

삼각퍼지수를 이용한 시계열모형

정혜영¹⁾, 최승희²⁾

요약

본 논문은 시간의 흐름에 따라 일정한 간격으로 관측된 시계열자료에 대한 통계적인 관계를 추정하기 위하여 삼각퍼지수를 이용한 퍼지시계열모형을 소개한다. 모든 관측치를 포함하는 전체집합을 분할하는 구간을 자료의 빈도수에 따라 결정하고 연속되는 두 시점에서 퍼지수가 일치하는 경우에는 관측된 자료의 차에 대한 정보를 이용하여 제안된 퍼지시계열모형을 추정한다. 예제를 이용하여 제안된 퍼지시계열모형의 정확성을 일반적인 시계열모형과 여러 가지 방법으로 추정된 퍼지시계열모형과 비교한다.

주요용어: 퍼지시계열모형; 퍼지논리; 삼각퍼지수.

1. 서론

시간의 흐름에 따라 일정한 간격으로 관측된 과거의 시계열자료를 이용하여 미래의 값을 예측하는 방법을 시계열분석이라 한다. 시계열분석은 정확하게 측정하여 실수로 기록된 시계열자료의 종속성이나 주기성에 근거하여 시계열모형을 추정한다. 시계열분석에서 사용되는 통계적인 모형은 시계열자료 수가 50개 이상일 때 추정 결과를 신뢰할 수 있다 (Tseng 등, 2001). 따라서 추정된 시계열모형의 정확성은 관측 자료의 수에 의존한다. 그러므로 주어진 시계열자료가 애매하여 문자로 표현되거나 혹은 자료의 수가 적은 경우에는 일반적인 시계열방법을 사용하여 추정할 수 없다.

관측된 시계열자료의 수가 충분하지 못하거나 혹은 실수로 표현할 수 없는 자료에 대한 패턴을 예측하기 위하여 Song과 Chissom (1993a)은 처음으로 퍼지시계열모형을 소개하였다. Sullivan (1994)은 퍼지시계열모형과 마코프모형을 비교하였고 Tseng 등 (2001)은 대만화폐교환율을 이용하여 시계열모형과 퍼지시계열모형에 대한 특징을 설명하였다. 퍼지수를 이용한 시계열모형은 대학 등록자수, 주식, 온도, 강수량, 경제지수 등 여러 분야에서 응용되고 있다.

퍼지시계열을 이용하여 미래의 값을 예측하기 위해 다음과 같은 절차가 필요하다.

- ① 관측된 자료에 대한 전체집합과 퍼지수 결정
- ② 퍼지수 사이의 퍼지논리관계 설정
- ③ 퍼지수 예측

1) (120-749) 서울시 서대문구 신촌동 134, 연세대학교 수학과, 강사. E-mail: hjung@yonsei.ac.kr.

2) (412-791) 경기도 고양시 화전동 200-1, 한국항공대학교 인문자연학부, 교수.

교신저자: shchoi@kau.ac.kr

관측된 자료를 퍼지화하기 위하여 관측된 모든 자료를 포함하는 전체집합을 구성하는 유한개의 폐구간을 결정하여야 한다. 이 때 사용된 폐구간의 길이와 구간의 수는 퍼지 시계열모형의 정확성에 영향을 준다. 따라서 퍼지시계열에서는 폐구간의 길이와 폐구간의 수를 적절히 조정하는 것이 중요하다. Chen (1996)과 Song과 Chissom (1994)은 길이가 동일한 폐구간을 이용하여 전체집합을 나누었고 Huarng (2001b)은 전체집합을 구성하는 폐구간의 수를 결정하기 위하여 관측치의 분포와 평균을 이용하였다. 그리고 Huarng (2001a)은 시계열모형의 정확성은 폐구간의 수에 의존함을 보였다.

퍼지시계열모형에서 사용되는 퍼지논리관계는 연속되는 두 시점에서 관측된 두 퍼지수의 관계를 설명한다. 두 퍼지수의 논리관계에 따라 퍼지수가 예측되므로 적절한 퍼지논리관계를 정의하고 사용하는 것은 퍼지시계열분석에서 중요하다. Chen (1996)과 Song과 Chissom (1993b)은 관측치를 퍼지화한 퍼지수를 이용하여 연속되는 두 시점 사이의 논리관계를 정의하였고 Huarng (2001b)은 경험적인 방법에 따라 관측치의 증가와 감소에 의존하는 퍼지논리관계를 제시하였다. 그리고 Yu (2005)와 Cheng 등 (2006)은 퍼지논리관계에 근거한 가중치를 이용하여 시계열모형을 추정하였다.

또한 Shah (2007)과 Huarng와 Yu (2003)은 고차 퍼지시계열모형을 소개하였고 Liu (2007)는 사다리꼴퍼지수를 이용하여 시계열모형을 설명하였다.

앞에서 언급된 모든 논문들은 관측치를 퍼지수로 변환하는 과정에서 서로 다른 퍼지수가 겹치는 방법으로 구간을 설정하였고, 연속하는 두 시점에 대한 퍼지수가 동일한 경우 관측된 자료의 변화에 대한 정보를 이용하지 않고 퍼지수를 예측하였다. 이와 같은 방법은 관측된 자료를 퍼지화하는 과정에서 하나의 관측치가 여러 개의 퍼지수로 표현될 수가 있다. 또한 관측된 시계열자료에서 유추할 수 있는 증가나 감소에 대한 정보가 자료를 퍼지화하는 과정에서 손실될 수 있다. 이것은 추정된 퍼지시계열모형의 정확성을 감소시키는 원인이 될 수 있다.

