

ESTIMATION OF ERROR USING VARIOUS MEASURES OF AVERAGES IN FORECASTING PROBLEMS

U. ABUTHAHIR

ABSTRACT. In general, there are many number of types to forecast items like weather, temperature, fuel, gold rate, enrollments and the economy. Depending on the forecasting results, people can very easily get benefits out of the forecasting activity. This paper presents a modified higher-order fuzzy logical relationship as per Root Mean Square Error using different averages like Arithmetic, Geometric, Harmonic, Heronian and Root Mean Square. This proposed method has maximum predicting accuracy rate when compared to the already present methods [5].

1. INTRODUCTION

In fact forecasting plays a key role in our day to day activity. In the last two decades, many number of approaches are developed for time series forecasting. In reality, forecasting is the major process by which we are able to predict the future in a fruitful way. Through this the decision makers find it easy to analyze the data and take the suitable decisions for the future. People can get to know many matters through forecasting. It helps as protect from various disasters. It also gives as various avenues for the betterment of our lives. It also helps know how problems arise and to stay updated. For all these, predicting really helps as to a great extent by helping as in various modes. Prediction helps as solve many problems and lead a fearless life and to learn many new things. However, the

2010 *Mathematics Subject Classification.* 94D05, 91B84.

Key words and phrases. Fuzzy time series, Fuzzy set, Forecasting and Fuzzy Relationship.

good old time series methods cannot focus on the predicting problems wherein the values of time series become linguistic terms shown by fuzzy sets [4]. To remove this drawback, Song and Chissom presented the famous theory of the fuzzy time series in [1] and [2]. Fuzzy time series models relate to both numerical and linguistic values.

Using the above said theory of fuzzy time series, [1] reported the predicting methods to forecast the enrollments of the University of Alabama. In lieu of complicated operations, S.M. Chen [4], used the simple arithmetic operation for time series forecasting. It contains the advantages to reduce the time and to simplify the calculating process. After which many such related research works were reported following their frame work and target to standardise the predicting accuracy and minimise the computational overhead. These works consist of enrolment forecasting [1], [2], [3] and [8,9], TAIEX prediction [5], and stock price forecasting [6] and [7] respectively.

In this paper, many number of mean techniques are used in the predicting enrolments of University of Alabama [5]. The remaining paper is arranged as follows. In section 2, brief review of some basic concepts of fuzzy time series are given. In section 3, the modified method for forecasting problems based on higher-order fuzzy logical relationship [8] is given. In section 4, numerical illustration by using different mean techniques such as Arithmetic, Geometric, Harmonic, Heronian and Root Mean Square are given. Finally the conclusions are presented in section 5.

2. PRELIMINARIES OF FUZZY TIME SERIES

The preliminaries are have been taken from the article [5] kindly refer this article in section 2.1 for more informations.

PREDICTION ERROR: (PE)

Prediction error is defined as

$$(2.1) \quad PE = \sqrt{\frac{\sum_{j=1}^n (FV_j - OV_j)^2}{n}}$$

where FV = Forecast value, OV = Observed Value and n = number of terms.

Using formula to forecast the enrolments for various averages:

Name of the Mean	Formula
Arithmetic Mean	$(a_1 + a_2 + \dots + a_m)/m$
Geometric Mean	$(a_1 \times a_2 \times \dots \times a_m)^{1/m}$
Harmonic Mean	$m/(1/a_1 + 1/a_2 + \dots + 1/a_m)$
Heronian Mean	$(a_1 + a_2 + \dots + a_m + (a_1 \times a_2 \times \dots \times a_m))/m + 1$
Root Mean Square	$((a_1^2 + a_2^2 + \dots + a_m^2)/m)^{1/2}$

3. MODIFIED FORECASTING METHOD AS PER HIGHER ORDER FUZZY LOGICAL RELATIONSHIPS

The Following is the modern forecasting method as per higher-order fuzzy logical relationship [8]. In major existing algorithms, the universe interval is considered as $[D_{min} - D_1, D_{max} + D_1]$. Here the universe interval that uses standard normal distribution range based definition is considered ie, $U = [\mu - 3\sigma, \mu + 3\sigma]$ where μ and σ are the Arithmetic Mean(AM) and standard deviation (SD) values of the data. In the existing type [5] the forecast variable is considered by taking into consideration all the values that include the values repeated, whereas in this method, repeated values are considered as a single value. We call the forecast variable as Modified forecast variable, as these modification the Prediction error of the method is minimum when compared with the already present methods [5].

