

MÔ HÌNH DỰ BÁO NGẮN HẠN

Lê Văn Duyệt^(*)

1. Mô hình dự báo ngắn hạn đa nhân tố

Phương pháp dự báo theo mô hình đa nhân tố động là phương pháp sử dụng hàm số toán học đa biến để mô tả mối quan hệ giữa hiện tượng (yếu tố) cần dự báo với các yếu tố có quan hệ với nó rồi sử dụng số liệu có liên quan ước lượng các thông số của mô hình, kiểm định mô hình nếu mô hình phù hợp sẽ sử dụng để dự báo.

Mô hình kinh tế lượng đơn giản nhất được sử dụng trong dự báo ngắn hạn là một phương trình hồi qui tương quan, có dạng:

$$(1) \quad Y = \sum_{i=1}^k a_i f_i(X) + \varepsilon$$

Ở đây Y và X là các chỉ tiêu kinh tế thống kê được; a_i là các tham số và f_i là hàm số của X ; ε là sai số thống kê.

Trong trường hợp đặc biệt $f_i(X) = X_i$, $i = 0, 1, \dots, k$; $X_0 = 1$; phương trình (1) có dạng

$$(2) \quad Y = a_0 + a_1 X_1 + \dots + a_k X_k$$

Ở đây Y là chỉ tiêu cần dự báo; X_1, \dots, X_k là các yếu tố ảnh hưởng đến Y ,

Thông qua các dãy số liệu về Y và X_i ($i=1, \dots, k$), người ta ước lượng các tham số a_i ($i=0, 1, \dots, k$). Sử dụng phương trình hồi qui này có thể tính toán các dự báo về Y cho các năm tiếp theo bằng cách thay các giá trị X_i tương ứng vào phương trình rồi tính.

Đối với các dự báo ngắn hạn, mô hình dự báo đa nhân tố có một ưu điểm rõ nét là dễ đưa ra kết luận về tương lai sát thực. Lý do là mô hình này cùng một lúc sử dụng nhiều thông tin khác nhau để dự báo (trong mô hình dự báo sử dụng nhiều biến độc lập

khác nhau để mô tả chỉ tiêu cần dự báo).

Đối với thực tế Việt Nam, trong quá trình dự báo gặp một số khó khăn sau:

- Số liệu thiếu tính hệ thống: Từ năm 1993, hệ thống thống kê của nước ta chính thức chuyển từ hệ MPS sang hệ SNA. Tuy nhiên do mạng lưới thống kê từ cơ sở vật chất đến đội ngũ cán bộ vẫn dựa trên nền tảng sẵn có nên trong thực tế các chỉ tiêu thống kê kinh tế không được điều tra một cách hệ thống. Do vậy toàn bộ các số liệu thống kê hiện có thiếu đi tính chất hệ thống. Mặt khác, một số chỉ tiêu thống kê lại ít được công bố.

- Số liệu thiếu tính tương thích: Sự tương thích giữa các chỉ tiêu thống kê nhiều khi không được đảm bảo, do vậy khi sử dụng các mối quan hệ kinh tế để kiểm tra các số liệu thường gặp khó khăn. Trong mô hình kinh tế lượng, các phương trình định nghĩa thường không đảm bảo khi thay các số liệu thống kê quá khứ.

- Độ dài của các chuỗi số liệu theo thời gian bị giới hạn: Thực tế từ 1993 đến nay, chiều dài của chuỗi thời gian chỉ có 13 năm. Tổng cục thống kê đã tính toán thêm số liệu một số chỉ tiêu từ năm 1986 (như vậy được khoảng 20 năm). Tuy nhiên giữa các dãy số liệu nhiều khi không đồng bộ, điều này ảnh hưởng nhiều đến việc xử lý số liệu xây dựng mô hình kinh tế lượng.