본 논문에서는 일정한 시간의 간격에 따라 관측된 자료에 대한 시계열모형을 추정하기 위하여 삼각퍼지수를 이용한 퍼지시계열을 소개한다. 제안된 퍼지시계열모형의 정확성을 높이기 위하여 모든 관측치를 포함하는 전체집합을 분할하는 구간을 자료의 빈도수에 따라 결정하고 연속되는 두 시점에서 퍼지수가 일치하는 경우에는 관측된 자료의 차에 대한 정보를 이용한다. 예제를 이용하여 제안된 퍼지시계열모형의 정확성을 일반적인 시계열모형과 여러 가지 방법으로 추정된 퍼지시계열모형과 비교한다.

2. 퍼지시계열

본 절에서는 시계열자료에 대한 모형을 추정하기 위해 삼각퍼지수를 이용한 시계열모형을 소개하고 제안된 퍼지시계열모형을 추정하기 위하여 7단계 방법을 설명한다. 애매한 자료나 문자로 표현된 시계열자료를 예측하기 위하여 Song과 Chissom (1993a)은 다음과 같은 정의를 이용하였다.

정의 2.1 퍼지시계열 $\{F(t) : t = 0, 1, 2, \dots\}$ 의 원소 $F(t)$ 는 퍼지집합 $A_i (i = 1, \dots, m)$ 의 조합으로 표현된다.

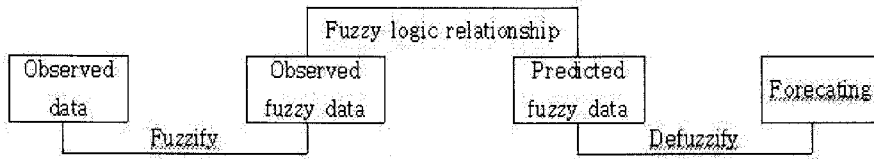


그림 2.1: 퍼지시계열 예측 과정

정의 2.2 연산자 \circ 에 대하여 $F(t) = F(t-1) \circ R(t, t-1)$ 을 만족하는 퍼지관계 $R(t, t-1)$ 이 존재하면 $F(t-1) \rightarrow F(t)$ 와 같이 표현한다.

정의 2.3 퍼지관계 $R(t, t-1)$ 이 시간 t 에 따라 변하는 경우와 변하지 않는 경우에 따라 $\{F(t)\}$ 를 각각 시불변퍼지시계열과 시변퍼지시계열이라 한다.

정의 2.4 $F(t-1) = A_i$ 이고 $F(t) = A_{t_j}$ 일 때 $F(t)$ 와 $F(t-1)$ 사이의 퍼지논리관계는 $A_i \rightarrow \{A_{t_j} : j = 1, \dots, k\}$ 와 같이 표현한다.

퍼지시계열의 목적은 주어진 시계열자료를 이용하여 미래를 정확하게 예측하는 것이다. 관측된 자료의 수가 충분하지 않거나 혹은 관측된 자료가 애매함이나 모호함을 포함하고 있는 시계열자료에 대한 퍼지시계열모형은 그림 2.1과 같은 순서에 따라 추정할 수 있다. 실수로 표현된 시계열자료는 퍼지수로 변형한 후 퍼지수사이의 퍼지논리관계성을 이용하여 퍼지수를 예측한다. 그리고 예측치를 추정하기 위하여 예측된 퍼지수를 실수로 변형하여야 한다. 시계열자료가 퍼지수인 경우에는 관측된 퍼지수사이의 퍼지논리관계성을 이용하여 퍼지시계열모형을 추정한다.

관측치를 퍼지수로 변형하기 위해서는 제일 먼저 모든 자료를 포함하는 전체집합을 결정한다. 다음으로 전체집합을 유한개의 폐구간으로 나누고 폐구간을 이용하여 관측치에 대한 퍼지수의 소속함수를 정의한다. 따라서 전체집합을 구성하는 폐구간의 수와 폐구간의 길이를 결정하는 것은 매우 중요하다. 특정한 폐구간에 포함된 관측치의 빈도수가 크면 다수의 관측치가 퍼지수로 변형되는 과정에서 동일한 소속함수를 갖는 경우가 발생한다. 따라서 서로 다른 다수의 관측치가 동일한 퍼지수로 대응되는 경우에는 특정한 구간을 다시 길이가 동일한 폐구간으로 나눈다. 그러면 서로 다른 관측치에 대한 퍼지수가 일치하는 문제를 해결할 수 있다. 또한 관측치를 퍼지화하는 과정에서 연속하는 두 시점에 대한 퍼지수가 동일한 경우가 발생할 수 있다. 현 시점과 전 시점에 대한 퍼지수가 동일한 경우는 변화가 없음을 의미하게 되는데 변동이 없는 시점이 많으면 많을수록 추정된 퍼지시계열모형의 정확성은 감소할 수 있다.

지금까지 발표된 퍼지시계열에 관한 많은 논문들은 하나의 관측치에 여러 개의 퍼지수가 대응될 수 있는 방법을 사용하였다. 또한 연속하는 두 시점에 대한 퍼지수가 동일한 경우에도 관측치에 대한 특별한 정보를 이용하지 않았다. 따라서 관측된 시계열의 차에 대한 정보를 이용하는 것이 추정된 퍼지시계열의 효율성을 높이는 방법이 될 수 있다.

본 논문에서는 일정한 시간에 따라 관측된 자료 $\{x_t : t = 1, 2, \dots\}$ 에 대한 퍼지시계열모형을 제시한다. 삼각퍼지수를 이용한 퍼지시계열모형을 추정하기 위하여 다음과 같은 절차가 필요하다.

