



Dự báo với mô hình hồi quy

Bởi:

Phạm Trí Cao

DỰ BÁO VỚI MÔ HÌNH HỒI QUY (Đọc thêm)PHÂN LOẠI CÁC PHƯƠNG PHÁP DỰ BÁO

Có hai nhóm phương pháp dự báo chính là nhóm định tính và nhóm định lượng. Trong giáo trình này chúng ta chủ yếu sử dụng phương pháp định lượng có kết hợp với các phán đoán định tính để dự báo.

Các phương pháp dự báo định tính

Các phương pháp dự báo định tính dựa vào phán đoán chủ quan và trực giác để đưa ra dự báo thay cho vì dựa vào các số liệu quá khứ. Phương pháp dự báo định tính hữu ích cho việc dự báo toàn cục và một số trường hợp mà số liệu quá khứ không hữu ích cho dự báo.

Các phương pháp dự báo định lượng

Các kỹ thuật dự báo định lượng dựa vào việc phân tích số liệu quá khứ để đưa ra dự báo. Giả định của phương pháp này là các nhân tố từng tác động lên biến được dự báo trong quá khứ vẫn tiếp tục ảnh hưởng đến biến này trong tương lai. Vậy dựa vào diễn biến dữ liệu trong quá khứ ta có thể dự báo cho tương lai. Các phương pháp dự báo định lượng lại được chia thành hai nhóm chính: dự báo định lượng mang tính nhân quả và dự báo định lượng mang tính thống kê.

Các phương pháp dự báo định lượng mang tính nhân quả

Đại diện của nhóm phương pháp này là phân tích hồi quy. Mô hình dự báo có hai nhóm biến số: các biến số được dự báo được gọi là biến độc lập, các biến số dùng để dự báo được gọi là biến phụ thuộc. Chúng ta đã nghiên cứu mô hình hồi quy ở phần 1, nay chúng ta tiếp tục nghiên cứu việc áp dụng mô hình hồi quy cho dự báo và một số kỹ thuật phân tích hồi quy với dữ liệu chuỗi thời gian.

Các phương pháp dự báo định lượng mang tính thống kê

Dự báo với mô hình hồi quy

Nhóm các phương pháp dự báo mang tính thống kê chỉ quan tâm đến quy luật biến thiên của biến cần dự báo trong quá khứ để đưa ra dự báo. Biến thiên của một biến số kinh tế được chia thành các thành phần: xu hướng, chu kỳ, thời vụ và ngẫu nhiên.

Nhóm các phương pháp dự báo mang tính thống kê lại chia thành hai nhóm chính.

Nhóm thứ nhất phân tích một thành phần hoặc kết hợp một số thành phần riêng biệt nêu trên như: đường xu hướng, san bằng số mũ, trung bình động.

Nhóm thứ hai sử dụng các khái niệm thống kê về dữ liệu chuỗi thời gian mà không chia biến động của dữ liệu thành các thành phần riêng biệt như ở phương pháp luận Box-Jenkins.

Dự báo với mô hình hồi quy thông thường

Mô hình hồi quy

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \epsilon_t$$

(6.1)

Chỉ số t chỉ thời kỳ thứ t.

Giả sử mô hình này thỏa mãn các điều kiện của phương pháp ước lượng theo bình phương tối thiểu. Các tham số ước lượng từ mô hình tương ứng là

$$\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$$

Ước đoán tốt nhất cho Y_{t+1} khi biết các $X_{i,t+1}$ là:

$$\hat{Y}_{t+1} = E(\hat{\beta}_1 + \hat{\beta}_2 X_{2,t+1} + \dots + \hat{\beta}_k X_{k,t+1})$$

(6.2)

Độ lệch chuẩn của ước lượng là

Đối với hồi quy hai biến

$$se(\hat{Y}_{t+1}) = \sigma \left[1 + \frac{1}{n} + \frac{(X_{t+1} - \bar{X})^2}{\sum_{i=1}^n X_i^2} \right]^{1/2}$$

Dự báo với mô hình hồi quy

(6.3)

Đối với hồi quy bội: công thức rất phức tạp và nằm ngoài phạm vi giáo trình này.

