International Journal of Physics and Mathematical Sciences ISSN: 2277-2111 (Online)
An Open Access, Online International Journal Available at http://www.cibtech.org/jpms.htm
2014 Vol. 4 (1) January-March, pp.20-25/Sivasamy and Ola Ama
Research Article

MIXED AVERAGE-BASED FUZZY TIME SERIES MODELS FOR FORECASTING FUTURE CIVILIAN FATALITIES BY TERRORIST ATTACKS IN SOUTH ASIA

***R. Sivasamy and Njoku Ola Ama**
Department of Statistics, University of Botswana, Gaborone
**Author for Correspondence*

ABSTRACT

Numerous fuzzy time series models have been widely used for forecasting the future outcomes when only small amounts of historical data are available over a short span of a time period. A primary assumption is that the given time series is a first-order fuzzy time series which is also time-invariant. This paper presents a new forecasting method by mixing both simple and weighted average values based on the frequency distribution within each 'Fuzzy Logical Relationship Group (FLRG)' over the historical data. The empirical results of the proposed method and that of other average based methods obtained from a hypothetical time series data set show that the Mixed Average-Based Model (MABM) can improve the fuzzy time series forecasting outputs than the existing simple average- or weighted average-based methods. Lastly MABM is fitted to the observed data on civilian fatalities caused by terrorist attacks in South Asia and is used to forecast the civilian fatalities of 2014.

Keywords: *Linguistic Variable, Fuzzy Set, Fuzzy Time Series, Fuzzy Logical Group (Flr), Forecasting*

INTRODUCTION

One particularly important group of models in the conventional time series models has been the family of Auto Regressive Integrated Moving Average (ARIMA) models of Box and Jenkins (1976). However, these ARIMA models have been applied to series with precise number of observations of at least 50 or preferably more than 100 observations (Tseng and Tzeng, 2001). The forecasts obtained by these models may easily encounter the problem of over-fitting. Further, because the existing ARIMA methods could not effectively analyze time series with small amounts of data and in particular historical data, fuzzy time series methods were developed by Zadeh (1975) to handle problems involving human linguistic terms.

Song and Chissom (1993, 1994) introduced a first order time-invariant model and a time-variant model of fuzzy time series in 1993 and thus fuzzified forecasting using logic max-min composition method applied to the observed enrolments at the University of Alabama. Chen (1996) simplified the Song and Chissom's method through arithmetic operations in place of logic operations. Subsequently and following these definitions, various fuzzy time series models have been proposed for several applications by several authors to forecast enrolments, stock indices, reactors, and temperature (Chen and Hwang, 2000; Xihao and Yimin, 2008). The first investigation of Huarng (2001) was on the effective lengths of intervals which showed that different lengths of intervals may result in different fuzzy relationships, and in turn different forecasting results.

In this paper a Mixed Average-Based Model (MABM) is proposed for forecasting of fuzzy time-series data of fatalities of South Asia from 2005 to 2014. One advantage of this methodology is that it adjusts the lengths of the intervals determined during the early stages of forecasting when the fuzzy relationships are formulated as discussed in Xihao and Yimin (2007) and Huarng (2001). The goal of this study is to propose a mixed average-based fuzzy time series model to improve forecasting accuracy of fuzzy time series. The annual fatality time series data of South Asia is downloaded and the same is 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. Section 3 explains the relevant definitions of lengths of intervals and proposes the mixed-average average-based fuzzy time series model. Section 4 elaborates on the use of model in forecasting fatality time series in South Asia and evaluates the models' performance while section 5 concludes the paper.

Fuzzy Time Series Revisited

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

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

$$A_i = \frac{fA_i(u_1)}{u_1} + \frac{fA_i(u_2)}{u_2} + \ldots\ldots\ldots + \frac{fA_i(u_n)}{u_n},$$

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

Definition 1: Y (t)(t = . . . , 0, 1, 2, . . .), is a subset of R. Let Y (t) be the universe of discourse defined by fuzzy set $f_1(t)$. If F(y) consisted of $f_1(t)$, $f_2(t)$, . . ., then F(t) is defined as a fuzzy time series on Y (t)(t =. . . , 0, 1, 2, . . .).

Definition 2: If there exists a fuzzy relationship R (t − 1, t), such that F (t) = F(t − 1) × R(t − 1, t), where × 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 LHS $A_i \rightarrow A_{j1}$, $A_i \rightarrow A_{j2}$, both FLRs can be grouped together into fuzzy logical relationship groups (FLRG) $A_i \rightarrow A_{j1}, A_{j2} \ldots$

Different fuzzy time-series models have been proposed by following Song and Chissom's definition of fuzzy time-series. 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.

