:

$$\beta_j = \frac{\partial E(Y_i)}{\partial X_{ji}} \approx \frac{\Delta E(Y_i)}{\Delta X_{ji}}$$

Với Δ thể hiện mức tăng của từng chỉ tiêu. Khi $\Delta X_{ji} = 1$ thì $\beta_j = \Delta E(Y_i)$

Với biểu thức này có thể giải thích ý nghĩa của β_j ($\forall j = \overline{2,k}$) như sau: Trong điều kiện các nhân tố khác không đổi, thì khi X_j tăng lên một đơn vị (theo đơn vị của X_j thì

E(Y) sẽ tăng bình quân β_j đơn vị (theo đơn vị của Y).

b) Mô hình hồi qui Log-Log

Dạng kỳ vọng có dạng:

$LnE(Yi) = \beta_1 + \beta_2 LnX_{2i} + ... + \beta_k LnX_{ki} \quad \forall i = \overline{1,n}$ (2)

Với Ln ký hiệu của logarit theo cơ số tự nhiên.

Dạng (2) chính là dạng hàm sản xuất Cobb-Douglas đã được tuyến tính hoá.

Liệu ý nghĩa kinh tế của các hệ số trong hàm hồi qui (2) có khác trong hàm hồi qui (1) hay không? Chúng ta không nhầm lẫn về kết quả để đưa ra những đánh giá đúng và hoạch định chính sách khoa học. Đối với mô hình (2), chúng ta có thể thực hiện đạo hàm riêng như sau:

$$\beta_j = \frac{\partial LnE(Y_i)}{\partial LnX_{ji}} = \frac{\frac{\partial E(Y_i)}{Y_i}}{\frac{\partial X_{ji}}{X_{ji}}}$$

Trong kinh tế, chúng ta có thể tính xấp xỉ như sau:

$$\beta_j = \frac{\frac{\partial E(Y_i)}{Y_i}}{\frac{\partial X_{ji}}{X_{ji}}} \approx \frac{\frac{\Delta E(Y_i)}{Y_i}}{\frac{\Delta X_{ji}}{X_{ji}}}$$

Với $\frac{\Delta E(Y_i)}{Y_i}$ và $\frac{\Delta X_{ji}}{X_{ji}}$ thể hiện tốc độ tăng của từng chỉ tiêu. Khi $\frac{\Delta X_{ji}}{X_{ji}} = 1$ thì $\beta_j = \frac{\Delta E(Y_i)}{Y_i}$

Như vậy, có thể nói β_j chính là hệ số co giãn của $E(Y_i)$ theo X_{ji}. Với biểu thức này có thể giải thích ý nghĩa của βj ($\forall j = \overline{2,k}$) như sau: Trong điều kiện các nhân tố khác không đổi, khi X_{ji} tăng lên 1% thì $E(Y_i)$ sẽ tăng bình quân β_j %.

c) Mô hình hồi qui Tuyến tính -Log

Mô hình hồi qui có dạng kỳ vọng

$E(Y_i) = \beta_1+\beta_2 LnX_{2i}+\ldots+\beta_k LnX_{ki} \quad \forall i = \overline{1,n}$ (3)

Trong hàm này, ý nghĩa của các hệ số được giải thích như thế nào? Để tìm cách giải thích ý nghĩa, chúng ta thực hiện đạo hàm riêng trong mô hình như sau:

$$\beta_j = \frac{\partial E(Y_i)}{\partial LnX_{ji}} = \frac{\partial E(Y_i)}{\frac{\partial X_{ji}}{X_{ji}}}$$

Trong kinh tế, chúng ta có thể tính xấp xỉ như sau:

$$\beta_j = \frac{\partial E(Y_i)}{\frac{\partial X_{ji}}{X_{ji}}} \approx \frac{\Delta E(Y_i)}{\frac{\Delta X_{ji}}{X_{ji}}}$$

Với $\Delta E(Y_i)$ thể hiện mức tăng của E(Yi) và $\frac{\Delta X_{ji}}{X_{ji}}$ thể hiện tốc độ tăng của Xj. Khi $\frac{\Delta X_{ji}}{X_{ji}} = 1$ thì $\beta_j = \Delta E(Y_i)$. Với biểu thức này có thể giải thích ý nghĩa của β_j ($\forall j = \overline{2,k}$) như sau: Trong điều kiện các nhân tố khác không đổi, thì khi X_{ji} tăng lên 1% thì $E(Y_i)$ sẽ tăng bình quân β_j đơn vị (theo đơn vị tính của Y).

d) Mô hình hồi qui Log-Tuyến tính

Mô hình có dạng kỳ vọng

$E(LnY_i) = \beta_1+\beta_2 X_{2i}+\ldots+\beta_k X_{ki} \quad \forall i = \overline{1,n}$ (4)