Step 1: Define the Big interval U as $U = [\mu - 3\sigma, \mu + 3\sigma]$ where μ and σ are the AM and SD deviation values of the data, and divide the Big interval U into n sub-intervals u_1, u_2, \dots, u_n of same length. The length of the interval is arbitrarily fixed in accordance with the required number of sub-intervals.

Step 2: Define linguistic terms, A_i , $i = 1$ to n , represented by fuzzy set $A_i = \sum_{j=1}^n \frac{\mu_{ij}}{u_j}$ where μ_{ij} is the grade value of u_j belonging to A_i and is defined by

$$\mu_{ij} = \begin{cases} 1 & \text{if } j = i \\ 0.5 & \text{if } j = i - 1 \text{ or } i + 1 \\ 0 & \text{otherwise} \end{cases}$$

Step 3: Fuzzify each datum into a fuzzy set derived in step 2. If each datum belongs to u_i and the maximum grade value of A_i occurs at u_i then the datum is fuzzified into A_i where $1 \leq i \leq n$.

Step 4: Establish the higher-order FLR from the fuzzified data of the practising data set.

Step 5: Transform every higher-order FLR $A_{x1}, A_{x2}, \dots, A_{kj}, \dots, A_{xn}, \rightarrow A_{x(n+1)}$ into the following form, $A_{x1+v(x1)+v(x2)}, \dots, A_{x1+v(x1)+\dots+v(x(n-1))} \rightarrow A_{x1+v(x1)+\dots+v(x(n-1))+v(xn)}$ where $v(x_1), v(x_2), \dots, v(x_n)$ are integers.

Step 6: Let the transformed higher-order FLRs from in step 5, having the same LHS form an higher-order FLRGs. For example, let us assume the following transformed III-order fuzzy logical relationships,

$$A_{a1}, A_{a1+v(a1)}, A_{a1+v(a1)+v(a2)} \rightarrow A_{a1+v(a1)+v(a2)+v(a3)} \\ A_{b1}, A_{b1+v(b1)}, A_{b1+v(b1)+v(b2)} \rightarrow A_{b1+v(b1)+v(b2)+v(b3)}, \dots, A_{k1}, A_{k1+v(k1)}, A_{k1+v(k1)+v(k2)} \\ \rightarrow A_{k1+v(k1)+v(k2)+v(k3)}$$

If $v(a_1) = v(b_1) = \dots = v(k_1)$ and $v(a_2) = v(b_2) = \dots = v(k_2)$ then these III-order fuzzy logical relationships could be grouped in to a transformed III-order FLRG , as given below:

$$A_x, A_{x+v(y1)}, A_{x+v(y1)+v(y2)} \rightarrow A_{x+v(y1)+v(y2)+v(a3)}, A_{x+v(y1)+v(y2)+v(b3)} \\ \dots A_{x+v(y1)+v(y2)+v(k3)} \text{ where } x = a_1, b_1, \dots k_1, \\ v(y_1) = v(a_1) = v(b_1) = \dots = v(k_1) \text{ and } v(y_2) = v(a_2) = v(b_2) = \dots = v(k_2)$$

Step 7: To predict $F(t)$, select a transformed higher-order FLRG for prediction. We know that $F(t-n) = A_{in}, F(t-(n-1)) = A_{i(n-1)}, \dots, F(t-2) = A_{i2}$ and $F(t-1) = A_{i1}$ where $A_{i1}, A_{i2}, \dots, A_{in}$ are fuzzy sets. Due to the transformed higher order FLRG obtained from the above step, select the relating transformed higher-order FLRG for forecast. Then the transformed higher-order FLRG is of the form, where $A_x = A_{in}, A_{x+v(y1)} = A_{i(n-1)}, \dots, A_{x+v(y1)+v(y2)+\dots+v(y_{n-1})} = A_{i1}$, then instead of x by the subscript in of the fuzzy set A_{in} to get the derived fuzzy sets