Tính chất “trễ” của dữ liệu chuỗi thời gian và hệ quả của nó đến mô hình

Khi chúng ta sử dụng mô hình (6.1) chúng ta giả định rằng các biến độc lập tác động tức thì lên biến phụ thuộc và biến phụ thuộc chỉ chịu tác động của biến độc lập. Đối với các biến số kinh tế các giả định này thường không đúng. Tác động của biến độc lập có thành phần tác động tức thời và có thành phần tác động trễ. Mặt khác, đôi khi bản thân biến phụ thuộc cũng có “quán tính” hay “sức ỳ” của nó. Có ba nguyên nhân gây ra “độ trễ” hay “sức ỳ” trong kinh tế là

Nguyên nhân tâm lý

Khi thu nhập của một người giảm tiêu dùng của người đó có thể không giảm ngay lập tức do thói quen duy trì mức sống cao. Nếu tình hình thu nhập vẫn không phục hồi trong thời gian dài, anh ta phải học cách chi tiêu tiết kiệm hơn.

Nguyên nhân kỹ thuật

Giả sử cầu nội địa đối với một mặt hàng tăng lên làm giá một mặt hàng này tăng. Sản lượng nội địa có thể không tăng tức thời vì để tăng sản lượng cần phải có thời gian xây dựng nhà máy, đầu tư máy móc thiết bị và đào tạo công nhân. Doanh nghiệp còn phải phân tích xem sự tăng cầu nội địa này có mang tính chất lâu dài hay chỉ là tức thời.

Nguyên nhân định chế

Các ràng buộc pháp lý là nguyên nhân của một số hiện tượng tác động trễ. Ví dụ nếu hợp đồng tài trợ Giải bóng đá chuyên nghiệp Việt Nam đã được ký kết có hiệu lực 2 năm thì Liên đoàn Bóng đá Việt Nam không thể huỷ hợp đồng để ký lại với một đối tác khác có số tiền tài trợ cao hơn. Giả sử số tiền tài trợ phụ thuộc tầm ảnh hưởng của giải đấu lên công chúng thể hiện qua số lượt khán giả đến sân và số lượt khán giả theo dõi qua truyền hình. Số khán giả đến sân tăng lên chỉ có thể tác động làm tăng số tiền tài trợ của lần ký kết ở 2 năm sau.

Khi có tính chất “trễ” nêu trên của dữ liệu chuỗi thời gian, mô hình (6.1) có sai số hồi quy không thỏa mãn các điều kiện của mô hình hồi quy tuyến tính cổ điển.(Tại sao?). Từ đó dự báo theo (6.2) sẽ không chính xác.

Mô hình tự hồi quy

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + \gamma_t$$

Dự báo với mô hình hồi quy

(6.4)

Mô hình (6.4) còn được gọi là mô hình động vì nó thể hiện mối liên hệ giữa giá trị của biến phụ thuộc với giá trị quá khứ của nó.

Mô hình có độ trễ phân phối

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \dots + \beta_k X_{t-k} + \varepsilon_t$$

(6.5)

Trong mô hình này k được gọi là độ trễ. Chúng ta phải xác định độ trễ k .

Cách tiếp cận của Alt và Tinberger F.F.Alt, “Distribution Lags”, *Econometrica*, quyển 10, 1942, trang 113-128. (Theo D.N.Gujarati, *Basis Econometrics 3rd Edition*, 1995, trang 591).:

Vì X_t là xác định và không tương quan với ε_t nên $X_{t-1}, X_{t-2}, \dots, X_{t-k}$ đều xác định và không tương quan với ε_t . Do đó chúng ta có thể áp dụng OLS để ước lượng tham số cho mô hình (6.5). Chúng ta sẽ xác định k bằng cách tăng dần độ trễ như sau:

Hồi quy Y_t theo X_t

Hồi quy Y_t theo X_t và $X_{t-1} \dots$

(k) Hồi quy Y_t theo $X_t, X_{t-1}, \dots, X_{t-k}$

(k+1) Hồi quy Y_t theo $X_t, X_{t-1}, \dots, X_{t-(k+1)}$

Quá trình này dừng ở độ trễ (k+1) hoặc (k+2) khi chúng ta nhận thấy các hệ số ứng với các biến trễ không có ý nghĩa thống kê hoặc đổi dấu.

Quá trình trên vướng phải bốn nhược điểm như sau:

Không có tiên liệu trước là độ trễ sẽ là bao nhiêu.

Mô hình có thêm một độ trễ thì mất đi một bậc tự do, nếu dữ liệu chuỗi thời gian không đủ dài thì ý nghĩa thống kê của mô hình ngày càng kém.

