

KNN예측방법을 리용한 환자시세의 예측가능성분석에 대한 리해

김 경 민

위대한 령도자 김정일동지께서는 다음과 같이 교시하시였다.

《우리가 자본주의시장을 대상으로 무역을 하는 조건에서 자본주의시장과 세계경제에 대한 파악이 없이는 무역을 잘할수 없습니다.》(《김정일선집》 증보판 제18권 369페이지)

국제금융시장에서 환자시세를 사전에 정확히 예측하는것은 금융시장의 변화추세에 맞게 대외경제활동과 국제금융거래를 주동적으로 벌려나가는데서 중요한 의의를 가진다.

국제금융시장에서 수시로 변하는 환자시세를 예측하는데서 나서는 중요한 문제는 높은 예측정확성을 보장할수 있는 예측방법을 옳게 적용하는것이다.

국제금융시장에서 환자시세의 예측은 지난 시기의 환자시세에 대한 자료를 분석하여 진행한다.

환자시세의 예측에 리용되는 지난 시기의 환자시세자료는 일정한 시계렬을 이룬다. 그러므로 과거의 시세자료를 리용한 환자시세의 예측은 시계렬예측으로 된다.

시계렬예측은 실생활에서 아주 의의있고 중요한 예측방법으로서 경제현상을 비롯한 여러 분야의 예측에서 널리 리용되고있다. 실례로 회귀모형, 유전알고리즘, 신경회로망, 웨블레트, 카오스예측 등을 들수 있다.

이러한 예측방법들가운데서 국제금융시장에서 환자시세의 변동을 예측하는데 효과적으로 리용할수 있는 방법은 카오스예측이다. 그것은 환자시세의 변동이 매우 복잡하고 무질서하며 예측하기 힘든 카오스적현상이기때문이다.

국제금융시장에서 환자시세의 변동은 많은 요인들의 영향을 받으며 그 요인들의 영향정도도 파악하기 매우 어렵다. 뿐만아니라 수많은 요인들가운데서 어느 시각에 어느 요인이 작용할것인지 알수 없으며 그 효과도 각이하다.

카오스예측에서 환자시세의 예측에 적용할수 있는 대표적인 방법은 KNN예측방법이다.

KNN예측방법은 과거의 시계렬변동을 가지고 앞으로의 시계렬의 움직임을 알아내는 방법이다.

이 방법을 실천에 적용하는데서 가장 큰 문제는 시계렬자료에 내재하고있는 본태적인 동력학적구조를 찾아내는것이다. 이 구조는 일정한 시간차이로 류사성을 가지고있으며 예측의 정확성과 직결되어있는 매우 중요한 문제로 된다.

이 문제를 해결하자면 반드시 2가지 단계를 거쳐야 한다.

하나는 국제금융시장에 대한 관측으로부터 얻은 환자시세의 1차원시계렬자료를 가지고 본래의 동력학적구조와 류사한 흡인체를 만드는것이고 다른 하나는 이 흡인체에 존재하는 우연성분을 제거함으로써 보다 높은 수준에서의 궤도의 류사성을 찾아내는것이다.

우연성분에는 국내외의 정치 및 군사정세의 변화, 개별적정치인들이나 경제인들, 기업가들, 경제평론가들의 발언, 개별적기업들의 경영상태와 화폐시장 및 주식시장들에서의 각종 투기행위 등이 포함될수 있다.

이러한 모든 형태의 우연성분들을 제거하고 환자시세변동의 고유한 법칙성을 찾아낸다면 궤도들의 류사성을 리용하여 앞날을 예측할수 있다.

KNN방법을 리용하여 국제시장에서 환자시세의 변동을 예측하기 위하여서는 이 방

법의 적용가능성 다시말하여 이 방법에 의한 환자시세예측가능성을 밝혀야 한다.

일반적으로 예측모형을 만들면서 일련의 추가적인 가정을 하기때문에 모형설계보다는 시계열자체를 가지고 그의 예측가능성을 연구하는것이 더 합리적이다. 그러므로 KNN방법에 의한 환자시세예측에서는 25년간의 두 나라 화폐사이의 일간환자시세변동자료들의 중간값을 가지고 예측가능성을 시험한 다음 그 결과를 분석하는 방식을 적용할수 있다.