2. Mô hình dự báo theo dãy số thời gian

Phương pháp dự báo theo dãy số thời gian dựa vào việc nghiên cứu quy luật phát triển của hiện tượng theo thời gian để mô tả

^(*) Viện Khoa học Thống kê

quy luật phát triển của hiện tượng được nghiên cứu, người ta sử dụng các hàm số toán học thích hợp và dựa vào số liệu phản ánh sự biến động theo thời gian của hiện tượng để ước lượng các thông số của nó sau đó kiểm định và dự báo nếu hàm số được chọn phù hợp. Hàm số được xác định các thông số được gọi là mô hình dự báo theo dãy số thời gian.

Phương pháp được sử dụng để dự báo ngắn hạn (dự báo cho tháng, quý và cho một hoặc hai năm) là mô hình dự báo theo dãy số thời gian. Dưới đây trình bày cơ sở phương pháp luận và một số mô hình hay được sử dụng trong quá trình dự báo theo dãy số thời gian.

Mô hình và giả thiết của dự báo

Giả sử có dãy số liệu động thái phản ánh sự phát triển theo thời gian của một chỉ tiêu kinh tế nào đó, ký hiệu nó như sau:

(3) $\{x_t\}_{t=1}^n$, trong đó x là chỉ tiêu thống kê, t biểu thị thời gian, còn n chỉ độ dài của dãy số; t chạy từ 1 đến n .

Dãy số liệu này được coi là sự thể hiện của một quá trình ngẫu nhiên, ví dụ X_t , nào đó. Quá trình này một mặt phụ thuộc vào yếu tố (biến) thời gian t , một mặt chịu sự chi phối của ba thành phần là: thành phần khuynh hướng $f(t)$, thành phần biến động thời vụ $\varphi(t)$ và thành phần ngẫu nhiên $z(t)$. Giữa ba thành phần này có mối quan hệ cộng tính hoặc nhân tính.

+ Quan hệ cộng tính:

$$(4) \quad X_t = f(t) + \varphi(t) + z(t)$$

+ Quan hệ nhân tính:

$$(5) \quad X_t = f(t)\varphi(t)z(t)$$

Do mô hình (5) có thể chuyển về dạng của mô hình (4) thông qua phép biến đổi lôga nên từ đây chỉ nghiên cứu ứng dụng mô hình (4).

Thành phần khuynh hướng $f(t)$ về mặt bản chất nó do nội lực của hiện tượng được nghiên cứu tác thành. Nó phát triển một cách có hệ thống và theo một quy luật nhất định tùy theo điều kiện nội và ngoại cảnh quyết định.

Thành phần biến động thời vụ $\varphi(t)$ được giả thiết là phát triển có hệ thống và theo một chu kỳ k nhất định. Điều này có nghĩa là cứ sau khoảng k thời gian hiện tượng lại được lặp lại: $\varphi(t) = \varphi(t + mk)$, trong đó k được gọi là chu kỳ dao động mùa vụ.

Thành phần ngẫu nhiên $z(t)$ là thành phần phản ánh sự tác động của các yếu tố ngẫu nhiên lên sự phát triển của hiện tượng được nghiên cứu. Ví dụ, sự cố đột nhiên mất điện hoặc đột nhiên có bão làm ảnh hưởng tới sản xuất,...

Thành phần ngẫu nhiên có đặc trưng sau:

+ Có kỳ vọng toán bằng 0: $E(z(t))=0$.

+ Không phụ thuộc vào biến thời gian t

Với các điều kiện trên có:

$$(6) \quad E(X_t) = f(t) + \varphi(t)$$

Như vậy về mặt lý thuyết, có cơ sở để coi dự báo là dự báo không chệch.

Để tiến hành dự báo phải dự báo được xu thế phát triển của thành phần khuynh hướng và thành phần biến động thời vụ. Công việc này được thực hiện theo 2 bước: bước thứ nhất là tách hai thành phần này ra khỏi dãy số động thái; bước thứ hai là dự báo sự phát triển của chúng trong tương lai. Tách hai thành phần đó ra khỏi dãy số nhằm phát hiện tính quy luật của chúng trên cơ sở đó xác định mô hình dự báo cho thích hợp. Tùy theo từng mô hình dự báo cụ thể, phương pháp tách thành phần khuynh hướng có khác nhau. Đối với thành phần mùa vụ, thường chấp nhận giả thuyết là biên độ biến động thời vụ không thay đổi, vì

vậy phương pháp tách thành phần mùa vụ ở các phương pháp đều tương tự nhau.