Step 1: 차에 대한 예측.

관측된 자료의 차 $\{d_t : d_t = x_t - x_{t-1}\}$ 을 이용하여 시계열의 차에 대한 예측치 \hat{d}_t 을 추정한다.

Step 2: 전체집합.

관측된 모든 자료를 포함하는 전체집합 $U = [x_{(1)} - c, x_{(n)} + d]$ 는 실수의 부분집합이다. 여기서 c 와 d 는 전체집합의 양 끝점을 단순화하는 역할을 하는 수로서 양수이고 $x_{(1)}$ 과 $x_{(n)}$ 은 각각 관측치중 가장 작은 수와 가장 큰 수이다.

Step 3: 전체집합의 분할.

전체집합 U 에 속하는 점 $e_j (j = 0, 1, \dots, m+1)$ 는 $e_j < e_{j+1}$ 이고 양 끝점은 $e_0 = x_{(1)} - c$ 와 $e_{(m+1)} = x_{(n)} + d$ 이다. 따라서 폐구간 $I_j = [e_{j-1}, e_j]$ 은 $\bigcup_{j=1}^{m+1} I_j = U$ 을 만족한다. 또한 주어진 구간 I_j 에 속하는 관측치의 수 n_j 는 양의 정수 r 보다 적다. 즉, $n_j = \#\{x_t : x_t \in I_j\} \leq r$ 이다. 그리고 각 구간의 끝점 e_j 는 시계열자료의 집합 $\{x_t : t = 1, 2, \dots\}$ 에 포함되지 않는 수이다. 즉, 모든 시점 t 에 대하여 $x_t \neq e_j (j = 0, 1, \dots, m+1)$ 이다. 본 논문에서는 Step 3에서 정의된 구간을 만들기 위하여 구간을 길이가 일정한 구간으로 $[e_0, e_{m+1}]$ 나눈다. 그리고 특정한 구간 I_j 에 속하는 관측치의 수가 많으면 구간 I_j 를 다시 길이가 일정한 폐구간 $I_j^{(k)}$ 로 나눈다.

Step 4: 관측된 자료의 퍼지화.

관측된 자료에 대한 삼각퍼지수를 정의하기 위하여 폐구간 $I_k = [e_{k-1}, e_k]$ 의 중점을 $c_k = (e_{k-1} + e_k)/2$ 라 하자. 폐구간 I_k 를 이용하여 정의된 삼각퍼지수 A_k 의 소속함수 μ_{A_k} 는

$$\mu_{A_k}(x) = \begin{cases} \frac{x - e_{k-1}}{c_k - e_{k-1}}, & \text{if } e_{k-1} \leq x \leq c_k, \\ \frac{x - e_k}{c_k - e_k}, & \text{if } c_k \leq x \leq e_k, \\ 0, & \text{if otherwise} \end{cases}$$

이다. 이 때 삼각퍼지수 A_k 는 $(c_k, l_k, r_k)_T$ 와 같이 표현하고 삼각퍼지수의 폭 l_k 와 r_k 는 각각 $c_k - e_{k-1}$ 과 $e_k - c_k$ 이다. 그리고 삼각퍼지수 A_k 의 지지 $\{x : \mu_{A_k}(x) > 0\}$ 는 구간 I_k 이다. 시점 t 에서 관측된 x_t 가 구간 I_j 에 속하면 시점 t 에서 관측된 퍼지수는 A_j 라 하고 $F(t) = A_j$ 이다.

Step 5: 퍼지논리관계

두 시점 $t-1$ 과 t 에 대한 퍼지논리관계는 정의 2.4를 이용하여 유도한다. 만약 두 시점 $t-1$ 과 t 에 대한 퍼지수가 각각 $F(t-1) = A_i$ 이고 $F(t) = A_{i_t}$ 이면 두 시점에 대

한 퍼지논리관계는

$$A_i \rightarrow \{A_{i_t} : t = 1, 2, \dots, k, i_t \in \{1, 2, \dots, m + 1\}\}$$

이고 $k \leq m$ 이다.

Step 6: 예측퍼지수

두 삼각퍼지수 $A_p = (c_p, l_p, r_p)_T$ 와 $A_q = (c_q, l_q, r_q)_T$ 의 합은

$$A_p + A_q = (c_p + c_q, l_p + l_q, r_p + r_q)_T$$

이다. 또한 실수 b 와 삼각퍼지수 $A_p = (c_p, l_p, r_p)_T$ 의 곱은

$$b \cdot A_p = \begin{cases} (bc_p, bl_p, br_p)_T, & \text{if } b \geq 0, \\ (bc_p, |b|r_p, |b|l_p)_T, & \text{if } b < 0 \end{cases}$$

이다. 시점 $t - 1$ 에서 퍼지수가 A_i 일 때 시점 t 에서 예측퍼지수 \hat{A}_t 는

$$\hat{A}_t = \begin{cases} \frac{1}{w} \sum_{j \in R(i)} A_j, & \text{if } F(t-1) \neq F(t), \\ \frac{1}{w} \sum_{j \in R(i)} A_j + \hat{D}_t, & \text{if } F(t-1) = F(t) \end{cases} \quad (2.1)$$

이고 $\hat{F}(t) = \hat{A}_t$ 이다. 여기서 \hat{D}_t 는 중심은 \hat{d}_t 이고 폭이 0인 삼각퍼지수이고 w 는 집합 $R(i) = \{j : A_i \rightarrow A_j\}$ 의 원소 수이다.