Effective Intervals for Fuzzy Time Series Models

Assume that the given time series is a first-order fuzzy time series which is also time-invariant. Then a key point in choosing effective lengths of intervals is that they should not be too large or small. The fluctuations in fuzzy time series can be represented by the absolute value of the first differences of any two consecutive data as in Huarng (2001) and Xihao and Yimin (2007) for which the mixed average-based length is proposed as follows:

Method-1 for the Selection of Intervals:

1. Calculate all the absolute differences between A_{i+1} and A_i (i = 1, . . . , n−1), as the first differences and the average of the first differences.
2. Take one half of the averages (in step 1) as the length.
3. According to the length (in step 2), determine the base for the length of intervals as in Table-1

Table 1: Base mapping table

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

Length is to be rounded according to the determined base from Table-1 as the effective length of intervals.

Method-2
Calculate the range of the given time series and divide it by the number of linguistic values of the linguistic variable under study to know the approximate length and then decide the appropriate length while rounding off the approximate length.
Forecasting of a Hypothetical Time Series with Seven Linguistic Values
Suppose that the time series data presented in column two of Table-2 represent the production level (called linguistic variable) of a business to meet the weekly demand over 21 weeks (column 1); where the linguistic values are A_1= very very low, A_2=very low, A_3= low, A_4 = neither low nor high, A_5=high, A_6= very high, A_7= very very high and these words are represented by fuzzy sets.

Table 2: Data on Production Level for 21 weeks, Fuzzy Series and the Forecasted Values

Week, t	Production Level: Y(t)	Fuzzy series	Forecasts Based on mean	Forecasts Based on weighted mean	Forecasts Based on mixed means
1	10.2	A_1			
2	12.7	A2	12.55	12.55	12.55
3	13.3	A2	13.40	13.12	13.12
4	12.8	A2	13.40	13.12	13.12
5	13.8	A3	13.40	13.12	13.40
6	14.2	A3	15.10	14.93	14.93
7	13.9	A3	15.10	14.93	14.93
8	14.6	A3	15.10	14.93	14.93
9	15.6	A4	15.10	14.93	15.10
10	14.9	A3	15.95	15.95	15.95
11	15.8	A4	15.10	14.93	15.10
12	17.5	A5	15.95	15.95	15.95
13	16.8	A5	18.50	18.50	18.50
14	17.9	A5	18.50	18.50	18.50
15	18.8	A6	18.50	18.50	18.50
16	18.2	A5	19.35	19.92	19.35
17	19.2	A6	18.50	18.50	18.50
18	20.9	A7	19.35	19.92	19.92
19	20.1	A6	20.20	20.20	20.20
20	21.4	A7	19.35	19.92	19.92
21	21.8	A7			
Error=sum of ABS{(Y-F)/Y}			0.939789	0.907804	0.834481

Step 1: *Defining the universe of discourse and intervals for fuzzy sets*
Here Y_{min}=10.2 and Y_{max}=21.8, and thus Range = Y_{max} - Y_{min} =21.8-10.2=11.6. Dividing 11.6 by 7 gives 1.657 (approximately 1.7). The effective length of each interval is 1.7 units. Now let the *universe of discourse be,* U= [10-21.9], then U can be partitioned into equal-length intervals u_1, u_2,. . . , u_7 as follows in Table-3:

Table 3: Illustration for forming fuzzy sets

Universe U	Elements of U :Class Interval	Fuzzy set	Mid- point
u_1	10.0 -- 11.7	A1	10.9
u_2	11.7 -- 13.4	A2	12.6
u_3	13.4 -- 15.1	A3	14.3
u_4	15.1 -- 16.8	A4	16.0
u_5	16.8 -- 18.5	A5	17.7
u_6	18.5 -- 20.2	A6	19.4
u_7	20.2 -- 21.9	A7	21.1

Step 2: *Defining fuzzy sets for observations*

Each linguistic observation, A_i, can be defined by the intervals $u_1, \ldots, u_7$, of the Universe of discourse U as follows:

$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7$;
$A_2 = 0.5/u1 + 1/u2 + 0.5/u3 + 0/u4 + 0/u5 + 0/u6 + 0/u7$;
$A_3 = 0/u1 + 0.5/u2 + 1/u3 + 0.5/u4 + 0/u5 + 0/u6 + 0/u7$
$A_4 = 0/u1 + 0/u2 + 0.5/u3 + 1/u4 + 0.5/u5 + 0/u6 + 0/u7$
$A_5 = 0/u1 + 0/u2 + 0/u3 + 0.5/u4 + 1/u5 + 0.5/u6 + 0/u7$
$A_6 = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0.5/u_5 + 1/u_6 + 0.5/u_7$
$A_7 = 0/u1 + 0/u2 + 0/u3 + 0/u4 + 0/u5 + 0.5/u6 + 1/u7$