ý nghĩa của các hệ số trong hàm hồi qui này được giải thích như thế nào? Thực hiện đạo hàm riêng theo biến X_j trong mô hình và kết quả như sau:

$$\beta_j = \frac{\partial LnE(Y_i)}{\partial X_{ji}} = \frac{\frac{\partial E(Y_i)}{Y_i}}{\partial X_{ji}}$$

Trong kinh tế, chúng ta có thể tính xấp xỉ như sau:

$$\beta_j = \frac{\partial LnE(Y_i)}{\partial X_{ji}} \approx \frac{\frac{\Delta E(Y_i)}{Y_i}}{\Delta X_{ji}}$$

Với $\Delta(X_{ji})$ thể hiện mức tăng của X_j và $\frac{\Delta Y_i}{Y_i}$ thể hiện tốc độ tăng của Y. Khi $\Delta X_{ji} = 1$ thì $\beta_j = \frac{\Delta E(Y_i)}{Y_i}$. Với biểu thức này có thể giải thích ý nghĩa của βj ($\forall j = \overline{2,k}$) như sau: Trong điều kiện các nhân tố khác không đổi, thì khi X_{ji} tăng lên 1 đơn vị (theo đơn vị tính của X_j) thì $E(Y_i)$ sẽ tăng bình quân β_j%.

e) Mô hình hồi qui Log-Tuyến tính có biến giả [3]

Xem xét mô hình hồi qui Log - Tuyến tính như sau:

$$\text{Ln}(Y_i) = \alpha + \sum_{j=1}^{k} \beta_j X_{ji} + \lambda D_i + u_i \quad (5)$$

Với X_j là các biến liên tục có hệ hồi qui là β_j và D là biến giả có hệ số hồi qui là λ, u_i là các sai số ngẫu nhiên có kỳ vọng bằng 0 phương sai hữu hạn.

Cách giải thích ý nghĩa của các β_j tương tự như mô hình trên còn hệ số λ cần phải tính antilog của λ đã được ước lượng và trừ cho 1.

4. Phương pháp ước lượng và sử dụng phần mềm hỗ trợ

Mô hình tốt hay xấu ảnh hưởng rất lớn đến chất lượng phân tích nên cần coi trọng vấn đề này. Hiện nay có nhiều phương pháp khác nhau để ước lượng các tham số của mô hình, nhưng phương pháp bình phương bé nhất nếu thoả mãn một số giả thiết sẽ cho kết quả tốt nhất. Thực hiện ước lượng các tham số của mô hình hồi qui rất phức tạp, đòi hỏi khối lượng tính toán rất lớn. Chính vì vậy, cần phải thực hiện các ước lượng bằng các phần mềm máy tính. Hiện nay, có nhiều phần mềm thống kê hay kinh tế lượng phục vụ ước lượng mô hình chuyên nghiệp như: EVIEWS, STATA, SPSS... Trong đó, EVIEWS có nhiều chức năng hơn trong phân tích kinh tế lượng như lựa chọn mô hình, đánh giá việc vi phạm các giả thiết cũng như cách khắc phục hậu quả khi mô hình không thoả mãn các giả thiết. STATA, SPSS phục vụ cho nhiều yêu cầu phân tích trong thống kê. Các phần mềm này có bán rộng rãi trên thị trường. Ngoài các phần mềm chuyên nghiệp này, hiện còn có nhiều phần mềm không chuyên cho phân tích hồi qui như: EXCEL, STATISTICA, MATHCAD...

Tuỳ thuộc vào đặc điểm của những hiện tượng khác nhau mà lựa chọn mô hình hồi qui phù hợp. Không có mô hình nào thích hợp cho mọi tình huống nên cần dựa vào những tiêu chuẩn khác nhau để xác định mô hình tốt. Lựa chọn mô hình phải dựa đồng thời vào phân tích lý luận và công cụ định lượng. Trong phân tích, ngoài việc lựa chọn mô hình tốt cần biết ý nghĩa kinh tế của mỗi mô hình. Bài viết này cũng chỉ trình bày một số tiêu chuẩn lựa chọn mô hình và cách giải thích ý nghĩa của các tham số trong một số mô hình. Hy vọng bài viết này sẽ là ý tưởng cho việc lựa chọn mô hình và giải thích ý nghĩa các tham số của mô hình hồi qui khác.

Tài liệu tham khảo

[1]. S. Charles Maurice, Charles W.Smithson (1990), Kinh tế quản lý, Trung tâm tài liệu thông tin ĐHKT Quốc dân, Hà Nội.

[2]. Jan Kmenta (1986), Elements of Econometrics, Second Edition, Macmillan, NewYork.

[3]. Guijarati (1988), Basic Econometrics, Mc Graw Hill Publishing, NewYork .

[4]. Maddala (1992), Introduction to Econometrics, Macmillan Publishing Company, NewYork.

[5]. William H.Greene (1991), Econometrics Analysis, Macmillan Publishing Company, NewYork.

[6]. Paul Newbold (1995), Statistics for Business & Economics, Fourth Edition, Prentice-Hall International, Inc.