A. Tách thành phần khuynh hướng và dự báo theo phương pháp xấp xỉ đoạn (Gia Quyền Điều Hoà- GQDH)

Có nhiều phương pháp để tách thành phần khuynh hướng ra khỏi dãy số động thái. Ở đây chúng tôi sẽ giới thiệu phương pháp **tách thành phần khuynh hướng bằng phương pháp xấp xỉ đoạn**. Nội dung của phương pháp như sau:

Ở phương pháp này người ta coi mỗi giá trị của thành phần khuynh hướng là một đại lượng ngẫu nhiên. Nó được xác định bằng một phương trình đường thẳng mà mỗi thông số của phương trình đường thẳng này lại là một đại lượng phụ thuộc vào thời gian t. Phương trình mô tả giá trị của từng điểm của khuynh hướng có dạng:

$$P_t = a_t + b_t t, \text{ với } t = 1, 2, \dots, n$$

Để tách thành phần khuynh hướng bằng phương pháp xấp xỉ đoạn, tiến hành các bước sau:

Đầu tiên chọn một số K nào đó (thường được chọn bằng chu kỳ của mùa vụ) làm cơ sở để phân đoạn đường cong (K cũng chính là số điểm nằm trên đoạn thẳng), sau đó tính các thông số của các phương trình đoạn thẳng:

$$(7) P_{it} = a_i t + b_i, \text{ với } i = 1, 2, \dots, n-K+1, t = 1, 2, \dots, n$$

Công thức (7) biểu thị các đoạn gập khúc của đường khuynh hướng. Có nhiều phương pháp để ước lượng các thông số a_i và b_i của các phương trình này. Ở đây giới thiệu cách ước lượng các thông số **a** và **b** bằng phương pháp qua 2 điểm. Cụ thể:

$$(8) a_i = \frac{\bar{X}_{i2} - \bar{X}_i}{\bar{t}_{i2} - \bar{t}_i}$$

$$(9) b_i = \bar{X}_i - a_i \bar{t}_i,$$

$$\text{trong đó: } \bar{t}_i = \frac{1}{K} \sum_{t=i}^{i+K-1} t, \bar{X}_i = \frac{1}{K} \sum_{t=i}^{i+K-1} x_t,$$

với $t=1, 2, \dots, n; i=1, 2, \dots, n-K+1$

$$\text{và } \bar{t}_{i2} = \frac{1}{m} \sum_{t=i}^{i+K-1} t, \bar{X}_{i2} = \frac{1}{m} \sum_{t=i}^{i+K-1} x_t,$$

với $\forall (t > \bar{t}_i)$ và m là số lượng t

Khi đã có các thông số của các đoạn thẳng ta ước lượng thông số của từng thời điểm t. Do số liệu ở từng thời kỳ (điểm) của dãy số thời gian tham gia vào mô tả các đoạn thẳng khác nhau nên việc tính các thông số của phương trình mô tả các sự biến động của dãy số thời gian ở từng thời kỳ (điểm) sẽ được tính dựa vào số lần tham gia vào mô tả các đoạn thẳng khác nhau. Công thức tính:

(10a)

$$\bar{a}_t = \begin{cases} \frac{1}{t} \sum_{i=1}^t a_i; & t = 1, 2, \dots, K \\ \frac{1}{K} \sum_{i=t-K+1}^t a_i; & t = K+1, \dots, n-K+1 \\ \frac{1}{n-t+1} \sum_{i=t-K+1}^{n-K+1} a_i; & t = n-K+2, \dots, n \end{cases}$$

(10b)

$$\bar{b}_t = \begin{cases} \frac{1}{t} \sum_{i=1}^t b_i; & t = 1, 2, \dots, K \\ \frac{1}{K} \sum_{i=t-K+1}^t b_i; & t = K+1, \dots, n-K+1 \\ \frac{1}{n-t+1} \sum_{i=t-K+1}^{n-K+1} b_i; & t = n-K+2, \dots, n \end{cases}$$

