A modified method for forecast gross domestic capital of India using fuzzy time series

U. Abuthahir\(^1\)\(^*\) and S. Ismail Mohideen\(^2\)

Abstract
Various methods and models have been proposed by researchers using fuzzy time series for forecasting. The purpose of forecasting is to improving the prediction accuracy. This paper presents a modified method to forecast the historical data as per the Mean and Standard deviation using fuzzy time-series model. And also this present method has maximum prediction accuracy rate when compared to the already existing method.

Keywords
Fuzzy time series, Fuzzy set, Mean and standard deviation.

AMS Subject Classification
94D05, 91B84.

1. Introduction
In fact forecasting plays an idle role in daily activity. 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. Forecasting is the process of predicting future outcomes, by which decision makers analyze the data and graph to take the best decisions for the future. There are number of approaches have been developed for the last few decades. However, the good old time series methods cannot focus on the predicting problems wherein the values of time series become linguistic terms shown by fuzzy sets [12]. To overcome this situations, Song and Chissom presented the famous theory of fuzzy time series [11].

Using the above said theory of fuzzy time series, Song and Chissom [9], [10] presented the forecasting methods to forecast the historical enrollments of the Alabama University. In lieu of complicated operations, S.M. Chen [5], 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 standardize the predicting accuracy and minimize the computational overhead, also time series prediction is a more useful and enamors application problems have been applied, which also include the TAIEX\(^3\)andinventories demand [5], weather prediction [2] enrollments forecasting [1], [3], [4],[6] and [7] respectively.

The aim of this paper is to propose a method to attain better prediction accuracy. The remaining paper is arranged as follows. In section 2, Preliminaries of fuzzy time series is given. In section 3, the proposed method for forecasting problems is presented. In section 4, numerical illustration of the proposed is given. Finally the conclusions are presented in section 5.


2. Preliminaries

In this section, some standard preliminaries of the fuzzy time series model is referred from the articles [8], [9] and [10]. Kindly refer for more information about the preliminaries for [8].

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

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

where $f_{A_i} : U \to [0, 1]$. $f_{A_i}(u_j)$ is the grade or membership value of $u_i$ in the fuzzy set $A_i$ and $1 \leq j \leq n, 1 \leq i \leq n$.

Let $Y(t)(i \in \varepsilon(\mathbb{Z}^+))$, be the universe of discourse where fuzzy sets $f_i(t)$ where $(i \in \varepsilon)$ are defined in the universe of discourse $Y(t)$. Assume that $F(t)$ is a collection of $f_i(t)(i \in \varepsilon)$ then $F(t)$ is called fuzzy time series on $Y(t)(i \in \varepsilon(\mathbb{Z}^+))$.

If $F(t)$ is predict only by $F(t-1)$ is represented by $F(t-1) \rightarrow F(t)$, then there is a fuzzy relationship between $F(t) & F(t-1)$ and can be expressed as the fuzzy relational equation $F(t) = F(t-1) \circ R(t-1, t)$ where the symbol $\circ$ represents the max-min composition operator. The relation $R$ is called I-order fuzzy time series model of $F(t)$.

Suppose $F(t-1) = A_i$ and let $F(t) = A_j$, where $A_i$ and $A_j$ are fuzzy sets, then the fuzzy logical relationship (FLR) between $F(t-1)$ and $F(t)$ can be denoted by $A_i \rightarrow A_j$.

If $F(t-n) = A_{in}, \ldots, F(t-2) = A_{i2}, F(t-1) = A_{i1}$ and $F(t) = A_j$, where $A_{in}, \ldots, A_{i2}, A_{i1}$ and $A_j$ are fuzzy sets, then the higher order fuzzy logical relationship is represented by $A_{in}, \ldots, A_{i2}, A_{i1} \rightarrow A_j$.

**Prediction Error:**

The Prediction Error becomes

$$FE = \left(\frac{\text{predicted value} - \text{actual value}}{\text{actual value}}\right) \times 100$$

3. Modified Methodology

(i) Following is the alternate forecasting method as per the higher order fuzzy logical relationship [7]. In major existing algorithms the interval is considered as $[D_{min} - D_1, D_{max} + D_1]$. Here the universe of discourse using standard normal distribution range based definition is counted $U = [\mu - 3\sigma, \mu + 3\sigma]$ where $\mu$ and $\sigma$ are the Arithmetic Mean (AM) and standard deviation values (SD) of the data.

(ii) Take the higher difference with second order difference. From these two modifications the prediction error value will become very low with comparing existing method [8]. The proposed method as per the higher-order fuzzy time series model.

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

Step 2: Define linguistic terms $(A_i), i = 1ton,$ it is the fuzzy set, denoted by

$(A_i) = \left(\sum_{j=1}^{n} \frac{\mu_{ij}}{u_j}\right)$ where $(\mu_{ij})$ is the membership value of $(u_j)$ belonging to $(A_i)$ and is defined by

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

Step 3: Fuzzify each datum into a fuzzy set defined 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 fuzzy logical relationships (FLR) from the above step 2.