Notice that there is no restriction on the number of the fuzzy sets defined. The elements of each fuzzy set A_j are $u_l, u_2, \ldots, u_7$. The memberships for each element $u_1, \ldots, u_7$, is in the respective fuzzy set specified above each u as $0/u_1$ or $0.5/u_4$ or as $1/u_7$. For practical purposes, one can also use $A_1, A_2, \ldots$ and A_7 as vectors whose elements are the corresponding memberships. By definition, we know that we have defined a time-invariant fuzzy time series.

Since fuzziness is contained in human language, it can be quantified in terms of a membership value in the interval [0, 1] differently by human judgement, evaluation and decisions.

Step 3: *Fuzzification of Observations on the Production Level*

Each production level can be fuzzified into a fuzzy set. Since Y(1)=10.2 falls in u_1 interval and u_1 has the highest membership of 1 in the fuzzy set A_1, the corresponding member of the fuzzy time series is A_1. As Y(2)=12.7 falls in u_2 interval and u_1 has the highest membership of 1 in the fuzzy set A_2, the corresponding member of the fuzzy time series is A_2 and son. The fuzzified historical data of the production level are shown in column 3 of Table 2. Here the FLR $A_i \sim A_k$ means that "if the production level of year i is A_j, then that of year i + 1 is A_k", where A_j is called the current state of the production level and A_k is called the next state of the production level. The frequency distribution of such FLRGs associated with each A_j for j=1,2,…,7 over the fuzzy time series can be obtained from Table-2 and is shown in Table-4 together with the simple mean and the weighted mean values of the production level for each FLG.

Table-4: Frequency Distribution of FLGs

SET	FLRG	Mid-point	Frequency	MEAN	W.MEAN
A_1	A_1~A_2	12.55	1	12.55	
A_2	A_2~A_2	12.55	2		
	A_2~A_3	14.25	1	13.4	13.12
A_3	A_3~A3	14.25	3		
	A_3~A4	15.95	2	15.1	14.93
A_4	A_4~A3	14.25	1		
	A_4~A5	17.65	1	15.95	
A5	A_5~A5	17.65	2		
	A_5~A6	19.35	2	18.5	
A6	A_6~A5	17.65	1		
	A_6~A7	21.05	2	19.35	19.92
A7	A_7~A6	19.35	1		
	A_7~A7	21.05	1	20.2	

Among the given 21 items , total transitions =20

The algorithm for MABM is outlined step by step below: Consider the FLRG of A_j from Table-4.

Step1: Find the number of transitions from table-2 and thus now A_1~A_2 is 1; $A_2 \sim A_2$ is 2 and that of $A_2 \sim A_3$ is 1; A_3~A_3 is 3 and that of A_3~A_4 is 2; A_4~A_3 is 1 and that of A_4~A_5 is 1; A_5~A_5 is 2 and that of A_5~A_6 is 2; A_6~A_5 is 1 and that of A_3~A_4 is 2; and A_7~A_6 is 1 and that of A_7~A_7 is 1.

Step-2: Find the simple mean which is 13.4 and the weighted mean (w.mean) which is 13.12 for the case of A2. Similarly, we find that the means for A_3, A_6 are 15.1 and 19.35 respectively while 'w.mean' values are 14.93 and 19.92. All other fuzzy sets of Table 4 have simple mean only.
Step3: As the given time series steadily increase with the passage time from the week-1 to week-21 given in Table-2, assign the min (13.12, 13.4)=13.2 for the first two destinations reached from A_2 as the forecasted values and assign 'w.mean' to the rest *i.e.* 13.4 is assigned to the third destination travelled from A_2 to A3 as shown below in Table 5:

Table 5: Illustration of the generation of forecasted values

t	Y(t) given	Fuzzy Series	Forecasts Based on mean	Forecasts Based on w.mean	Forecasts Based on mixed
2	12.7	A2			
3	13.3	A2	13.40	13.12	13.12
4	12.8	A2	13.40	13.12	13.12
5	13.8			13.12	13.40

Table 5 shows a simple improvement in the observed series and the forecast series calculated under the mixed based method as well as mean based and weighted mean based forecasts.
Step4: Repeat the step-2 with other FLRGs wherever possible. From Table-2, it is noted that FLRGs of A_2, A_3 and A_6 have both simple mean and weighted mean values.
Using the illustration of Table-5, it is a simple job to forecast the future events. Using this MABM algorithm **step-1** to **step-4**, those forecast values obtained by implementing the mixed and non-mixed algorithms have been reported in Table-2. Further, it is seen that forecasts of the mixed method produces smaller error than the forecasts produced by non-mixed averages methods.