$$A_{in+v(y1)+v(y2)+\dots+v(y_{n-1})+v(an)}, A_{in+v(y1)+v(y2)+\dots+v(y_{n-1})+v(bn)}, \dots, \\ A_{in+v(y1)+v(y2)+\dots+v(y_{n-1})+v(kn)} \text{ for prediction.}$$

$$\text{Let } A_{j1} = A_{in+v(y1)+v(y2)+\dots+v(y_{n-1})+v(an)}; A_{j2} = A_{in+v(y1)+v(y2)+\dots+v(y_{n-1})+v(bn)} \dots, \\ A_{jk} = A_{in+v(y1)+v(y2)+\dots+v(y_{n-1})+v(kn)}.$$

The maximum membership values of $A_{i1}, A_{j1}, A_{j2} \dots$ and A_{jk} occur at the intervals u_{i1}, u_{j1}, u_{j2} and u_{jk} respectively and $m_{i1}, m_{j1}, m_{j2} \dots$ and m_{jk} are the midpoints of the intervals u_{i1}, u_{j1}, u_{j2} and u_{jk} respectively. Here the midpoints $m_{i1}, m_{j1}, m_{j2} \dots$ and m_{jk} of the intervals u_{i1}, u_{j1}, u_{j2} and u_{jk} and are calculated using different Mean like i)Arithmetic, ii)Geometric, iii)Harmonic, iv) Heronian and v) Root mean square.

Then the alternate forecast variable [MFVar(t)] of the year t is determined is given below:

$$(3.1) \quad \mathbf{MFVar(t)} = \text{Average of distinct values of } m_{i1}, m_{j1}, m_{j2} \dots m_{jk} - m_{i1}$$

Here also five different corresponding averages are used to calculate average of distinct values.

The Forecast or predict value FV(t) determined is given below

$$(3.2) \quad \mathbf{FV(t)} = RV(t - 1) + MFVar(t)$$

where $RV(t - 1)$ is actual value of the previous year $(t - 1)$.

4. NUMERICAL ILLUSTRATION

The student enrolment of the Alabama University from the year 1971 to 1992 is considered as historical data by several authors [1], [2] and [3] for forecasting. The same historical enrolments of the Alabama University is considered and given in Table 1.

Year	Data	Year	Data
1971	13055	1982	15433
1972	13563	1983	15497
1973	13867	1984	15145
1974	14696	1985	15163
1975	15460	1986	15984
1976	15311	1987	16859
1977	15603	1988	18150
1978	15861	1989	18970
1979	16807	1990	19328
1980	16919	1991	19337
1981	16388	1992	18876

TABLE 1

The present method of forecasting due to the higher-order fuzzy logical relationship is applied to the data in Table-1. The historical enrolments data from Table-1 is arbitrarily divided into two parts in which among the two parts, one of the part is considered as the practising data set and the remaining part is the testing data set. Here the enrolments of the year 1972 to 1988 are considered as the practising data set and 1989 to 1992 as the testing data set, as in [5].

Step 1: μ and σ are calculated for the data in Table-1. Here $\mu = 16194.18$ and $\sigma = 1774.725$ the Big interval $U = [\mu - 3\sigma, \mu + 3\sigma] = [10869.8, 21518.356] \approx [10870, 21519]$. Let us assume that the length of each sub-interval from the universe interval U be 1100 as in reference article [5]. Number of interval is calculated as $n = (\text{Upper limit of } U - \text{Lower limit of } U) / \text{length of interval}$.

i.e, $u_i = [10870 + (i-1) 1100, 10870 + i]$, for $i = 1, 2, \dots, 10$ (4)

Step 2: Let us define the fuzzy set relating or concerning to each interval u as $A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + \dots + 0/u_{10}$, $A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + \dots + 0/u_{10}$, $\dots A_{10} = 0/u_1 + 0/u_2 + 0/u_3 + \dots + 0.5/u_9 + 1/u_{10}$

Step 3: Determination of fuzzified data:

In Table 1, the enrolment of the year 1971 is 13055, which is present in the interval u_2 and hence is fuzzified into A_2 . Similarly the enrolment of the year 1972 is 13563 which is belonging to the interval u_3 and hence is fuzzified to in A_3 . In the same way the fuzzified practising data and fuzzified testing data are estimated and given in Table 2 and Table 3.