KNN방법에 의한 예측가능성분석에서는 실지로 국제금융시장에서 많이 거래되는 전 환성화폐들사이의 환자시세를 분석대상으로 선택하여야 하며 충분히 긴 기간 실례로 25 년간에 대하여 매일매일의 환자시세자료를 수집한 다음 그것을 리용하여 예측가능성을 시험하여야 한다.

KNN방법을 리용하여 전환성을 가진 두 나라 화폐의 환자시세변동자료(1990-2014 년)를 가지고 그의 예측가능성을 시험할수 있는 방안은 다음과 같다.

우선 정확성을 높이기 위하여 예측의 전처리단계를 설정하고 여기에서 예측의 정확 성을 떨어뜨리는 우연성분을 제거하기 위한 시험을 진행한다.

일반적으로 환자시세자료는 1차원스칼라시계열자료로 주어진다.

환자시세자료의 분석에서는 이러한 1차원스칼라시계열을 발생시키는 함수가 몇차원 공간에서 정의되는 함수인가를 밝혀야 한다.

상관적분법을 리용한 계산에 의하면 두 나라 화폐의 환자시세자료가 10차원공간에서 정의되는 함수로부터 얻어진다는것을 알수 있다.

이미 알려진것처럼 상관적분 $C(\varepsilon)$ 은 다음과 같이 정의된다.

$$C(\varepsilon) = \frac{1}{N^2} \sum_{i,j} \theta(\varepsilon) - \|X_i - X_j\| \quad (1)$$

$\varepsilon \rightarrow 0, N \rightarrow \infty$ 일 때 우연성분이 없는 경우에는

$$C(\varepsilon) \sim \varepsilon^d \quad (2)$$

로 된다. 여기서 d 는 흡인체의 상관차원이다.

$$d = \lim_{\substack{\varepsilon \rightarrow 0 \\ N \rightarrow \infty}} d(\varepsilon), \quad d(\varepsilon) = \frac{d \ln c(\varepsilon)}{d \ln c} \quad (3)$$

10차원공간에서 정의되는 력학계로부터 두 나라 화폐의 환자시세자료가 얻어지며 그 가운데서 어느 한 자리표측에 대한 사영성분으로서의 환자시세자료를 관측할수 있다.

또한 상관적분법을 리용하여 계산된 환자시세자료의 상관적분곡선족을 얻는다.

분석에 의하면 두 나라 화폐의 환자시세자료의 상관차원값이 9보다는 크고 10을 넘 지 못한다.

분석을 통하여 실지로 두 나라 화폐의 환자시세자료가 명백히 카오스성을 가지고있 다는것을 알수 있다. 그리고 두 나라 화폐의 환자시세자료를 발생시키는 력학계가 포함 하는 경제적요인들이 무엇인지는 모르지만 그것들의 개수가 10개정도일것이라는것은 알 수 있다.

타켄즈의 문힘정리를 리용하여 1차원관측시계열만을 가지고도 본래의 동력학적구조 와 유사한 구조를 가진 흡인체를 구성할수 있다.

1차원스칼라시계열이 $X_i (i=1, N)$ 라고 하자.

이때 첫번째 10차원 공간벡토르는 $X_1 = (X_1, X_2, \dots, X_{10})$
 두번째 10차원 공간벡토르는 $X_2 = (X_2, X_3, \dots, X_{11})$

...
 n 번째 10차원 공간벡토르는 $X_n = (X_n, X_{n+1}, \dots, X_{n+9})$ 로 된다.

이것이 타켄즈의 문힘정리에서 정의된 지연자리표계이다.

이 자리표계를 리용하면 1차원스칼라시계열만을 가지고도 쉽게 n 차원공간의 력학계를 구성할수 있으며 이를 통하여 1차원시계열을 발생시키는 력학계의 구조를 알수 있다.

또한 환자시세자료를 그대로 리용하지 않고 다음과 같은 연산과정을 거친 다음 리용한다.

$$r_i = \ln \frac{X_i}{X_{i-1}} \quad (4)$$

시계열 $\{X_i\}$ 대신 시계열 $\{r_i\}$ 를 리용하는것이 합리적이다.