Ước lượng giá trị khuynh hướng từng thời điểm t

Các giá trị khuynh hướng ở từng thời điểm t được ước lượng dựa vào công thức:

$$(11) \hat{P}_t = \bar{a}_t t + \bar{b}_t, \text{ với } t = 1, 2, \dots, n$$

Thành phần khuynh hướng tách được sẽ là cơ sở để dự báo sự phát triển của hiện tượng được nghiên cứu ở phương pháp dự báo gia quyền điều hoà và mức độ gia tăng của thành phần khuynh hướng là thành tố quan trọng xác định mức độ gia tăng của nó trong tương lai.

B. Tách thành phần khuynh hướng và dự báo bằng phương pháp san số mũ

Trong dự báo theo dãy số thời gian, thông thường người ta đi xác định dạng của thành phần khuynh hướng. Trong thực tế, rất khó xác định dạng của hàm mô tả khuynh hướng của hiện tượng được nghiên cứu, vì vậy thường người ta sử dụng đa thức bậc p để mô tả sự biến động của nó. Đa thức bậc p có đặc điểm là mô tả rất uyển chuyển sự biến động "thất thường" của hiện tượng nghiên cứu. Theo quan điểm này mô hình khuynh hướng có dạng:

$$(12)$$

$$Y_t = \lambda_0 + \lambda_1 t + \lambda_2 t^2 + \lambda_3 t^3 + \dots + \lambda_p t^p + \varphi(t) + \xi(t)$$

Ở thời điểm t , người ta muốn dự báo ở thời điểm $t+\tau$ nào đó (τ là số nguyên dương lớn hơn 1). Theo nguyên lý dự báo không chệch ta có:

$$(13)$$

$$E(Y_{t+\tau}) = \lambda_0 + \lambda_1(t+\tau) + \lambda_2(t+\tau)^2 + \lambda_3(t+\tau)^3 + \dots + \lambda_p(t+\tau)^p + \varphi(t)$$

Như vậy giá trị dự báo được xác định bằng giá trị khuynh hướng cộng với giá trị thời vụ ở thời kỳ dự báo.

Mô hình sử dụng trong phương pháp san số mũ

Do ở phương pháp san số mũ cho phép các thông số của đa thức dự báo thay đổi theo thời gian nên mô hình (39) được thay bằng mô hình sau:

$$(14)$$

$$Y_t = \lambda_{0t} + \lambda_{1t}t + \lambda_{2t}t^2 + \lambda_{3t}t^3 + \dots + \lambda_{pt}t^p + \varphi(t) + \xi(t)$$

Mô hình (14) khác với mô hình (12) ở chỗ là các thông số có thêm chỉ số t để biểu thị là chúng thay đổi theo thời gian.

Trong phương pháp san số mũ, dự báo được tiến hành theo một phương pháp đặc biệt do vậy người ta biểu diễn mô hình (14) ở dạng:

$$(15)$$

$$Y_t = a_{0t} + a_{1t}t + \frac{1}{2!}a_{2t}t^2 + \frac{1}{3!}a_{3t}t^3 + \dots + \frac{1}{p!}a_{pt}t^p + \varphi(t) + \xi(t)$$

Trong đó, $\lambda_{0t} = a_{0t}$, $\lambda_{1t} = a_{1t}$, $\lambda_{2t} = \frac{1}{2!}a_{2t}, \dots,$

$$\lambda_{pt} = \frac{1}{p!}a_{pt}$$

Mô hình (14) được quyền chuyển đổi thành mô hình (15) vì có thể xây dựng một ánh xạ 1-1 cho các hệ số tương ứng của hai mô hình này. Mô hình (15) giúp chúng ta dễ dàng hơn trong nghiên cứu vì khi lấy vi phân bậc i của các hệ số chúng ta thu được một kết quả rất đơn giản đó là a_{it} :

$$\left(\frac{1}{i!}a_{it}t^i\right)^{(i)} = a_{it}, \text{ ở đây (i) biểu thị vi phân bậc } i.$$

C. Tách thành phần khuynh hướng và dự báo bằng phương pháp tự hồi quy

Mô hình tự tương quan có dạng:

(16)

$$X_t = b_1 X_{t-1} + b_2 X_{t-2} + b_3 X_{t-3} + \dots + b_p X_{t-p} + U_t$$

Trong đó t là biến chỉ thời gian, p là độ trễ và U_t là thành phần ngẫu nhiên.