Year	Data	Fuzzy set	Year	Data	Fuzzy set
1971	13055	A_2	1980	16919	A_6
1972	13563	A_3	1981	16388	A_6
1973	13867	A_3	1982	15433	A_5
1974	14696	A_4	1983	15497	A_5
1975	15460	A_5	1984	15145	A_4
1976	15311	A_5	1985	15163	A_4
1977	15603	A_5	1986	15984	A_5
1978	15861	A_5	1987	16859	A_6
1979	16807	A_6	1988	18150	A_7

TABLE 2. Fuzzified Training Data of the Enrolments

Year	Data	Fuzzy set
1989	18970	A_8
1990	19328	A_8
1991	19337	A_8
1992	18876	A_8

TABLE 3. Fuzzified testing data of the enrolments

Step 4: Estimation of fuzzy logical relationship:

The fuzzified practising data of the historical enrolment 1972, 1973 and 1974 are A_3, A_3 and A_4 respectively and hence the concerning second order FLR $A_3, A_3 \rightarrow A_4$. In Proceeding in this way the FLR is established for the fuzzified training data and are given in Table-4.

Year	Fuzzy logical relationships	Fuzzy variable logical relationships
1971, 1972 \rightarrow 1973	$f'_2, f'_3 \rightarrow f'_3$	$f'_2, f'_{2+1} \rightarrow f'_{2+1+0}$
1972, 1973 \rightarrow 1974	$f'_3, f'_3 \rightarrow f'_4$	$f'_3, f'_{3+0} \rightarrow f'_{3+0+1}$
1973, 1974 \rightarrow 1975	$f'_3, f'_4 \rightarrow f'_5$	$f'_3, f'_{3+1} \rightarrow f'_{3++1}$
1974, 1975 \rightarrow 1976	$f'_4, f'_5 \rightarrow f'_5$	$f'_4, f'_{4+1} \rightarrow f'_{4+1+0}$
1975, 1976 \rightarrow 1977	$f'_5, f'_5 \rightarrow f'_5$	$f'_5, f'_{5+0} \rightarrow f'_{5+0+0}$
1976, 1977 \rightarrow 1978	$f'_5, f'_5 \rightarrow f'_5$	$f'_5, f'_{5+0} \rightarrow f'_{5+0+0}$
1977, 1978 \rightarrow 1979	$f'_5, f'_5 \rightarrow f'_6$	$f'_5, f'_{5+0} \rightarrow f'_{5+0+1}$
1978, 1979 \rightarrow 1980	$f'_5, f'_6 \rightarrow f'_6$	$f'_5, f'_{5+1} \rightarrow f'_{5+1+0}$
1979, 1980 \rightarrow 1981	$f'_6, f'_6 \rightarrow f'_6$	$f'_6, f'_{6+0} \rightarrow f'_{6+0+0}$
1980, 1981 \rightarrow 1982	$f'_6, f'_6 \rightarrow f'_5$	$f'_6, f'_{6+0} \rightarrow f'_{6+0-1}$
1981, 1982 \rightarrow 1983	$f'_6, f'_5 \rightarrow f'_5$	$f'_6, f'_{6-1} \rightarrow f'_{6-1+0}$
1982, 1983 \rightarrow 1984	$f'_5, f'_5 \rightarrow f'_4$	$f'_5, f'_{5+0} \rightarrow f'_{5+0-1}$
1983, 1984 \rightarrow 1985	$f'_5, f'_4 \rightarrow f'_4$	$f'_5, f'_{5-1} \rightarrow f'_{5-1+0}$
1984, 1985 \rightarrow 1986	$f'_4, f'_4 \rightarrow f'_5$	$f'_4, f'_{4+0} \rightarrow f'_{4+0+1}$
1985, 1986 \rightarrow 1987	$f'_4, f'_5 \rightarrow f'_6$	$f'_4, f'_{4+1} \rightarrow f'_{4+1+1}$
1986, 1987 \rightarrow 1988	$f'_5, f'_6 \rightarrow f'_7$	$f'_5, f'_{5+1} \rightarrow f'_{5+1+1}$