Step 7: 예측퍼지수의 비퍼지화

관측치가 실수인 경우에는 퍼지시계열모형을 이용하여 추정된 퍼지수를 실수로 변형하여야 한다. 시점 t 에서 예측치 $\hat{x}_t(\alpha)$ 는 예측퍼지수 \hat{A}_t 의 α -수준집합 $\{x : \mu_{\hat{A}_t}(x) > \alpha\}$ 의 중점이다. 여기서 α 는 예측치의 신뢰도이다.

위에서 소개된 7단계를 이용하여 주어진 시계열자료 $\{x_t : t = 1, 2, \dots\}$ 에 대한 퍼지시계열모형을 추정할 수 있다. 본 논문에서 제안된 삼각퍼지수를 이용한 시계열모형을 추정하는 과정을 구체적으로 설명하기 위하여 Song과 Chissom (1994)이 소개한 대학 등록자 수와 Liu (2007)가 제안한 대만의 특허 수를 사용한 실제적인 예제를 다음 절에서 소개한다. 또한 논문에 제안된 방법을 이용하여 추정한 시계열모형의 정확성을 일반적인 방법으로 추정한 시계열 모형과 여러 가지 방법으로 추정한 시계열모형과 비교한다.

3. 퍼지시계열모형의 정확성

본 절에서는 삼각퍼지수를 이용한 퍼지시계열모형을 설명하기 위해 대학 등록자 수와 대만의 특허수와 같이 여러 논문에서 자주 언급된 시계열자료를 이용한다. 그리고 본 논문

에서 제안된 퍼지시계열모형의 정확성을 일반적인 시계열모형과 다른 방법으로 추정된 퍼지시계열모형과 비교한다.

일반적인 시계열분석에서는 예측모형의 정확성을 측정하기 위해 평균제곱오차와 평균절대비율오차를 사용한다. 퍼지시계열에서는 관측퍼지수 A_t 와 예측퍼지수 \hat{A}_t 에 대한 제곱오차와 절대비율오차를 각각

$$\int_0^1 [c\{A_t(\alpha)\} - c\{\hat{A}_t(\alpha)\}]^2 d\alpha$$

와

$$\int_0^1 \frac{|c\{A_t(\alpha)\} - c\{\hat{A}_t(\alpha)\}|}{|c\{A_t(\alpha)\}|} d\alpha$$

로 정의할 수 있다. 여기서 $c\{A_t(\alpha)\}$ 와 $c\{\hat{A}_t(\alpha)\}$ 는 각각 퍼지수 A_t 와 \hat{A}_t 의 α -수준집합의 중심이다. 따라서 추정된 퍼지시계열모형에 대한 정확성은 다음과 같은 퍼지평균제곱오차(Fuzzy Mean Square Error: FMSE)와 퍼지평균절대비율오차(Fuzzy Mean Absolute Percentage Error: FMAPE)를 이용하여 설명할 수 있다.

$$\text{FMSE} = \frac{1}{n} \sum_{t=1}^n \int_0^1 [c\{A_t(\alpha)\} - c\{\hat{A}_t(\alpha)\}]^2 d\alpha, \quad (3.1)$$

$$\text{FMAPE} = \frac{1}{n} \sum_{t=1}^n \int_0^1 \frac{|c\{A_t(\alpha)\} - c\{\hat{A}_t(\alpha)\}|}{|c\{A_t(\alpha)\}|} d\alpha. \quad (3.2)$$

만약 관측된 시계열이 모두 실수이면 주어진 자료의 오른쪽과 왼쪽 폭이 0이므로

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n (c_t - \hat{c}_t)^2 \quad (3.3)$$

와

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \frac{|c_t - \hat{c}_t|}{|c_t|} \quad (3.4)$$

는 각각 일반적인 시계열분석에서 사용하는 평균제곱오차와 평균절대비율오차와 일치한다.

추정된 퍼지수와 관측된 퍼지수의 소속함수가 만나는 부분의 면적이 넓으면 넓을수록 두 측도의 값은 감소한다. 그리고 평균제곱오차와 평균절대비율오차의 최소값은 0이므로 두 측도 (3.1)과 (3.2)의 값이 0에 근접하면 근접할수록 추정된 시계열모형은 더 정확하다고 할 수 있다. 따라서 두 측도 (3.1)과 (3.2)를 이용하여 추정된 시계열모형의 정확성을 비교할 수 있다.

다음 예제는 Liu (2007)가 제안한 대만의 특허 수에 대한 퍼지시계열모형을 삼각퍼지수를 이용하여 추정한다.

예제 3.1 사다리꼴퍼지수를 이용한 퍼지시계열모형을 설명하기위해 Liu (2007)는 표 3.1의 자료를 제안하였다. 특허 수에 대한 예측치를 구하기 위해 앞 절에서 제시된 7단계 방법을 이용하면 다음과 같다.

Step 1: 표 3.1에 제시된 두 관측치의 차에 대한 예측치는 경험적 퍼지시계열을 이용하여 얻은 결과이다 (참고 Huarng, 2001b).

Step 2: 2000년 특허 수는 $38,665(x_{(21)})$ 으로 가장 많고 1981년 특허 수는 $6,264(x_{(1)})$ 으로 가장 적다. 두 양수 $246(c)$ 와 $1,335(d)$ 을 이용하여 천 단위를 조정하면 전체 집합은

$$U = [6246 - 264, 38665 + 1335] = [6000, 40000]$$

이다

Step 3: 폐구간 $I_j = [e_{j-1}, e_j]$ 의 점 $e_j = 6000 + 2000j$ 이다. 관측치를 가장 많이 포함하고 있는 폐구간은 I_1 와 I_{12} 이고 구간 I_{12} 에 속하는 관측치의 수(n_{12})는 네 개다. 따라서 $\bigcup_{j=1}^{17} I_j = U$ 이고 $\max\{n_j : j = 1, \dots, 17\} \leq 4$ 이다.