우의 식 (4)에서 r_i 는 서로 이웃하고있는 환자시세자료들사이의 비를 자연로그로 처리한것이다. 이것은 오늘의 환자시세가 어제의 환자시세에 비해볼 때 얼마나 증가하였는가 혹은 감소하였는가를 보여주는 량이다.

우의 식을 리용하여 두 나라 화폐의 환자시세자료와 그에 대응한 대리시계열 $\{r_i\}$ 를 얻을수 있다.

대리시계열 $\{r_i\}$ 를 리용하여 환자시세자료속에 들어있는 시간적차이를 줄이고 궤도의 류사성을 얻어낼수 있는 가능성을 찾을수 있다.

또한 궤도들간의 류사성을 얻어내기 위하여 우연성분을 국부사영방법을 리용하여 제거한다.

국부사영방법을 리용하여 로렌즈계에서 비선형우연성분을 제거하는 방법은 다음과 같다.

로렌즈계는 카오스성을 가진 비선형동력학계를 연구하는데 리용되는 전형적인 계로서 다음과 같이 서술된다.

$$\left\{ \begin{array}{l} \frac{dx}{dt} = -ax + ay \\ \frac{dy}{dt} = -xz + cx - y \\ \frac{dz}{dt} = xy - bz \end{array} \right. \quad (5)$$

이 식을 리용하면 궤도들의 형태들사이에 류사성이 존재한다는것을 알수 있으며 이러한 류사성을 리용하면 앞날을 예측할수 있다. 즉 상공간에서 관찰점으로부터 가장 가까운 k 개의 점들을 선택하고 그 점들로부터의 시간진화특성을 리용하여 앞으로의 궤도를 예측한다.

이러한 예측모형의 기본내용은 다음의 식들에서 명백히 알수 있다.

벡토르렬 $\{\overline{X_{n+1}}\}$, $i = \overline{1, n}$ 에 대하여

$$\{\overline{X_{n+1}}\} = F(\overline{X_{i_1+1}}, \overline{X_{i_2+1}}, \dots, \overline{X_{i_k+1}}) \quad (6)$$

로 쓸수 있다.

가장 단순한 방법은 k 개의 이웃점들을 평균하여 예측값을 얻는것이다.

$$\{\overline{X_{n+1}}\} = \frac{1}{k} (\overline{X_{i_1+1}} + \overline{X_{i_2+1}} + \dots + \overline{X_{i_k+1}}) \quad (7)$$

로렌즈계에 대한 적용결과를 보면 예측결과가 800걸음이상까지도 높은 수준의 정확성을 유지한다는것을 알수 있다.

이 예측방법을 두 나라 화폐의 환자시세자료에 적용한 결과는 다음과 같다.

날자	실지값	예측값	날자	실지값	예측값
2014.6.1	0.596	0.597	2014.6.17	0.59	0.589
2014.6.2	0.597	0.598	2014.6.18	0.589	0.589
2014.6.3	0.598	0.597	2014.6.19	0.59	0.59
2014.6.4	0.595	0.597	2014.6.20	0.591	0.589
2014.6.5	0.594	0.595	2014.6.21	0.59	0.591
2014.6.6	0.596	0.595	2014.6.22	0.591	0.589
2014.6.7	0.597	0.594	2014.6.23	0.588	0.591
2014.6.8	0.596	0.595	2014.6.24	0.589	0.588
2014.6.9	0.594	0.595	2014.6.25	0.587	0.588
2014.6.10	0.594	0.596	2014.6.26	0.588	0.595
2014.6.11	0.593	0.593	2014.6.27	0.593	0.6
2014.6.12	0.595	0.593	2014.6.28	0.592	0.6
2014.6.13	0.593	0.592	2014.6.29	0.587	0.589
2014.6.14	0.586	0.59	2014.6.30	0.587	0.58
2014.6.15	0.585	0.586	2014.7.1	0.586	0.585
2014.6.16	0.587	0.586	2014.7.2	0.585	0.57

표에서 보는바와 같이 예측결과는 20일정도까지 그 정확성을 유지하고있다. 이것을 통하여 환자시세자료의 예측가능성을 확인할수 있다.

우리는 KNN예측방법을 리용한 환자시세의 예측가능성분석에 대하여 잘 알고 국제 금융거래를 주동적으로, 능란하게 벌려나감으로써 사회주의경제강국건설에 적극 이바지 하여야 할것이다.