TẠP CHÍ KHOA HỌC TÍNH TOÁN VÀ ĐIỀU KHIỂN TẬP III, (1987) Số 2 (4–8)

PHƯƠNG PHÁP SAN SỐ MŨ TRONG DỰ ĐOÁN DÃY CHỈ TIÊU KINH TẾ

TRẦN THỌ ĐẠT, VŨ QUỐC HUY
Trường đại học kinh tế quốc dân

1. ĐẶT VẤN ĐỀ

Trong khi nghiên cứu các quá trình kinh tế, dãy thời gian $\{x_t\}$ thường được xem như là tổng của hai thành phần:

$$x_t = \xi_t + \varepsilon_t$$

trong đó ξ_t là hàm biến thời gian không ngẫu nhiên, phản ánh xu hướng cơ bản của quá trình và là kết quả tác động của các nhân tố chủ yếu đến quá trình đó. Thành phần ε_t phản ánh tác động ngẫu nhiên của các nhân tố khác.

Nhờ phân tích số liệu thống kê, ta xác định ξ_t thông qua việc xấp xỉ nó bằng một hàm P_t. Như thường lệ, các hệ số của P_t tìm được bằng phương pháp bình phương nhỏ nhất và nó đặc trưng cho mức độ trung bình của toàn bộ quá trình nghiên cứu. Để phù hợp hơn với thực tế, khi tiến hành dự đoán ta cần lưu ý rằng, kết quả dự đoán thường chịu ảnh hưởng của những quan sát cuối cùng nhiều hơn là những quan sát ở xa về quá khứ. Vì vậy, các hệ số của P_t cần phải thay đổi theo một quy luật nào đó để đáp ứng yêu cầu trên.

Phương pháp san số mũ là một biến dạng của phương pháp bình phương nhỏ nhất, cho phép tìm được các hệ số đặc trưng cho xu hướng được hình thành tới thời điểm quan sát cuối cùng của dãy số.

2. PHƯƠNG PHÁP SAN SỐ MŨ

Cho dãy thời gian: ..., x_{T-i}, x_{T-i+1}, ..., x_{T-1}, x_T (1)

Vấn đề đặt ra là tìm ước lượng P_{T-i} của xu thế ξ_t, làm cực tiểu hàm sai số có trọng số mũ:

$$e = \alpha \sum_{i=0}^{\infty} \beta^i (x_{T-i} - P_{T-i})^2 \quad 0 < \alpha < 1,\ \beta = 1 - \alpha \tag{2}$$

Quy ước biểu thức: $$P_{T+t} = \sum_{n=0}^{N} a_n^{[T]} t^n \tag{3}$$

là giá trị của hàm tại điểm t, ký hiệu $[T]$ của hệ số $a_n^{[T]}$ để chỉ hệ số tương ứng của hàm được xác định tại điểm gốc T theo trục thời gian t.

Gọi: $S_T^{[1]}(x) = \alpha \sum_{i=0}^{\infty} \beta^i x_{T-i}$ là trung bình mũ bậc 1 đối với dãy (1)

– Trung bình mũ bậc hai đối với dãy (1) là:

$$S_T^{[2]}(x) = \alpha \sum_{i=0}^{\infty} \beta^i \quad S_{T-i}^{[1]}(x).$$

— Cuối cùng $S_T^{[k]}(x) = \alpha \sum_{i=0}^{\infty} \beta^i S_{T-i}^{[k-1]} x$ là trung bình mũ bậc k đối với dãy (1).

Trong [1], R.G.Brown và R.F.Meyer đã chứng minh một định lý (gọi là định lý cơ bản của phương pháp san số mũ), cho phép biểu diễn các hệ số của (3) qua các trung bình mũ tương ứng.

— Định lý cơ bản của phương pháp san số mũ.

Để cho đa thức $P_{T+i} = \sum_{n=0}^{N} a_n^{[T]} i^n$ làm cực tiểu hàm sai số $e = \alpha \sum \beta^i (x_{T-i} - P_{T-i})^2$

điều kiện cần và đủ là:

$$S_T^{[k+1]}(x) = S_T^{[k+1]}(P) \quad (k = 0, 1, \ldots, N) \tag{4}$$

trong đó $S_T^{[k+1]}(P)$ là trung bình mũ cấp k+1 của dãy các giá trị của (3): $\ldots P_{T-i}, \ldots, P_T$.