Step 4: 삼각퍼지수 $A_j = (5000 + 2000j, 1000, 1000)_T$ 의 지지는 폐구간 I_j 이다. 그리고 시점 t 에서 퍼지수 $F(t)$ 는 관측치(x_t)에 의존한다. 예를 들면 1982년과 1997년의 특허 수는 각각 $7460(x_3)$ 과 $29356(x_{18})$ 이므로 $x_3 \in I_1$ 이고 $x_{18} \in I_{12}$ 이다. 따라서 1982년과 1997년에 대한 퍼지수는 각각 $F(3) = A_1$ 와 $F(18) = A_{12}$ 이다.

Step 5: 표 3.1에 제시된 시계열자료로부터 유도한 각 시점의 특허 수에 대한 퍼지논리관계(Fuzzy Logic Relationship: FLR)는 표 3.2와 같다.

Step 6: 표 3.1과 3.2에서 $A_7 \rightarrow \{A_9, A_{12}\}$ 와 $F(10) \neq F(11)$ 이므로 식 (2.1)로부터 시점 11(1990년)에 대한 예측퍼지수는

$$\hat{F}(11) = \frac{1}{2}(A_9 + A_{12})$$

이다. 또한 시점 3과 4의 퍼지수는 $F(3) = F(4) = A_1$ 이고 퍼지수 A_1 에 대한 퍼지논리관계는 $A_1 \rightarrow \{A_1, A_2\}$ 이므로 시점 4(1983년)에 대한 추정치는

$$\hat{F}(4) = \frac{1}{2}(A_1 + A_2) + (-375, 0, 0)_T$$

이다. 표 3.2의 퍼지논리관계와 식 (2.1)을 이용하여 표 3.1의 예측퍼지수를 추정할 수 있다. Step 2에서 정의된 폐구간의 길이가 일정함으로 표 3.1의 예측퍼지수의 폭도 일정하다.

Step 7: 매년 허가된 특허 수(x_t)는 정수이므로 예측치도 정수로 표현하여야 한다. 표 3.1의 예측치는 예측퍼지수 $\hat{F}(t) = \hat{A}_t$ 의 α -수준집합에 대한 중심값이다. 모든 예측퍼지수가 대칭이므로 중심값은 α 에 의존하지 않는다.

표 3.1: 특허 수에 대한 자료

Year	Patents	Fuzzy number	Forecasted difference	Forecasted fuzzy number	Forecasted patent
1980	6633	A_1			
1981	6264	A_1	-369	$(7631, 1000, 1000)_T$	7631
1982	7460	A_1	1281	$(9281, 1000, 1000)_T$	9281
1983	7096	A_1	-375	$(7625, 1000, 1000)_T$	7625
1984	8592	A_2	1281	$(8000, 1000, 1000)_T$	8000
1985	9427	A_2	500	$(10500, 1000, 1000)_T$	10500
1986	10526	A_3	1344	$(10000, 1000, 1000)_T$	10000
1987	10615	A_3	-1500	$(10500, 1000, 1000)_T$	10500
1988	12355	A_4	1344	$(12000, 1000, 1000)_T$	12000
1989	19265	A_7	6500	$(19000, 1000, 1000)_T$	19000
1990	22601	A_9	3500	$(26000, 1000, 1000)_T$	26000
1991	27281	A_{11}	4500	$(23000, 1000, 1000)_T$	23000
1992	21264	A_8	1500	$(21000, 1000, 1000)_T$	21000
1993	22317	A_9	1063	$(23000, 1000, 1000)_T$	23000
1994	19032	A_7	-1500	$(23000, 1000, 1000)_T$	23000
1995	29707	A_{12}	10500	$(26000, 1000, 1000)_T$	26000
1996	29469	A_{12}	-375	$(30625, 1000, 1000)_T$	30625
1997	29356	A_{12}	-2312	$(28688, 1000, 1000)_T$	28688
1998	25051	A_{10}	-2312	$(31000, 1000, 1000)_T$	31000
1999	29144	A_{12}	4500	$(29000, 1000, 1000)_T$	29000
2000	38665	A_{17}	1500	$(31000, 1000, 1000)_T$	31000

표 3.2: 특허에 대한 퍼지논리관계

Index	FLR	Index	FLR
1	$A_1 \rightarrow \{A_1, A_2\}$	6	$A_8 \rightarrow \{A_9\}$
2	$A_2 \rightarrow \{A_2, A_3\}$	7	$A_9 \rightarrow \{A_7, A_{11}\}$
3	$A_3 \rightarrow \{A_3, A_4\}$	8	$A_{10} \rightarrow \{A_{12}\}$
4	$A_5 \rightarrow \{A_7\}$	9	$A_{11} \rightarrow \{A_8\}$
5	$A_7 \rightarrow \{A_9, A_{12}\}$	10	$A_{12} \rightarrow \{A_{10}, A_{12}, A_{17}\}$

다음 예제는 Liu (2007)가 사용한 특허 수에 대한 시계열모형의 정확성을 비교한다.

