ISSN 1 746-7233, England, UK
World Journal of Modelling and Simulation
Vol. 4 (2008) No. 2, pp. 104-111

Average-based fuzzy time series models for forecasting Shanghai compound index*

Sun Xihao[1,2] , Li Yimin[1†]

[1] Faculty of Science, Jiangsu University, China
[2] Lake Tai College, Southern Yangtze University, China

(*Received July 3 2007, Accepted November 14 2007*)

Abstract. Effective lengths of intervals affect forecasting results in fuzzy time series. In this paper, we introduce the average-based lengths of interval and applied to Chen's model. By using the Shanghai compound index as the forecasting target, the empirical results show that the average-based model can improve the fuzzy time series forecasting.

Keywords: fuzzy time series, forecast, Shanghai compound index, average-based

1 Introduction

It is obvious that forecasting activities play important role in our daily life. We usually forecast many things concerned with our daily life, such as economy, stock index, population growth, weather, etc. To make a forecast with 100% accuracy may be impossible, but we can try our best to reduce the forecasting errors[4]. To solve the forecasting problems, many researchers have proposed many different methods and models, such as uni-variant (using past data), multi-variant (using the relationship among many variants) and qualitative analysis (using researcher's judgment)[13]. However, the time series is widely used. The conventional time series models, one particularly important group of models has been the family of auto regressive integrated moving average (ARIMA) models[1].However, there are many conditions that need to be fulfilled to successfully apply these models. For example, at least 50 and preferably more than 100 observations are required to apply ARIMA[11], and zero mean and zero variance are required to apply autoregressive (AR) models[2], etc. Meanwhile, the forecasts obtained by these models may easily encounter the problem of over-fitting[5]. Furthermore, these models have been applied to model precise observations.

Fuzzy set theory was originally developed to handle problems involving human linguistic terms[16]. Because the existing statistical time series methods could not effectively analyze time series with small amounts of data and historical data, fuzzy time series methods were developed. Song and Chissom proposed a first-order time-invariant model and a time-variant model of fuzzy time series in 1993[8]. They fuzzified the enrollment at the University of Alabama in 1993 in the first application of fuzzy time series to forecasting[7]. Then, in 1994, they proposed a new fuzzy time series and compared three different defuzzification models[9]. Chen considered that the Song and Chissom's method is too complicated to apply; he therefore presented arithmetic operations instead of the logic max-min composition[3]. The arithmetic operations have a robust specification and are superior to those applied in Song and Chissom model. Following these definitions, fuzzy time series models have been proposed for various applications, such as enrollment[3, 6–9], stock indices[6, 14, 15], reactors[12], and temperature forecasting[4].

* The authors would like to acknowledge the supported by the National Natural Science key Foundation of China (60234010).
† *E-mail address*: llym@ujs.edu.cn.

However, the issue of lengths of intervals had never been discussed until the effective lengths of intervals were proposed[6]. That study showed that different lengths of intervals may result in different fuzzy relationships, and in turn different forecasting results. And the effective lengths of intervals did improve forecasting results in various empirical analyses. Thus, we introduce the average-based lengths to improve the forecasting.

The goal of this study is to propose average-based fuzzy time series models to improve fuzzy time series forecasting, with the daily stock index in Shanghai being used in the empirical analysis. The rest of this paper is organized as follows. In Section 2, the concepts of fuzzy time series are reviewed. In Section 3, we explain the relevant definitions of lengths of intervals and propose the average-based fuzzy time series model. Section 4 elaborates on the use of model in forecasting Shanghai Compound Index. Section 5 evaluates the models' performance and Section 6 concludes the paper.

2 Fuzzy time series revisited

Song and Chissom first proposed the definitions of fuzzy time series in 1993[8]. The concepts of fuzzy time series are described as follows.

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

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

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

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 $F(y)$ consisted of $f_1(t), f_2(t), \ldots$, $F(t)$ is defined as a fuzzy time series on $Y(u)(t = \ldots, 0, 1, 2, \ldots)$.*

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 $F(t-1)$.*

Let $F(t) = A_i$ and $F(t-1) = A_j$. The relationship between $F(t)$ and $F(t-1)$ (referred to as a fuzzy logical relationship, FLR) can be denoted by $A_i \rightarrow A_j$; where A_i is called the left-hand side (LHS) and A_j the right-hand side (RHS) of the FLR.