Từ các định nghĩa về trung bình mũ các bậc như trên, ta có:

$$S_T^{[k]}(x) = \alpha S_T^{[k-1]}(x) + \beta S_{T-1}^{[k]}(x) \tag{5}$$

Nhờ công thức truy toán này và định lý cơ bản của phương pháp san số mũ ta rút ra công thức tính toán hệ số đối với mô hình tuyến tính và mô hình parabol (hai mô hình thường hay được sử dụng trong thực tế dự đoán) như sau:

— Đối với mô hình tuyến tính:

Khi đó N = 1 và $P_{T-i} = \sum_{n=0}^{1} a_n^{[T]} (-i)^n = a_0^{[T]} - i\, a_1^{[T]}$

Các hệ số được xác định là:

$$a_0^{[T]} = 2 S_T^{[1]}(x) - S_T^{[2]}(x)$$

$$a_1^{[T]} = \frac{\alpha}{\beta} \left(S_T^{[1]}(x) - S_T^{[2]}(x) \right)$$

— Đối với mô hình parabol:

Khi đó N = 2 và $P_{T-i} = \sum_{n=0}^{2} a_n^{[T]} (-i)^n = a_0^{[T]} - i\, a_1^{[T]} + i^2 a_2^{[T]}$.

Các hệ số được xác định là:

$$a_0^{[T]} = 3 S_T^{[1]}(x) - 3 S_T^{[2]}(x) + S_T^{[3]}(x)$$

$$a_1^{[T]} = \frac{\alpha}{2\beta^2} \left[(6-5\alpha) S_T^{[1]}(x) - 2(5-4\alpha) S_T^{[2]}(x) - (4-3\alpha) S_T^{[3]}(x) \right] \tag{6}$$

$$a_2^{[T]} = \frac{\alpha^2}{\beta^2} \left[S_T^{[1]}(x) - 2 S_T^{[2]}(x) + S_T^{[3]}(x) \right. \tag{6}$$

3. LƯỢC ĐỒ TÍNH TOÁN TRONG DỰ ĐOÁN

Để ví dụ, ta xét mô hình parabol:

Giả sử đã biết nghiệm của bài toán san số mũ đối với dãy:

$$\ldots, x_{t-p}, x_{t-p+1}, \ldots, x_{t-1}$$

là $\hat{a}_0^{[t-1]}$, $\hat{a}_1^{[t-1]}$, $\hat{a}_2^{[t-1]}$ thì nghiệm $\hat{a}_0^{[t]}$, $\hat{a}_1^{[t]}$, $\hat{a}_2^{[t]}$ của bài toán san số mũ đối với dãy: ..., x_{t-p}, x_{t-p+1}, ..., x_t được tính như sau:

Từ (6), thay $T = t-1$ rồi giải hệ đó, tìm được:

$S_{t-1}^{[1]}(x)$, $S_{t-1}^{[2]}(x)$, $S_{t-1}^{[3]}(x)$. Theo (5), ta sẽ tính được $S_t^{[1]}(x)$, $S_t^{[2]}(x)$, $S_t^{[3]}(x)$ rồi cuối cùng lại thay vào (6) sẽ tìm được $\hat{a}_0^{[t]}$, $\hat{a}_1^{[t]}$, $\hat{a}_2^{[t]}$. Bây giờ, giả sử cần tiến hành dự đoán nếu có dãy thời gian hữu hạn:

$$x_{T-r}, x_{T-r+1}, ..., x_T$$

Chúng ta giải bài toán san số mũ đối với dãy:

$$..., 0, 0, x_{T-r+1}, ..., x_T$$

Đối với dãy: ..., x_{T-r-2}, x_{T-r-1}, thì do tất cả các số hạng của nó đều bằng không nên rõ ràng:

$$\hat{a}_0^{[T-r-1]} = \hat{a}_1^{[T-r-1]} = \hat{a}_2^{[T-r-1]} = 0$$

Áp dụng lược đồ đã nói ở trên, tiếp tục giải bài toán san số mũ đối với dãy: ..., 0, 0, x_{T-r} và tìm được: $\hat{a}_0^{[T-r]}$, $\hat{a}_1^{[T-r]}$, $\hat{a}_2^{[T-r]}$. Cứ tiến hành như thế cho đến cuối cùng thu được $\hat{a}_0^{[T]}$, $\hat{a}_1^{[T]}$, $\hat{a}_2^{[T]}$ là nghiệm của bài toán san số mũ đối với dãy: ...0, 0, x_{T-r}, x_{T-r+1}, ..., x_T