예제 3.2 삼각퍼지수를 이용하여 대만의 특허 수에 대한 시계열모형을 예제 3.1에서 추정하였다. Liu (2007)는 특허 수에 대한 퍼지시계열모형을 추정하기 위하여 사다리꼴퍼지수를 사용하였다. 그리고 자신의 결과를 Chen (1996), Hwang 등 (1998), Lee와 Chou (2004) 등 여러 사람들이 제시한 퍼지시계열모형과 비교한 후 자신이 제시한 퍼지시계열모형이 더 효율적임을 보였다. 표 3.1에 제시된 특허 수에 대한 일반적인 방법의 시계열모형을 추정하기 위하여 관측치를 로그 변환하여 얻은 결과는

$$\ln x_t = 9.04914 + \ln x_{t-1}$$

표 3.3: 추정된 모형의 정확성

Index	Ordinal method	Chen's method	Hwang's method	Lee's method	Liu's method	Proposed method
MSE	17220953	14524448	25021153	18172556	8378697	8165093
MAPE(%)	15.08	18.94	13.67	19.44	10.45	10.01

와 같은 AR(1)모형이다. 여기서 $\ln x = \log_e x$ 이다.

표 3.3은 Liu (2007)가 소개한 네 가지 모형에 대한 두 척도 (3.1)과 (3.2)를 보여준다. 삼각퍼지수를 이용한 시계열모형에 대한 평균제곱오차(MSE)와 평균절대비율오차(MAPE)가 일반적인 시계열모형과 여러 가지 방법으로 제시된 퍼지시계열모형의 오차값보다 크지 않음을 표 3.3은 보여준다. 이것은 관측치에 대한 차를 이용하는 것이 추정된 시계열모형의 정확성을 높일 수 있음을 보여준다.

다음 예제는 Song과 Chissom (1993a)이 제안한 대학 등록자 수에 대한 퍼지시계열모형을 추정하는 과정을 설명한다.

예제 3.3 Song과 Chissom (1993a)은 퍼지시계열모형을 처음으로 제안하고 표 3.4에 소개된 대학 등록자 수에 대한 퍼지시계열모형을 추정하였다. 이후 많은 사람들이 예측된 등록자 수와 표 3.4에서 주어진 등록자 수 사이의 오차를 줄이기 위한 여러 방법들을 소개하였다. 2절에서 소개된 7단계 방법을 이용하여 등록자 수에 대한 퍼지시계열모형을 구하는 과정은 다음과 같다.

Step 1: 등록자 수에 대한 차($d_t = x_t - x_{t-1}$)에 대한 추정치(\hat{d}_t)는 예제 3.1과 같이 경험적 퍼지시계열방법을 이용하여 추정하였다.

Step 2: 등록자 수의 최대값은 19,337($x_{(22)}$)이고 최소값은 13,055($x_{(1)}$)이므로 두 양수 55(c)와 163(d)을 이용하면 전체집합은 $U = [13055 - 55, 19337 + 163] = [13000, 19500]$ 이다.

Step 3: 폐구간 $I_j = [e_{j-1}, e_j]$ 의 끝점을 $e_j = 13000 + 500j$ 라 하면 $\bigcup_{j=1}^{13} I_j = U$ 이다. 또한 6개의 관측치가 폐구간 I_5 에 포함되고 폐구간 I_6 와 I_8 은 3개의 관측치를 포함하고 있다. 한 구간에 많은 관측치가 포함되면 시계열자료에 대한 증가나 감소에 대한 정보를 이용할 수 없기 때문에 추정된 시계열모형의 정확성이 떨어질 수 있다. 따라서 폐구간 I_6 을 길이가 250인 폐구간 $I_6^{(k)} = [15500 + 250(k-1), 15500 + 250k]$ ($k = 1, 2$)을 이용하여 나누고 폐구간 I_8 을 네개의 폐구간으로 나누기 위해 폐구간 $I_8^{(k)} = [16500 + 125(k-1), 16500 + 125k]$ ($k = 1, 2, 3, 4$)을 이용한다. 그리고 폐구간 I_5 를 네개의 폐구간 $I_5^{(k)}$ ($k = 1, 2, 3, 4$)로 나누었다. 각각의 폐구간은 $I_5^{(1)} = [15000, 15250]$, $I_5^{(2)} = [15250, 15375]$, $I_5^{(3)} = [15375, 15438]$, $I_5^{(4)} = [15438, 15500]$ 와 같다. 그러면 각 폐구간에 속하는 최대의 관측치의 수는 두 개($r = 2$)다. 즉, $\max\{n_j : j = 1, \dots, 20\} \leq 2$ 이다.

표 3.4: 특허에 대한 퍼지논리관계

Year	Enrollment	Fuzzy number	Forecasted difference	Forecasted fuzzy number	Forecasted enrollment
1971	13055	A_1			
1972	13563	A_2	508	$(13750, 250, 250)_T$	13750
1973	13847	A_2	150	$(14400, 250, 250)_T$	14400
1974	14696	A_4	675	$(14250, 250, 250)_T$	14250
1975	15460	$A_5^{(4)}$	641.7	$(15406.5, 31, 31)_T$	15406.5
1976	15311	$A_5^{(2)}$	-100	$(15218.8, 125, 125)_T$	15218.8
1977	15603	$A_6^{(1)}$	150	$(15625, 125, 125)_T$	15625
1978	15861	$A_6^{(2)}$	675	$(15875, 125, 125)_T$	15875
1979	16807	$A_8^{(3)}$	675	$(16812.5, 62.6, 62.5)_T$	16812.5
1980	16919	$A_8^{(4)}$	150	$(17593.8, 250, 250)_T$	17593.8
1981	16388	A_7	-500	$(15406.5, 31.5, 31.5)_T$	16250
1982	15433	$A_5^{(3)}$	-900	$(16250, 250, 250)_T$	15406.5
1983	15497	$A_5^{(4)}$	50	$(15469, 31, 31)_T$	15469
1984	15145	$A_5^{(1)}$	8.3	$(15218.5, 94.5, 94.5)_T$	15218.8
1985	15163	$A_5^{(1)}$	50	$(15175, 125, 125)_T$	15175
1986	15984	$A_6^{(2)}$	8.3	$(15500, 125, 125)_T$	15500
1987	16859	$A_8^{(3)}$	641.7	$(16812.5, 62.5, 62.5)_T$	16812.5
1988	18150	A_{11}	1300	$(17593.8, 156.3, 156.3)_T$	17593.8
1989	18970	A_{12}	825	$(18750, 250, 250)_T$	18750
1990	19328	A_{13}	641.7	$(19250, 250, 250)_T$	19250
1991	19337	A_{13}	50	$(19300, 250, 250)_T$	19300
1992	18876	A_{12}	8.3	$(19000, 250, 250)_T$	19000

