ELSEVIER
Fuzzy Sets and Systems 123 (2001) 387–394
FUZZY sets and systems
www.elsevier.com/locate/fss

# Effective lengths of intervals to improve forecasting in fuzzy time series

Kunhuang Huarng

*Department of Finance, Chaoyang University of Technology, 168 GiFeng E. Rd., WuFeng, Taichung County, Taiwan, ROC*

Received 14 December 1998; received in revised form 12 November 1999; accepted 9 March 2000

**Abstract**

Length of intervals affects forecasting results in fuzzy time series. Unfortunately, the issue of how to determine effective lengths of intervals has not been touched in previous studies. This study proposes distribution- and average-based length to approach this issue. Distribution-based length is the largest length smaller than at least half the first differences of data. Average-based length is set to one half the average of the first differences of data. Empirical analyses show that distribution- and average-based lengths are simple to calculate and can greatly improve forecasting results; in particular, they are superior to the randomly chosen lengths used in previous studies. 

*Keywords:* Enrollments; Forecasting; Fuzzy sets; Fuzzy time series; Stock market indices

## 1. Introduction

The concept of fuzzy time series was first proposed by Song and Chissom [2]. Since then, several other studies have appeared [1,3,4]. However, there are still many critical issues open. The determination of effective lengths of intervals is one of these.

Length of intervals greatly affects forecasting results in fuzzy time series. Hence, an effective length of intervals can significantly improve the forecasting results. This study proposes distribution- and average-based length to approach this issue. Chen's model gave the best results among previous studies [1,3,4] and hence is selected as the target for comparison. The yearly data on enrollments at the University of Alabama as well as the daily data from Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) are used to demonstrate the impact of effective lengths of intervals on forecasting results. Empirical analyses show that both distribution- and average-based lengths are simple to calculate and can greatly improve on the forecasting results obtained from previous models.

Section 2 briefly introduces fuzzy time series. Section 3 explains the relevant definitions of lengths of intervals and proposes approaches to determine effective lengths. Sections 4 and 5 compare the forecasting results of university enrollments and TAIEX by various lengths of intervals. Section 6 offers some conclusions.

## 2. Fuzzy time series

Let $U$ be the universe of discourse, where $U=\{u_1,u_2,\ldots,u_n\}$. A fuzzy set $A_i$ of $U$ is defined by

$$A_i = f_{Ai}(u_1)/u_1 + f_{Ai}(u_2)/u_2 + \cdots + f_{Ai}(u_n)/u_n,$$

*E-mail address:* huarng@mail.cyut.edu.tw (K. Huarng).

0165-0114/01/$ - see front matter 
PII: S0165-0114(00)00057-9

where $f_{Ai}$ is the membership function of fuzzy set $A_i$, $f_{Ai}: U \to [0,1]$. $u_k$ is the element of fuzzy set $A_i$, and $f_{Ai}(u_k)$ is the degree of belongingness of $u_k$ to $A_i$. $f_{Ai}(u_k) \in [0,1]$ where $1 \leqslant k \leqslant n$.

The concept of fuzzy time series was first proposed by Song and Chissom [2]:

**Definition 1.** $Y(t)$ $(t = \ldots, 0, 1, 2, \ldots)$, is a subset of $R$. Let $Y(t)$ be the universe of discourse defined by fuzzy set $f_i(t)$. If $F(t)$ consists of $f_i(t)$ $(i = 1, 2, \ldots)$, $F(t)$ is defined as a fuzzy time series on $Y(t)$ $(t = \ldots, 0, 1, 2, \ldots)$.

Following Definition 1, relevant definitions are proposed.

**Definition 2.** If there exists a fuzzy relationship $R(t-1, t)$, such that $F(t) = F(t-1) \times R(t-1, t)$ where $\times$ represents an operator, then $F(t)$ is said to be caused by $F(t-1)$. (Note that the operator can be either max–min [3], min–max [4], or arithmetic operator [1].) When

$$F(t-1) = A_i \quad \text{and} \quad F(t) = A_j,$$

the relationship between $F(t-1)$ and $F(t)$ (called a fuzzy logical relationship in [3]) is denoted by

$$A_i \to A_j.$$

**Definition 3.** Fuzzy logical relationships with the same fuzzy set on the left-hand side can be further grouped into a fuzzy logical relationship group [1]. Suppose there are fuzzy logical relationships such that

$$A_i \to A_{j1},$$
$$A_i \to A_{j2},$$
$$\ldots$$

They can be grouped into a fuzzy logical relationship group

$$A_i \to A_{j1}, A_{j2}, \ldots$$

Following Chen's model, the same fuzzy sets can only show up once on the right-hand side of the fuzzy logical relationship group.

## 3. Lengths of intervals

For enrollment forecasting, Song and Chissom choose 1000 as the length of intervals, without specifying any reason [3]. Since then, 1000 has been used as the length of intervals in further studies [4,1]. How the lengths of intervals affect forecasting results was left unanswered in these studies. In fact, different lengths of intervals may lead to different forecasting results.

A time-series forecasting example is given with two lengths of intervals to show that different lengths of intervals may lead to different forecasting results and forecasting errors.

Suppose we have the following time-series data:

$$6, 10, 12, 6, 4$$

The range $U$ is set as $[3, 13]$. If the length of intervals is chosen as 5, there are two intervals: $u_1 = [3, 8]$ and $u_2 = [8, 13]$. According to Chen's model, the forecasting mean squared error (MSE) is 10. On the other hand, if the length of interval is set to 2, there are five intervals: $u_1 = [3, 5]$, $u_2 = [5, 7]$, $u_3 = [7, 9]$, $u_4 = [9, 11]$, and $u_5 = [11, 13]$. The MSE is 4.5. Obviously, different lengths of intervals result in different forecasting errors.

Hence, the determination of the lengths of intervals, especially effective ones, is never trivial in the forecasting of fuzzy time series. An efficient way to choose effective lengths of intervals is therefore critical to improve forecasting in fuzzy time series. A key point in choosing effective lengths of intervals is that they should not be too large or small. When an effective length of intervals is too large, there will be no fluctuations in the fuzzy time series. On the other hand, when the length is too small, the meaning of fuzzy time series will be diminished. In oder to reflect fluctuations properly and to keep fuzzy time series meaningful, the heuristic is set in such a way that at least half the fluctuations in the time series are reflected by the effective lengths of intervals.

The fluctuations in fuzzy time series can be represented by the absolute value of the first differences of any two consecutive data (the first differences hereafter). Hence, the heuristic can reflect at least half the first differences. Based on this idea, two approaches are proposed: distribution- and average-based length. Distribution-based length is calculated