Sau đó, tiến hành ngoại suy t bước:

$$\hat{x}_T^{(t)} = \hat{a}_0^{[T]} + t\,\hat{a}_1^{[T]} + \frac{1}{2} t^2\, \hat{a}_2^{[T]}$$

4. TÍN HIỆU KIỂM TRA MÔ HÌNH DỰ ĐOÁN

Ở trên, chưa nói về việc lựa chọn tham số san α như thế nào. Rõ ràng là một khi mô hình đã được lựa chọn thì tính chất của quá trình san và cả chất lượng dự đoán nữa đều do tham số san quyết định. Việc lựa chọn được tiến hành trên cơ sở phân tích giá trị quá khứ của chỉ tiêu dự đoán và thường là dẫn đến sự nhân nhượng giữa tính nhạy cảm và tính ỳ của mô hình dự đoán. Trong các sách báo hiện nay, thường đưa ra tham số san không lớn lắm, tức là vẫn thiên về tính ỳ. Tuy nhiên, hướng cơ bản của việc hoàn thiện phương pháp san số mũ là nâng cao các tính chất thích nghi của nó. Muốn thực hiện được điều này, phải giải quyết hai vấn đề:

— Thứ nhất là phát hiện ra sự thay đổi của dãy thời gian mà tham số hiện tại của mô hình không còn thỏa mãn nữa

— Thứ hai là cải tiến tham số một cách thích ứng. Vấn đề thứ nhất được D.W.Trigg [2] giải quyết bằng cách xây dựng một tín hiệu nhằm phát hiện thời điểm khi chất lượng dự đoán không còn chính xác nữa và cần phải có tham số khác.

Tín hiệu được xác định như sau:

$$K_t = \frac{\hat{e}_t}{\tilde{e}_t} \qquad (7)$$

Trong đó

$$\hat{e}_t = (1-\gamma)\,\hat{e}_{t-1} + \gamma e_t \qquad 0 < \gamma < 1 \qquad (8)$$
$$\tilde{e}_t = (1-\gamma)\,\tilde{e}_{t-1} + \gamma |e_t|$$

trừ $\hat{e}_t$ là tổng các sai số dự đoán được san, còn $\tilde{e}_t$ là tổng các trị tuyệt của sai số dự đoán được san.

Tín hiệu này được sử dụng khá rộng rãi vì các giá trị của nó chỉ biến thiên trong các giới hạn từ -1 đến $+1$. Nếu mô hình dự đoán tỏ ra không phù hợp với quá trình nghiên cứu thì các sai số sẽ cùng dấu và tín hiệu kiểm tra tiến đến -1 hay $+1$. Tham số γ trong (8) nói chung khác với α mặc dù chúng có cùng bản chất và cùng giới hạn.

Khi sử dụng tín hiệu K_t, điều quan trọng là phải lựa chọn các giới hạn cho nó để sao cho khi vượt ra ngoài giới hạn đó thì có thể phát tín hiệu về sự không chính xác của dự đoán thu được. Có thể chứng minh rằng, các giá trị xấp xỉ của nó là $\pm 2{,}4\sqrt{\frac{\gamma}{2-\gamma}}$. Còn vấn đề cải tiến tham số của mô hình dự đoán tức là tăng giá trị của tham số san thì lại thường không được quy định bằng một quy tắc nào đó mà do từng nhà nghiên cứu thực hiện trong từng trường hợp cụ thể.

Trong [3], D.W.Trigg và A.G.Leach đã sử dụng chính ngay $|K_t|$ làm tham số cải tiến vì nó có cùng giới hạn biến thiên như tham số san và sự thay đổi của dãy càng lớn thì nó càng lớn.

Như vậy, ở đây xuất hiện mối liên hệ ngược tự động. Tuy nhiên, thực tế đã chứng minh rằng điều này không phải bao giờ cũng đem lại kết quả tốt. Trong trường hợp khi tham số san khá nhỏ mà tín hiệu tính ra lại khá hơn thì nên tăng tham số san lên từ từ để tín hiệu lại lọt vào các giới hạn cho phép là được. Điều đó có thể dễ dàng được thực hiện thông qua quá trình đối thoại giữa người và máy.