Step 4: 폐구간 I_j 와 $I_j^{(k)}$ 에 대응하는 삼각퍼지수는 각각 A_j 와 $A_j^{(k)}$ 이다. 시점 t 에서 퍼지수 $F(t)$ 는 관측치 x_t 에 속하는 폐구간에 따라 결정한다. 예를 들면 1982년의 등록자 수는 15,433(x_{12})이고 $x_{12} \in I_5^{(3)}$ 이므로 시점 12에서 퍼지수는 $F(12) = A_5^{(3)}$ 이다. 표 3.4은 각 시점의 관측치를 퍼지수로 변형한 결과이다.

Step 5: 표 3.4에 근거한 각 시점의 등록자 수에 대한 퍼지논리관계(Fuzzy Logic Relationship: FLR)는 표 3.5와 같다.

Step 6: 시점 2 (1972년)와 3 (1973년), 14 (1984년)와 15 (1985년), 20 (1990년)과 21 (1991년)에서 퍼지수가 변화가 없으므로 차에 대한 예측치를 이용하여 예측 퍼지수를 추정한다. 식 (2.1)과 $A_5^{(1)} \rightarrow \{A_5^{(1)}, A_6^{(2)}\}$ 로부터 시점 15(1985년)에 대한 예측퍼지수는

$$\hat{F}(15) = \frac{1}{2} (A_5^{(1)} + A_6^{(2)}) + (50, 0, 0)_T$$

이다. 표 3.4와 식 (2.1)을 이용하여 등록자 수에 대한 예측퍼지수를 표 3.2와 같이 추정하였다. 전체집합의 부분집합인 폐구간의 길이가 다르므로 예측퍼지수의 폭도 다르다.

표 3.5: 등록자에 대한 퍼지논리관계

Index	FLR	Index	FLR
1	$A_1 \rightarrow \{A_2\}$	8	$A_6^{(1)} \rightarrow \{A_6^{(2)}\}$
2	$A_2 \rightarrow \{A_2, A_4\}$	9	$A_6^{(2)} \rightarrow \{A_8^{(3)}\}$
3	$A_4 \rightarrow \{A_5^{(4)}\}$	10	$A_7 \rightarrow \{A_5^{(3)}\}$
4	$A_5^{(1)} \rightarrow \{A_5^{(1)}, A_6^{(2)}\}$	11	$A_8^{(3)} \rightarrow \{A_8^{(4)}, A_{11}\}$
5	$A_5^{(2)} \rightarrow \{A_6^{(1)}\}$	12	$A_8^{(4)} \rightarrow \{A_7\}$
6	$A_5^{(3)} \rightarrow \{A_5^{(4)}\}$	13	$A_{11} \rightarrow \{A_{12}\}$
7	$A_5^{(4)} \rightarrow \{A_5^{(1)}, A_5^{(2)}\}$	14	$A_{12} \rightarrow \{A_{13}\}$
15	$A_{13} \rightarrow \{A_{12}, A_{13}\}$		

표 3.6: 추정된 모형의 정확성

Index	Ordinal method	Chen's method	Hwang's method	Lee's method	Liu's method	Proposed method
MSE	220536	398051	299634	255227	108097	89565
MAPE(%)	2.41	3.08	2.94	2.69	1.33	1.32

Step 7: 표 3.4의 등록자 수는 실수이므로 예측치도 실수로 표현하여야 한다. 등록자 수 (x_t) 에 대한 예측치는 예측퍼지수 $\hat{F}(t) = \hat{A}_t$ 의 α -수준집합에 대한 중심값이다. 표 3.4의 예측퍼지수가 대칭이므로 예측치는 α 에 상관없이 일정하다.

다음 예제는 여러 방법으로 추정된 대학 등록자수에 대한 시계열모형의 정확성을 비교한다.

예제 3.4 퍼지시계열을 설명하기 위한 자료로 소개된 대학 등록자 수에 대한 시계열모형은 Chen (1996), Hwang 등 (1998), Lee와 Chou (2004) 등 많은 사람들에 의해 소개되었다. 이렇게 많은 사람들에 의해 다루어진 대학 등록자 수에 대한 일반적인 방법의 시계열모형은 1차 차분계열에 의해 AR(1)모형이 얻어지므로 원시계열인 등록자 수에 대한 시계열모형은

$$x_t = 124.4836 + 1.53171x_{t-1} - 0.53171x_{t-2}$$

와 같다. Liu (2007)는 사다리꼴퍼지수를 이용한 퍼지시계열모형을 대학 등록자 수에 응용하여 자신의 결과와 다른 사람들의 결과를 비교하였다. 표 3.6은 Chen (1996), Hwang 등 (1998), Lee와 Chou (2004), Liu (2007)가 추정된 시계열모형과 일반적인 시계열모형에 대한 두 척도 (3.1)과 (3.2)를 보여준다. 표 3.6을 통해 여러 가지 방법으로 추정된 시계열모형에 대한 평균제곱오차(MSE)와 평균절대비율오차(MAPE)가 본 논문에서 추정된 시계열모형의 오차 값보다 적지 않음을 알 수가 있다. 이것은 전체집합의 부분집합인 폐구간에 속하는 관측치의 수를 적절히 분산하고 기록된 자료로부터 유도할 수 있는 변화의 정보를 이용하는 것이 추정된 시계열모형의 정확성을 높일 수 있음을 보여준다.