TABLE 4

Step 5: Transformation of II-order FLR into fuzzy variable logical relationship(FVLR):

If the II-order FLR is , $A_2, A_3 \rightarrow A_4$ then we can transform it into, $A_{x_1}, A_{x_1+v(x_1)} \rightarrow A_{x_1+v(x_1)+v(x_2)}$. Here the subscripts x_1, x_2 and x_3 of the fuzzy sets, A_2, A_3 and A_4 are 2,3 and 3 respectively, i.e, $x_1 = 2, x_2 = 3$ and $x_3 = 3$. Hence $v(x_1) = x_2 - x_1 = 3 - 2 = 1, v(x_2) = x_3 - x_2 = 3 - 3 = 0$, Now the II-order FLR $A_2, A_3 \rightarrow A_3$ can be modified as $f'_2, f'_{2+1} \rightarrow f'_{2+1+0}$. Proceeding in this way we can transform the II-order FLR as given in Table-4.

Step 6: Determination of second order fuzzy logical relationship groups

Let the transformed second order FVLR obtained in step 5 having the same left hand side form a second order FLRG shown in Table-4, now their transformed II-order FLR are $G_f, G_{f+1} \rightarrow f'_{f+1+0}, G_f, G_{f+1} \rightarrow f'_{f+1+0}, G_f, G_{f+1} \rightarrow f'_{f+1+1}, G_f, G_{f+1} \rightarrow f'_{f+1+0}, G_f, G_{f+1} \rightarrow f'_{f+1+1}$ and ' $G_f, G_{f+1} \rightarrow f'_{f+1+1}$ respectively. As these transformed II- order FLR have the same left hand side " G_f, G_{f+1} " we can group them into a transformed second order fuzzy logical relationship group $G_f, G_{f+1} \rightarrow f'_{f+1+0}, f'_{f+1+0}, f'_{f+1+1}, f'_{f+1+0}, f'_{f+1+1}, f'_{f+1+1}$. Table-5 shows the transformed second order FLRGs are estimated from Table-4.

Groups	Transformed second – order fuzzy logical relationships
G_1	$G_f, G_{f-1} \rightarrow f'_{f-1+0}, f'_{f-1+0}$
G_2	$G_f, G_{f+0} \rightarrow f'_{f+0+1}, f'_{f+0+0}, f'_{f+0+0}, f'_{f+0+1}, f'_{f+0+0}, f'_{f+0-1}, f'_{f+0-1}, f'_{f+0+1}$
G_3	$G_f, G_{f+1} \rightarrow f'_{f+1+0}, f'_{f+1+0}, f'_{f+1+1}, f'_{f+1+0}, f'_{f+1+1}, f'_{f+1+1}$

TABLE 5

Step 7: Prediction: Let us consider that we want to predict the enrolment for the year 1989 by using Geometric mean, in the II-order FLR. From the Table 2, the fuzzified enrolment for the year 1987 and 1988 are A_6 and A_7 respectively. The fuzzy set A_6 and A_7 can be transformed into G_f and G_{f+1} , where $f = 6$. Based on Table 5, group 3 is selected for prediction.i.e, $G_f, G_{f+1} \rightarrow f'_{f+1+0}, f'_{f+1+0}, f'_{f+1+1}, f'_{f+1+0}, f'_{f+1+1}, f'_{f+1+1}$ putting $f = 6$, it becomes $f'_7, f'_8, f'_7, f'_7, f'_8, f'_8$ respectively. Now distinct element of $f'_7, f'_8, f'_7, f'_7, f'_8, f'_8$ are f'_7, f'_8 as the maximum grade values of the fuzzy sets A_7 and A_8 occur in the interval u_7 and u_8 . Now the Geometric mean of the interval ($f'_7 =$) u_7 and ($f'_8 =$) u_8 are 18011.60 and 19112.08. Geometric mean of the values 18011.60 and 19112.08

is 18553.68. The fuzzy set for the year 1988 (previous year of forecasting variable) is A_7 and the corresponding interval is u_7 the Geometric mean of u_7 is 18011.60. Then by using equation (3.1), the Modified forecast variable of the year 1989 is, $MFvar(1989) = 18553.68 - 18011.60 = 542.08$ Now using (3.2) the forecast enrolment for the year 1989 is $18150 + 542.08 = 18692.1$ Similarly the forecast values for the other testing years are calculated and given in the Table-6.