5. ỨNG DỤNG

Do khối lượng tính toán tương đối lớn nên phương pháp này ít được sử dụng trong dự đoán ở nước ta. Trong thời gian qua, chúng tôi đã sử dụng nó để dự đoán một số chỉ tiêu kinh tế trên máy tính APPLE – II. Để minh họa, xin dẫn ra một ví dụ về dự đoán sản lượng khai thác than tại khu mỏ X của miền Bắc nước ta như sau:

Năm	Sản lượng khai thác than (nghìn tấn)	Năm	Sản lượng khai thác than (nghìn tấn)
1965	21,14	1975	95,06
1966	25,3	1976	106,06
1967	28,8	1977	119,48
1968	33,1	1978	126,4
1969	39,36	1979	133,74
1970	48,88	1980	148,52
1971	51,7	1981	162,58
1972	58,76	1982	172,47
1973	69,06	1983	177,52
1974	80,74	1984	182,18

Xu thế của dãy thời gian được mô tả bằng đa thức bậc 2

$$x(t) = 9{,}9 + 5{,}52t + 0{,}18t^2$$

Kết quả tính toán (dự đoán 1 bước) đối với một số giá trị trong bảng sau:

α	Sai số tuyệt đối trung bình	Sai số bình phương trung bình
0,05	4,2	25,71
0,1	4,22	27,67
0,2	4,06	20,2
0,25	3,46	18,3
0,3	3,60	21,21
0,35	3,51	19,26
0,4	3,57	17,94
0,45	3,63	17,15

Lấy $\gamma = 0,1$ và tính các giới hạn đối với tín hiệu, ta được các giới hạn đó là $\pm 0,55$.

Thử xét tín hiệu thay đổi thế nào trong trường hợp xấu nhất là khi $\alpha = 0,1$. Kết quả tính cho thấy bắt đầu từ bước thứ 5, tín hiệu $K_5 = -0,72$, tức là đã vượt ra ngoài giới hạn cho phép. Ta cần thay đổi α (tăng lên từ từ) sao cho tín hiệu lại lọt vào giới hạn đó là được.

Nhận ngày 15-4-1986

TÀI LIỆU THAM KHẢO

1. R.G. Brown, R.F. Meyer, The fundamental theorem of exponential smoothing. Oper Res. Quart.. 1961, vol 9, n.5.

2. D.W. Trigg, Monitoring a forecasting system Oper. Res. Quart., 1964, vol 15, n.3.

3. D.W. Trigg, A.G. Leoh, Exponential smoothing with an adaptive response rate. Oper. Res. Quart., 1967, vol, 18, n.1.

4. Адаптация при прогнозировании экономических показателей методом экспоненциального сглаживания «Экономика и математические методы». Том XII, 6-1981.

ABSTRACT

This article concentrates on the content of the method of exponential smoothing in forecasting economic series, the use of signal of D.W. Trigg and A.G. Leach.

However, in many cases, such a direct use can make big errors. The author asks to use the signal to controll the forecasting model and the correction of adaptive parameter is made by the conversation between men and computers.

PHÉP TÁCH SUY BIẾN...

(Tiếp theo trang 3)

TÀI LIỆU THAM KHẢO

1. W.W. Armstrong, C. Delobel, Decompositions and functional dependencies in relations, ACM TODS 5, 4 (1980), 401—430.

2. C. Beeri, R. Fagin, D. Maier, M. Yannakakis, On the desirability of acyclic database schemes, U.ACM 30, 3(1983), 479—513.

3 L.T. Vuong, Untersuchung zur ternaeren Dekomposition einer Relation und zur Anwendung der unscharfen Mengen im CRM, Diss. TU—Dresden, 1983.

4. L.T. Vuong, Ueber n-fache Dekomposition einer Relation im Coddxhen Relationenmodell Közlemenyek 31(1984), 95-113.

ABSTRACT

DEGENERATE DECOMPOSITION OF A RELATION IN DATABASE

In this paper we introduced the notion of ternary degenerate decomposition of a relation and inverstigated its properties.

By applicatioo of this degenerate decomposition it can be captured more sematics of data in a relational database.