4. 결론

본 논문에서는 시계열자료에 대한 모형을 추정하기 위하여 삼각퍼지수를 이용한 퍼지 시계열을 소개하였다. 특정한 구간에 속하는 자료의 빈도수에 근거하여 전체집합을 유한 개의 폐구간으로 나누었다. 다음으로 전체집합의 부분집합인 폐구간을 이용하여 삼각퍼지수를 정의하였다. 또한 연속되는 두 시점에서 퍼지수가 일치하는 경우에는 관측된 자료의 차에 대한 정보를 이용하여 제안된 퍼지시계열모형을 추정하였다. 기록된 자료의 차에 대한 정보를 이용하고 관측치의 빈도수에 따라 폐구간의 수와 길이를 조절하는 것이 추정된 시계열모형의 정확성을 증가시킬 수 있음을 확인하였다. 예제를 이용하여 제안된 퍼지시계열모형에 대한 정확성이 자료의 수와 종속성에 영향을 받는 일반적인 시계열모형과 여러 가지 방법으로 추정된 퍼지시계열모형보다 좋음을 보였다.

참고문헌

- Chen, S. M. (1996). Forecasting enrollments based on fuzzy time series, *Fuzzy Sets and Systems*, **81**, 311-319.
- Cheng, C. H., Chen, T. L. and Chiang, C. H. (2006). Trend-weighted fuzzy time series model for TAIEX forecasting, *Lecture Notes in Computer Science*, **4234**, 469-477.
- Huang, K. (2001a). Effective lengths of intervals to improve forecasting in fuzzy time series, *Fuzzy Sets and Systems*, **123**, 387-394.
- Huang, K. (2001b). Heuristic models of fuzzy time series for forecasting, *Fuzzy Sets and Systems*, **123**, 369-386.
- Huang, K. and Yu, H. K. (2003). An N-th order heuristic fuzzy time series model for TAIEX forecasting, *International Journal of Fuzzy Systems*, **5**, 247-253.
- Hwang, J. R., Che, S. M. and Lee, C. H. (1998). Handling forecasting problems using fuzzy time series, *Fuzzy Sets and Systems*, **100**, 217-218.
- Lee, H. S. and Chou, M. T. (2004). Fuzzy forecasting based on fuzzy time series, *International Journal of Computer Mathematics*, **81**, 781-789.
- Liu, H. T. (2007). An improved fuzzy time series forecasting method using trapezoidal fuzzy numbers, *Fuzzy Optimization Decision Making*, **6**, 63-80.
- Shah, M. (2007). Fuzzy time series: A realistic method to forecast gross domestic capital of India, *Analysis and Design of Intelligent Systems using Soft Computing Techniques*, **41**, 255-264.
- Song, Q. and Chissom, B. S. (1993a). Forecasting enrollments with fuzzy time series-part I, *Fuzzy Sets and Systems*, **54**, 1-9.
- Song, Q. and Chissom, B. S. (1993b). Fuzzy time series and its models, *Fuzzy Sets and Systems*, **54**, 269-277.
- Song, Q. and Chissom, B. S. (1994). Forecasting enrollments with fuzzy time series-part II, *Fuzzy Sets and Systems*, **62**, 1-8.
- Sullivan, J. (1994). A comparison of fuzzy forecasting and Markov modeling, Ph. D. Dissertation, The University of Alabama.
- Tseng, F. M., Tzeng, G. H., Yu, H. C. and Yuan, B. J. C. (2001). Fuzzy ARIMA model for forecasting the foreign exchange market, *Fuzzy sets and Systems*, **118**, 9-19.

Yu, H. K. (2005). Weighted fuzzy time series models for TAIEX forecasting, *Physica A: Statistical Mechanics and Its Applications*, **349**, 609–624.

[2008년 4월 접수, 2008년 5월 채택]

Time Series Using Fuzzy Logic

Hyeyoung Jung¹⁾, Seunghoe Choi²⁾

Abstract

In this paper we introduce a time series model using the triangle fuzzy numbers in order to construct a statistical relation for the data which is a sequence of observations which are ordered in time. To estimate the proposed fuzzy model we split of a universal set includes all observation into closed intervals and determine a number and length of the closed interval by the frequency of events belong to the interval. Also we forecast the data by using a difference between observations when the fuzzified numbers equal at successive times. To investigate the efficiency of the proposed model we compare the ordinal and the fuzzy time series model using examples.

Keywords: Fuzzy time series; fuzzy logic; triangle fuzzy number.

1) Lecture, Department of Mathematics, Yonsei University, Seoul 120-749, Korea.
E-mail: hjung@yonsei.ac.kr.

2) Professor, School of Liberal arts and Sciences, Korea Aerospace University, Koyang, Kyungkido 412-791, Korea. Correspondence: shchoi@kau.ac.kr