Definition 3. *Given two FLRs with the same fuzzy sets on the LHSs $A_i \rightarrow A_{j1}, A_i \rightarrow A_{j2}$. Both FLRs can be grouped together into fuzzy logical relationship groups (FLRG) $A_i \rightarrow A_{j1}, A_{j2}$.*

Different fuzzy time-series models have been proposed by following Song and Chissom's definition of fuzzy time-series. For example, there have been Song and Chissom's time-invariant model[7], Song and Chissom's time-variant model, the Markov model[10] Chen's model[3], weighted model[15], etc.

It is common for these models to include the following steps: (1) define the universe of discourse and the intervals for the observations; (2) partition the universe based on the intervals; (3) define the fuzzy sets for the observations; (4) fuzzify the observations; (5) establish the fuzzy logic relationship and fuzzy logic relationship group; (6) perform the forecast; and (7) defuzzify the forecasting results.

3 Average-based fuzzy time series model

3.1 Lengths of intervals

In previous models, all the lengths of the intervals were determined at Step 1 of the forecasting process. The interval length affects the formulation of fuzzy relationships, and the fuzzy relationships affect the forecasting results. In addition to capture proper fuzzy relationships, the determination of proper interval lengths is also critical in fuzzy time-series forecasting. A key point in choosing effective lengths of intervals is that they should not be too large or small[6]. While the 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. Thus, the lengths should reflect at least half the fluctuations in the time series[6].

However, the fluctuations in fuzzy time series can be represented by the absolute value of the first differences of any two consecutive data[6]. Hence, the average-based length is proposed.

3.2 Average-based lengths

In this section, we propose an average-based fuzzy time-series model, which can be used to adjust the lengths of the intervals determined during the early stages of forecasting, when the fuzzy relationships are formulated.

Following the discussion in Section 3.1, the algorithm for average-based lengths is demonstrated as follows[6]:

(1). Calculate all the absolute differences between A_{i+1} and $A_i (i = 1, \ldots, n-1)$, as the first differences and the average of the first differences.

(2). Take one half the average (in step 1) as the length.

(3). According to the length (in step 2), determine the base for the length of intervals by following Tab. 1.

Table 1. Base mapping table

Range	Base
0.1-1.0	0.1
1.1-10	1
11-100	10
101-1000	100

(4). Round the length according to the determined base as the length of intervals.

In order to show how to calculate the average-based length, an example is given. Suppose we have the following time series data: 30,50,80,120,110, and 70. The algorithm for average-based length is implemented step by step below:

(1). The first differences are

20, 30, 40, 10, 40

The average of the first differences is 28.

(2). Take half of the average as the length, which is 14.

(3). According to the length (in step 2), the base for length of intervals is determined as 10 by Tab. 1.

(4). Round the length 14 by the base 10, which is 10. So 10 is chosen as the length of interval.

4 SCSI forecasting

To demonstrate how average-based can assist forecasting in fuzzy time series, Shanghai compound index (SCSI) is used as target. First, we forecast the SCSI using Chen's model with average-based length of interval. Since weighted model gave better results than other model in terms of MSE[15]. Thus, weighted model is used as the target model to compare the average-based model.

4.1 Forecasting with average-based model

The daily stock index is used to demonstrate the forecasting performed by fuzzy time series model. Data covering the period 2004/01/02 ∼ 2004/10/29 are used in the estimation and the period 2004/11/1 ∼ 2004/12/31 is forecast. Closing has been considered in the forecasting of the stock index[6, 14, 15], we also choose closing index as target here.

Step 1. Defining the universe of discourse and intervals.

The universe of discourse for observations, U, is defined as $[D_{\min} - D_1, D_{\max} - D_2]$, where $D_{\min}$ and $D_{\max}$ is the minimum and maximum of known historical data, D_1, D_2 are two proper positive numbers. According to the SCSI for 2004, we can see that $D_{\min} = 1260.316$ and $D_{\max} = 1777.516$. Thus, the universe of discourse is defined as $U = [1260, 1780]$.

Average-based length is calculated:

(1). The first differences are calculated and the average of the first difference is 31.94138.

(2). Take half the average as the length, which is 15.97069.