Các biến giải thích thực chất là giá trị của một biến X theo thời gian, điều này gây ra sự tương quan giữa các biến giải thích trong mô hình, tức là có hiện tượng đa cộng tuyến. Ước lượng các tham số của mô hình trong trường hợp có đa cộng tuyến sẽ cho kết quả kém chính xác.

Dự báo với mô hình hồi quy

Việc xác định độ trễ k của mô hình (6.5) theo cách thức trên là một dạng của “đào mỏ dữ liệu”.

Mô hình Koyck

Giả định:

Tất cả các hệ số ứng với biến trễ có cùng dấu

Các hệ số tuân theo cấp số nhân giảm dần:

$$\beta_k = \beta_0 \lambda^k$$

với $0 < \lambda < 1$.

Chúng ta viết lại mô hình (6.5) như sau

$$Y_t = \alpha + \beta_0 X_t + \beta_0 \lambda X_{t-1} + \beta_0 \lambda^2 X_{t-2} + \dots + \varepsilon_t$$

(6.6)

Tương tự

$$Y_{t-1} = \alpha + \beta_0 X_{t-1} + \beta_0 \lambda X_{t-2} + \beta_0 \lambda^2 X_{t-3} + \dots + \varepsilon_{t-1}$$

(6.7)

Nhân (6.7) với λ

$$\lambda Y_{t-1} = \alpha \lambda + \beta_0 \lambda X_{t-1} + \beta_0 \lambda^2 X_{t-2} + \beta_0 \lambda^3 X_{t-3} + \dots + \varepsilon_{t-1}$$

(6.8)

Lấy (6.6) trừ (6.7)

$$Y_t - \lambda Y_{t-1} = \alpha(1 - \lambda) + \beta_0 X_t + (\varepsilon_t - \lambda \varepsilon_{t-1})$$

(6.9)

Kết quả cuối cùng

$$Y_t = \alpha(1 - \lambda) + \beta_0 X_t + \lambda Y_{t-1} + \gamma_t$$

(6.10)

Với

Dự báo với mô hình hồi quy

$$\gamma_t = \epsilon_t - \lambda \epsilon_{t-1}$$

, γ_t còn được gọi là trung bình trượt của λ_t và λ_{t-1} .

Mô hình (6.10) được gọi là mô hình chuyển dạng Koyck. Chúng ta đã chuyển mô hình về phân phối thành mô hình tự hồi quy.

Mô hình kỳ vọng thích nghi

Giả sử mô hình xác định cầu tiền có dạng như sau

P.Cagan, “The Monetary Dynamics of Hyperinflations”, in M.Friedman (ed.), “*Studies in the Quantity Theory of Money*”, University of Chicago Press, 1956.

$$Y_t = \beta_0 + \beta_1 X_t^* + \epsilon_t$$

(6.11)

Y : Cầu tiền

X*: Giá trị kỳ vọng

Giá trị kỳ vọng ở đây mang ý nghĩa giá trị được mong đợi, không mang ý nghĩa giá trị trung bình thực.
của lãi suất danh nghĩa

λ : Sai số hồi quy

Lãi suất kỳ vọng của năm nay (năm t) không thể quan sát được một cách trực tiếp mà được xác định như sau

$$X_t^* - X_{t-1}^* = \lambda(X_t - X_{t-1}^*)$$

với $0 < \lambda \leq 1$.

Biểu thức này hàm ý kỳ vọng của người ta thay đổi (thích hợp) theo lãi suất thực tế, hay nói cách khác người ta học hỏi từ sai lầm.

$$X_t^* = \lambda X_t + (1 - \lambda) X_{t-1}^*$$

(6.12)

Thay (6.12) vào (6.11)

$$Y_t = \beta_0 + \beta_1 [\lambda X_t + (1 - \lambda) X_{t-1}^*] + \epsilon_t$$

Dự báo với mô hình hồi quy

Qua một số phép biến đổi tương tự như mô hình Koyck ta có

$$Y_t = \gamma\beta_0 + \gamma\beta_1 X_t + (1-\gamma)Y_{t-1} + \gamma_t$$

(6.13)

Với

$$\gamma_t = \epsilon_t - (1-\gamma)\epsilon_{t-1}$$

Mô hình hiệu chỉnh từng phần

Mô hình hiệu chỉnh từng phần phù hợp với phân tích hồi quy có độ trễ do lý do kỹ thuật và định chế.