Year	Observed value	Forecasted value	(Forecasted value-Actual value) ²
1989	18970	18692.1	77228.41
1990	19328	19512.5	34040.25
1991	19337	19306.8	912.04
1992	18876	19315.8	193424.04
Total			305604.74

TABLE 6. The observed value and forecast value for the year 1989 to 1992 using Geometric mean

$$PE = \sqrt{\frac{305604.74}{4}} = 276.4 \text{ by using equation (2.1).}$$

Similarly, to determine the Prediction Error by using different averages like Arithmetic, Harmonic, Heronian and Root mean square are used and presented in Table 7. Also the PE are calculated and compared with existing methods by Chen [5]. For this example, it is observed that the accuracy of error is less when Harmonic mean is applied.

		forecast values by using various mean techniques with length 1100					
Year	Observed value	S.M.Chen et al [5]	Arithmeric Mean	Geometric Mean	Harmonic Mean	Heronian mean	Root mean square
1989	18970	18810	18700	18692.1	18684.2	18697.4	18708.3
1990	19328	19630	19520	19512.5	19504.9	19517.6	19527.9
1991	19337	19450	19328	19306.8	19285.9	19322.6	19349.1
1992	18876	19459	19337	19315.8	19294.3	19331.7	19358.1
RMSE		342.6	283.8	276.4	269.5	282.01	291.9

TABLE 7

5. CONCLUSION

The measures of various averages namely Arithmetic, Geometric, Harmonic, Heronian and Root Mean Square are used to predict the historical enrolments of the Alabama University. From the result shows that the present method gets the smallest average predicting error of comparing with Chen method [5] for forecasting the enrolments of the Alabama University. As well as the result verified the Property that harmonic mean \leq Geometric mean \leq Arithmetic mean.

REFERENCES

- [1] S. M. CHEN: *Forecasting enrolments based on fuzzy time-series*, Fuzzy. Set. Syst., **81**(1996), 311–319.
- [2] S. M. CHEN: *Forecasting enrolments based on high-order fuzzy time series*, Cybern. Syst. Int. J., **33**(1) (2002), 1–16.
- [3] S. M. CHEN, C. C. HAS: *A new method to forecast enrolments using fuzzy time series*, Int. J. Appl. Sci. Eng., **2**(3) (2004), 234–244.
- [4] S. M. CHEN, J. R HWANG: *Temperature prediction using fuzzy time series*, IEEE Trans. Syst. Man. Cybern. B. Cybern., **30**(2) (2000), 263–275.
- [5] C. D. CHEN, S. M. CHEN: *Handling forecasting problems based on high-order fuzzy logical relationships*, Expert Syst. Appl., **38**(2011), 3857–3864.
- [6] Y. Y. HSU, S. M. TSE, B. WU: *A new approach of bivariate fuzzy time series analysis to the forecasting of a stock index*, Int. J. Uncertain. Fuzz., **11**(6) (2003), 671–690.
- [7] K. HUANG: *Effective lengths of intervals to improve forecasting in fuzzy time-series*, Fuzzy. Set. Syst., **123**(3) (2001a), 387–394.
- [8] S. ISMAIL MOHIDEEN, U. ABUTHAHIR: *A modified method for forecasting problems based on higher-order fuzzy logical relationship*, Advance Fuzzy Set. Syst., **19**(1) (2015), 45–61.
- [9] U. ABUTHAHIR, S. ISMAIL MOHIDEEN: *A modified method for forecast gross domestic capital of India using fuzzy time series*, Malaya Journal of Matematik, **5**(1)(2020), 663–667.

DEPARTMENT OF MATHEMATICS
 JAMAL MOHAMED COLLEGE (AUTONOMOUS)
 TIRUCHIRAPPALLI-620020, TAMILNADU, INDIA
 E-mail address: abugous2004@yahoo.co.in