(3). Given the length chosen in step 2, the base for length of intervals is determined as 10 by following Tab. 1.

(4). Round the length by base 10, which is 20. 20 is chosen as the length.

Now average-based length is applied to Chen's model.

Then U can be partitioned into equal-length intervals $u_1, \ldots, u_{26}$, and the midpoints of these intervals are $m_1, \ldots, m_{26}$, respectively, where $u_1 = [1260, 1280], \ldots, u_2 = [1760, 1780]$.

Step 2. Defining fuzzy sets for observations.

Each linguistic observation, A_i , can be defined by the interval $u_1, \ldots, u_{26}$, as follows:

$$
\begin{aligned}
&A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + \ldots + 0/u_{26};\\
&A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + \ldots + 0/u_{26};\\
&A_2 = 0 + 0.5/u_2 + 1 + 0.5/u_4 + \ldots + 0/u_{26};\\
&\ldots\\
&A_{25} = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + \ldots + 0.5/u_{24} + 1/u_{25} + 0.5/u_{26};\\
&A_{26} = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + \ldots + 0/u_{24} + 0.5/u_{25} + 1/u_{26};
\end{aligned}
$$

Step 3. Fuzzifying observations.

Each SCSI can be fuzzified into a fuzzy set. To facilitate the explanations that follow, some SCSI and their corresponding fuzzy set A_i are listed in Tab. 2.

Step 4. Establishing FLRs and FLRGs.

Table 2. Some fuzzy SCSI

Date	SCSI	Fuzzy Sets
...	...	...
2004/10/08	1422.929	A_9
2004/10/11	1413.154	A_8
2004/10/12	1384.439	A_7
2004/10/13	1386.721	A_7
2004/10/14	1332.938	A_4
2004/10/15	1330.518	A_4
2004/10/18	1335.394	A_4
2004/10/19	1337.618	A_4
2004/10/20	1330.584	A_4
2004/10/21	1310.545	A_3
2004/10/22	1329.355	A_4
2004/10/25	1311.151	A_3
2004/10/26	1324.783	A_4
2004/10/27	1342.799	A_5
2004/10/28	1341.737	A_5
2004/10/29	1320.536	A_4

Based on the fuzzy set in step 3, the FLRs are established, as in Tab. 3. The FLRs can be rearranged to FLRGs, as in Tab. 4.

Step 5. Forecasting.

Table 3. Fuzzy Logic relationships

$\ldots, A_9 \to A_8, A_8 \to A_7, A_7 \to A_7, A_7 \to A_4, A_4 \to A_4, A_4 \to A_3,$
$A_3 \to A_4, A_4 \to A_5, A_5 \to A_5, A_5 \to A_4, \ldots$

Table 4. Fuzzy Logic relationship Group

...
$A_3 \rightarrow A_2, A_4, A_5$
$A_4 \rightarrow A_3, A_4, A_5$
$A_5 \rightarrow A_4, A_5, A_6$
$A_6 \rightarrow A_6, A_8, A_5$
...

Table 5. Forecasts of the SCSI in the Year 2004

Date	Actual	Average-Based	Weighted
2004/11/01	1305.291	1330	1326.3
2004/11/02	1301.528	1310	1330
2004/11/03	1326.749	1310	1330
2004/11/04	1304.776	1330	1326.3
2004/11/05	1305.128	1310	1330
2004/11/08	1304.229	1310	1330
2004/11/09	1307.429	1310	1330
2004/11/10	1354.386	1310	1330
2004/11/11	1347.070	1350	1342.2
2004/11/12	1352.217	1350	1342.2
2004/11/15	1370.046	1350	1342.2
2004/11/16	1370.389	1376.667	1378.3
2004/11/17	1356.138	1376.667	1378.3
2004/11/18	1367.828	1350	1342.2
2004/11/19	1379.960	1376.667	1378.3
2004/11/22	1383.020	1376.667	1378.3
2004/11/23	1371.244	1373.333	1388.3
2004/11/24	1359.125	1376.667	1378.3
2004/11/25	1358.333	1350	1342.2
2004/11/26	1356.727	1350	1342.2
2004/11/29	1337.434	1350	1342.2
2004/11/30	1340.771	1330	1326.3
2004/12/01	1334.944	1350	1342.2
2004/12/02	1333.090	1330	1326.3
2004/12/03	1337.197	1330	1326.3
2004/12/06	1339.644	1330	1326.3
2004/12/07	1323.752	1330	1326.3
2004/12/08	1326.438	1330	1326.3
2004/12/09	1338.810	1330	1326.3
2004/12/10	1317.717	1330	1326.3
2004/12/13	1309.695	1310	1330
2004/12/14	1307.553	1310	1330
2004/12/15	1313.045	1310	1330
2004/12/16	1305.018	1310	1330
2004/12/17	1290.490	1310	1330
2004/12/20	1275.459	1280	1276.6
2004/12/21	1275.170	1310	1310
2004/12/22	1307.570	1310	1310
2004/12/23	1282.719	1310	1330
2004/12/24	1285.036	1280	1276.6
2004/12/27	1280.272	1280	1276.6
2004/12/28	1278.943	1280	1276.6
2004/12/29	1274.313	1310	1310
2004/12/30	1273.709	1310	1310
2004/12/31	1266.496	1310	1310