Giả sử mức đầu tư tư bản tối ưu ứng với một mức sản lượng X cho trước là Y^* . Mô hình hồi quy đơn giản Y^* theo X như sau:

$$Y_t^* = \beta_0 + \beta_1 X_t + \epsilon_t$$

(6.14)

Thực tế chúng ta không trực tiếp quan sát được Y_t .

Giả định Y_t được xác định như sau:

$$Y_t - Y_{t-1} = \delta(Y_t^* - Y_{t-1})$$

với $0 < \delta \leq 1$. (6.15)

Trong đó

$$Y_t - Y_{t-1} = I$$

: Thay đổi lượng tư bản thực tế, cũng chính là đầu tư trong kỳ

$$Y_t^* - Y_{t-1}$$

: Thay đổi lượng tư bản mong muốn

Từ (6.14) và (6.15) sau một vài phép biến đổi chúng ta nhận được

$$Y_t = \delta\beta_0 + \delta\beta_1 X_t + (1-\delta)Y_{t-1} + \delta\epsilon_t$$

(6.17)

Một lần nữa chúng ta lại nhận được mô hình tự hồi quy.

Dự báo với mô hình hồi quy

Ước lượng mô hình tự hồi quy

Trong cả ba mô hình vừa xét, chúng ta đều nhận được mô hình cuối cùng có dạng tự hồi quy.

Koyck:

$$Y_t = \alpha(1-\lambda) + \beta_0 X_t + \lambda Y_{t-1} + (\epsilon_t - \lambda \epsilon_{t-1})$$

(6.18)

Kỳ vọng thích nghi

$$Y_t = \gamma \beta_0 + \gamma \beta_1 X_t + (1-\gamma) Y_{t-1} + [\epsilon_t - (1-\gamma) \epsilon_{t-1}]$$

(6.19)

Hiệu chỉnh từng phần

$$Y_t = \delta \beta_0 + \delta \beta_1 X_t + (1-\delta) Y_{t-1} + \delta \epsilon_t$$

(6.20)

Dạng chung của ba mô hình này là

$$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Y_{t-1} + \gamma_t$$

(6.21)

Có hai vấn đề cần lưu tâm đối với mô hình (6.21):

Thứ nhất, có sự hiện diện của biến ngẫu nhiên trong các biến độc lập, đó là Y_{t-1} . Điều này vi phạm điều kiện của mô hình hồi quy tuyến tính cổ điển.

Thứ hai, có khả năng xảy ra hiện tượng tương quan chuỗi.

Để tránh các hệ quả bất lợi do Y_{t-1} gây ra người ta sử dụng một biến thay thế cho Y_{t-1} với đặc tính biến này tương quan mạnh với Y_{t-1} nhưng không tương quan với X_t . Biến độc lập có đặc tính vừa kể được gọi là biến công cụ

N.Levitan có đề xuất dùng X_{t-1} làm biến công cụ cho Y_{t-1} và đề xuất một hệ phương trình chuẩn đặc biệt cho ước lượng hệ số, nhưng vấn đề đa cộng tuyến của mô hình cũng không được khắc phục triệt để. (Theo Gujarati, Basic Econometrics, 3rd Edition, Mc Graw-Hill Inc, 1995, trang 604-605).

Dự báo với mô hình hồi quy

6.6. Phát hiện tự tương quan trong mô hình tự hồi quy

Trị thống kê h

$$h = \hat{\rho} \sqrt{\frac{n}{1 - n[\text{var}(\hat{\alpha}_2)]}}$$

(6.22)

Trong đó: n = cỡ mẫu; $\text{var}(\hat{\alpha}_2)$ = phương sai hệ số ước lượng của Y_{t-1} .

là hệ số tự tương quan mẫu bậc nhất được xác định từ công thức

$$\hat{\rho} = \frac{\sum_{t=1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum_{t=1}^n \hat{\varepsilon}_t^2}$$

(6.23)

h có phân phối chuẩn hoá tiệm cận. Từ phân phối chuẩn hoá chúng ta có

$$P(-1,96 < h < 1,96) = 0,95$$

Quy tắc quyết định:

Nếu $h < -1,96$, chúng ta bác bỏ H_0 cho rằng mô hình không có tự tương quan bậc 1 nghịch.

Nếu $h > 1,96$, chúng ta bác bỏ H_0 cho rằng mô hình không có tự tương quan bậc 1 thuận.

Nếu $-1,96 < h < 1,96$: chúng ta không thể bác bỏ H_0 cho rằng không có tự tương quan bậc nhất.