Forecasting of Fatalities of South Asia: 2005-2014

The above mixed type methodology has been applied to the data downloaded from the web site on the number of civilian fatalities by terrorist attacks in South Asia that took place from 2005 to 2013 . The original data Y (t) represent the number of civilian fatalities by terrorist attacks of South Asia. The outcomes of this analysis in terms of forecast number of civilian fatalities by terrorist attacks of South Asia have been calculated and reported in Table-7 as F(t) for t =2006 to 2014. The value for 2009 has been omitted in the calculations because it is an outlier and is capable of marring the results.

Table 7: Number of civilian fatalities Y (t) by terrorist attacks of South Asia and forecast values F(t) by fuzzy theory

t	Y(t)		Fuzzy sets	Forecast, F(t)	Error
2005	2063		A_1		
2006	2803		A_3	2860	0.020335
2007	3128		A_4	3180	0.016624
2008	3653		A5	3393	0.071174
2009	14197	OMIT			
2010	2571		A_2	2220	0.136523
2011	3173		A_4	3180	0.002206
2012	3270		A_4	3393	0.037615
2013	3533		A_5	3340	0.054628
2014				2540	0.339105

Average forecasting errors of time-invariant series is calculated using

$$\text{Error}=\frac{|\text{Actual Y(t)-Forecasted F(t)}|}{\text{Actual Y(t)}}$$

The necessary details on the partition of the universe of discourse U =[2060,3660], its effective and equal length intervals u_1, u_2, ..., u_5 , and the fuzzy sets A_1,A_2, ... ,A_5 which are linguistic values of the linguistic variable " civilian fatalities " are provided in Table-8.

Table 8: Universe of discourse, U and fuzzy sets A_j

U	LL	UL	A_j	Midpoint
u_1	2060	2380	A_1	2220
u_2	2380	2700	A_2	2540
u_3	2700	3020	A_3	2860
u_4	3020	3340	A_4	3180
u_5	3340	3660	A_5	3500

For this case, the fuzzy logical relationships obtained from Table-8 are shown in Table-9

Table 9: Fuzzy logical relationships derived from theTable-8

Fuzzy set	FLR	frequency		
A_1	A_3	1		
A_2	A_4	1		
A_3	A_4	1		
A_4	A_4	1	mean	w.mean
	A_5	2	3340	3393.333
A_5	A_2	1		

Applying the proposed mixed method to the data in Table -7 the forecast number of civilian fatalities by terrorist attacks of South Asia would be around 2540 with an error co-efficient of $\pm$ 34%.

Conclusion

The paper has demonstrated that by connecting approaches from two different first-order and time-invariant models—the simple average and the fuzzy weighted average, a generalization that combines the good properties of the two models, a new best fitting of Fuzzy time series through MABM and one that produces smaller estimated error has been obtained. The proposed MABM is as good of those of Xihao and Yimin (2007, 2008) and can be better, producing smaller error. The results of the analysis based on the data in Table 2 and Table7 are clear indication of this assertion and the utility of the MABM. The application is simple and does not create any methodological problems and is therefore recommended.

REFERENCES

Tseng F and Tzeng G (2001). Fuzzy ARIMA model for forecasting the foreign exchange market. *Fuzzy Sets and System* **118** 9–19.

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

Huarng K (2001). Effective lengths of intervals to improve forecasting in fuzzy time series. *Fuzzy Sets and Systems* **123** 387–394.

Zadeh L (1975). The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences* **8** 199–249.

Song Q and Chissom B (1993). Fuzzy time series and its models. *Fuzzy Sets and Systems* **54** 269–277.

Song Q and Chissom B (1994). Forecasting enrollments with fuzzy time series-part 2. *Fuzzy sets and Systems* **62** 1–8.

Chen S (1996). Forecasting enrolments based on fuzzy time series. *Fuzzy Sets and Systems* **81** 311–319.

Chen S and Hwang J (2000). Temperature prediction using fuzzy time series. *IEEE Transactions on Systems, Man, and Cybernetics* **30** 263–275.

Sun Xihao and Yimin Li (2008). Average-based fuzzy time series models for forecasting Shanghai compound index. *World Journal of Modelling and Simulation* **4** (2008) No. 2, pp. 104-111, ISSN 1 746-7233, England, UK

Hsu Y, Tse S and Wu B (2003). A new approach of bivariate fuzzy time series analysis to the forecasting of stock index. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **11** 671-690.