Mô hình (16) cho thấy, giá trị xưa X ở thời điểm t là hàm số của chính bản thân nó ở các thời điểm trước đó (vì thế mới có tên gọi là "tự tương quan?"). Như vậy, mô hình tự hồi quy là một dạng của mô hình đa nhân tố, chỉ khác là các biến độc lập lại là chính bản thân nó. Vì vậy, việc ước lượng thông số của mô hình được thực hiện giống như ở mô hình đa nhân tố.

D. Tách thành phần mùa vụ

Sau khi đã ước lượng được giá trị khuynh hướng ta tiến hành tách thành phần mùa vụ. Để việc tính toán sau này được thuận tiện, chúng tôi tách thành phần mùa vụ theo dạng chỉ số. Cách làm đó như sau:

Đầu tiên tính tỷ số (M) giữa giá thực tế và giá trị khuynh hướng đã tách được:

$$(16) M_t^j = \frac{x_t^j}{\hat{x}_t^j}, \text{ với } t = 1, 2, \dots, n; j = 1, 2, \dots, k \text{ và } k \text{ là số thời vụ}$$

E. Các giả thiết cơ bản của dự báo

Khi tiến hành dự báo, thường xuất hiện các tình huống sau:

a) Trong thời gian từ t đến $t+\tau$ (thời kỳ dự báo), các hệ số của mô hình dự báo thay đổi.

b) Trong thời gian từ t đến $t+\tau$, bậc của đa thức thay đổi, và

c) Phương sai của thành phần ngẫu nhiên không phải là một hàm tăng theo thời gian.

d) Chu kỳ biến động thời vụ không thay đổi theo thời gian.

Trường hợp a) và b) dẫn chúng ta đến tình huống là các giá trị dự báo sẽ không

được xác định bằng kỳ vọng toán của biến được dự báo. Và như vậy dự báo của chúng ta là dự báo chệch.

Trong trường hợp thành phần ngẫu nhiên là một hàm tăng theo thời gian, dự báo vẫn cho kết quả không chệch song lại mắc sai số hệ thống. Điều này làm cho hiệu quả dự báo giảm đi đáng kể.

Thông thường khi tiến hành dự báo, người ta giả thiết rằng các trường hợp a), b), c) không xảy ra. Các giả thiết như vậy được gọi là *các giả thiết cơ bản của dự báo*.

D. Tiến hành dự báo

Để dự báo theo dãy số thời gian, việc làm trước tiên là dự báo cho thành phần khuynh hướng, sau đó nhân nó với hệ số mùa vụ tương ứng để có được kết quả dự báo của chỉ tiêu cần dự báo.

Tóm lại, dự báo là dựa vào các thông tin phản ánh sự phát triển của hiện tượng được quan tâm trong quá khứ và hiện tại để dự đoán khả năng phát triển của nó trong tương lai. Dự báo có một giả thiết rất cơ bản là quy luật phát triển của hiện tượng trong quá khứ và hiện tại vẫn giữ nguyên trong tương lai. Vì vậy, để việc dự báo sát với thực tiễn, người sử dụng kết quả dự báo cần xem xét, đánh giá tình hình để xem "liệu trong thời gian tới có gì đột biến không" để chỉnh lý kết quả đôi chút ■

Tài liệu tham khảo

1. Viện Khoa học Thống kê: Báo cáo tổng hợp kết quả nghiên cứu đề tài "Nghiên cứu ứng dụng các phương pháp dự báo ngắn hạn để dự báo một số chỉ tiêu thống kê kinh tế chủ yếu".

2. Oskar Lange: Wstep do Econometria

3. Siegel: Pratical Business Statistics