Forecasting is conducted by the following rules:

Rule 1: If the current fuzzy set is A_i, and the fuzzy logical relationship group of A_i is empty, i.e., $A_i \rightarrow$, the forecast is m_i, the midpoint of u_i.

$$Forecasting = m_i$$

Rule 2: If the current fuzzy set is A_i, and the fuzzy logical relationship group of A_i is one-to-one, i.e., $A_i \rightarrow A_j$, the forecast is m_j, the midpoint of u_j.

$$Forecasting = m_j$$

Rule 3: If the current fuzzy set is A_i, and the fuzzy logical relationship group of A_i is one-to-many, i.e., $A_i \rightarrow A_{j1}, A_{j2}, \ldots, A_{jn}$, the forecast is equal to the average of $m_{j1}, m_{j2}, \ldots, m_{jn}$, the midpoints of $u_{j1}, u_{j2}, \ldots, u_{jn}$, respectively.

$$Forecasting = \frac{\sum_{i=1}^{n} m_{ji}}{n}$$

In this paper, we use MSE to compare the forecasting performance.

$$MSE = \frac{\sum_{i=1}^{n}(Forecasting\_Value_i - Actual\_Value_i)^2}{n}$$

where there are altogether n data.

4.2 Defuzzifying the forecast

In the following, we only illustrate the process of some SCSIs.The similar procedure can be applied to other SCSIs.

[2004/11/01] Because the fuzzified SCSI of 2004/10/29 shown in Tab. 2 is A_4, and from Tab. 3, we can see that the corresponding FLRGs to A_4 is as follows:

$$A_4 \rightarrow A_3, A_4, A_5$$

Following the forecasting rules 3, the forecasted SCSI of 2004/11/01 is equal to

$$\frac{1}{3}(m_3 + m_4 + m_5) = \frac{1}{3}(1310 + 1330 + 1350) = 1330$$

[2004/11/11] Because the fuzzified SCSI of 2004/11/10 shown in Tab. 2 is A_5, and from Tab. 3, we can see that the corresponding FLRGs to A_5 is as follows:

$$A_5 \rightarrow A_4, A_5, A_6$$

Following the forecasting rules 3, the forecasted SCSI of 2004/11/11 is equal to

$$\frac{1}{3}(m_6 + m_4 + m_5) = \frac{1}{3}(1370 + 1330 + 1350) = 1350$$

[2004/12/24] Because the fuzzified SCSI of 2004/12/23 shown in Tab. 2 is A_2, and from Tab. 3, we can see that the corresponding FLRGs to A_2 is as follows:

$$A_2 \rightarrow A_1, A_2$$

Following the forecasting rules 3, the forecasted SCSI of 2004/11/01 is equal to

$$\frac{1}{2}(m_1 + m_2) = \frac{1}{2}(1270 + 1290) = 1280$$

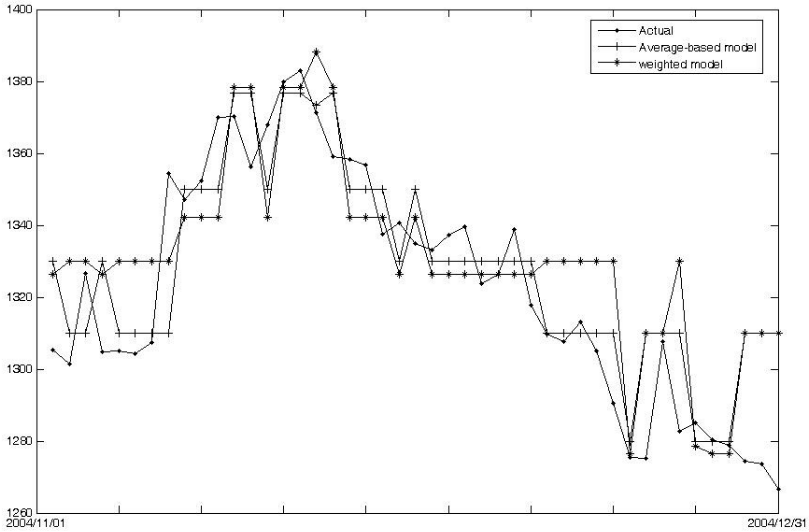

Fig. 1. Performance comparison

Table 6. MSE comparison

	Average-Based Model	Weighted model
MSE	292.3224	436.227

5 Empirical analyses

Following section 4, the out-of -sample defuzzified forecasts for average-based model and weighted model[15] in the 2004 (from 11/01 to 12/31) are listed in Tab. 5 as an illustration. These defuzzified forecasts are also depicted in Fig. 1.

For comparison purpose, we list the MSE of 2004 in Tab. 6.

6 Conclusion

In this paper, we introduce average-based fuzzy time series model, where the data of historical SCSI are adopted to illustrate the forecasting process. The forecasting results show that the proposed model outperforms weighted model. The average-based models can thus be applied to improve fuzzy time series forecasting. Furthermore, the average-based lengths are simple to calculate. In summary, the proposed model is as simple as Chen's method, but more accurate in forecasting.

References

[1] G. Box, G. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-day Inc., San Francisco, CA, 1976.

[2] C. Brooks. *Introductory Econometrics for Finance*. UK, Cambridge Univ. Press, Cambridge, 2002.

[3] S. Chen. Forecasting enrollments based on fuzzy time series. *Fuzzy Sets and Systems*, 1996, **81**: 311–319.

[4] S. Chen, J. Hwang. Temperature prediction using fuzzy time series. *IEEE Trans Syst Man Cybern*, 2000, **30**: 263–275.

[5] Y. Hsu, S. Tse, B. Wu. A new approach of bivariate fuzzy time series analysis to the forecasting of a stock index. *Int. J. Uncertainty,Fuzziness. Knowl.-Based Systems*, 2003, **11**: 671–690.

[6] K. Huarng. Effective lengths of intervals to improve forecasting in fuzzy time series. *Fuzzy Sets and Systems*, 2001, **123**: 387–394.

[7] Q. Song, B. Chissom. Forecasting enrollments with fuzzy time series-part 1. *Fuzzy sets and Systems*, 1993, **54**: 1–9.

[8] Q. Song, B. Chissom. Fuzzy time series and its models. *Fuzzy sets and Systems*, 1993, **54**: 269–277.
[9] Q. Song, B. Chissom. Forecasting enrollments with fuzzy time series-part 2. *Fuzzy sets and Systems*, 1994, **62**: 1–8.
[10] J. Sullivan, W. Woodall. A comparison of fuzzy forecasting and markov modeling. *Fuzzy Sets Systems*, 1994, **64**: 279–293.
[11] F. Tseng, G. Tzeng, etc. Fuzzy arima model for forecasting the foreign exchange market. *Fuzzy Sets and System*, 2001, **118**: 9–19.
[12] M. Versaci, F. Morabito. Fuzzy time series approach for disruption prediction in tokamak reactors. *IEEE Trans. Magnetics*, 2003, **29**(1503-1506).
[13] C. Wang. Predicting tourism demand using fuzzy time series and hybrid grey theory. *Tourism Management*, 2004, **25**(367-374).
[14] H. Yu. A refined fuzzy time-series model for forecasting. *Physica A*, 2005, **346**: 657–681.
[15] H. Yu. Weighted fuzzy time series models for taiex forecasting. *Physica A*, 2005, **349**: 609–624.
[16] L. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences*, 1975, **8**: 199–249.