Step 5: Select a suitable parameter $w$, where $w(>1)$, calculate $(R^w)(t-1, t)$ and forecast the historical data as follows:

$$F(t) = F(t-1)(\times R^w)(t-1, t),$$

where $F(t)$ denotes the forecast fuzzy value of year $t$, $F(t-1)$ denotes the fuzzified data of the year $t-1$, and

$$R^w(t, t-1) = (F^T)^{(t-2)} \times F(t-1)(\cup F^T)(t-3) \times F(t-2)(\cup \ldots (\cup F^T)(t-w) \times F(t-w+1),$$

where $w$ is called the “model parameter” denoting the number of years previous $t$, “$\times$ ” is the Cartesian product operator, and $T$ is the transpose operator.

Step 6: Defuzzify the forecast fuzzy historical enrolments using neural nets. In [8], used the following Markov model to predict the given historical data. The time invariant fuzzy time series model is represented by

$$P_{t+1} = P_t \times (R_m).$$

The time-variant fuzzy time-series model is

$$P_{t+1} = (P_t) \times (R^k_m), k = 1, 2 \ldots$$

for more information kindly refer [8].

4. Numerical Illustration

Step 1: $\mu$ and $(\sigma)$ are calculated for the data in Table-1. Here $\mu = 61108.33$ and $(\sigma) = 37629.79$ the Big interval $U = [\mu - 3\sigma, \mu + 3\sigma] = [-51781.55173997.72] \approx [-51781173998]$. Let us assume that the length of each sub-interval in the universe of discourse $U$ be 20000 as in reference article [8].

Number of interval is calculated as $n = (\text{Upper limit of } U) - \text{Lower limit of } U/\text{length of interval}$.

$n = (1739985 + 51781)/20000 = 11.288 \approx 12$. Divide the universe interval of $U$ into $n = (12)$ sub-intervals of equal length.
of 20000 namely \((u_1), (u_2), \ldots, (u_{12})\).

\[
U_k = [-51781 + (i - 1) 20000, -51781 + i 20000],
\]

for \(k = 1, 2, \ldots, 12\) Therefore, the Big interval becomes \(U = [-51781, 173998]\). Let us define the fuzzy set corresponding to each interval \(u\) as \((U_1 = [-51781, -31781], U_2 = [\cdots, U_3] = [-11781, 8219], U_4 = [8219, 28219], U_5 = [28219, 48219], U_6 = [48219, 68219], U_7 = [68219, 88219], U_8 = [88219, 108219], U_9 = [108219, 128219], U_{10} = [128219, 148219], U_{11} = [148219, 168219]\) and \((U_{12} = [168219, 188219]\).

The historical data of gross domestic capital of India as presented in the following table:

<table>
<thead>
<tr>
<th>Year</th>
<th>Historical data in million repee</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980-1981</td>
<td>12105</td>
</tr>
<tr>
<td>1981-1982</td>
<td>16986</td>
</tr>
<tr>
<td>1982-1983</td>
<td>20139</td>
</tr>
<tr>
<td>1983-1984</td>
<td>21265</td>
</tr>
<tr>
<td>1984-1985</td>
<td>25600</td>
</tr>
<tr>
<td>1985-1986</td>
<td>29990</td>
</tr>
<tr>
<td>1986-1987</td>
<td>34772</td>
</tr>
<tr>
<td>1987-1988</td>
<td>33757</td>
</tr>
<tr>
<td>1988-1989</td>
<td>40136</td>
</tr>
<tr>
<td>1989-1990</td>
<td>46405</td>
</tr>
<tr>
<td>1990-1991</td>
<td>53099</td>
</tr>
<tr>
<td>1991-1992</td>
<td>57633</td>
</tr>
<tr>
<td>1992-1993</td>
<td>63997</td>
</tr>
<tr>
<td>1993-1994</td>
<td>70834</td>
</tr>
<tr>
<td>1994-1995</td>
<td>88206</td>
</tr>
<tr>
<td>1995-1996</td>
<td>90977</td>
</tr>
<tr>
<td>1996-1997</td>
<td>96187</td>
</tr>
<tr>
<td>1997-1998</td>
<td>100653</td>
</tr>
<tr>
<td>1998-1999</td>
<td>114545</td>
</tr>
<tr>
<td>1999-2000</td>
<td>134484</td>
</tr>
<tr>
<td>2000-2001</td>
<td>131505</td>
</tr>
</tbody>
</table>

Table-I

Step 3: Describing every fuzzy set \((A_i)\) as per the corrected interval from step 2, where the fuzzy set \((A_i)\) denotes a linguistic value of the historical data. i.e.,

\[
(A_1) = (low^7), (A_2) = (low^6), (A_3) = (low^5), (A_4) = (low^4),
(A_5) = (low^3), (A_6) = (low^2), (A_7) = (low), (A_8) = (medium^4),
(A_9) = (medium^3), (A_{10}) = (medium^2), (A_{11}) = (medium^1),
(A_{12}) = (high^1), (A_{13}) = (high^0), (A_{14}) = (high^0), (A_{15}) =
(high^0), (A_{16}) = (high^0), (A_{17}) = (high^0), (A_{18}) = (high^0),
(A_{19}) = (high^0).
\]

Then for every fuzzy set \((A_i)\) represented by the following linguistic values:

\[
(A_1) = 1/(u_1) + 0.5/(u_2) + 0/(u_3) + 0/(u_4) + 0/(u_5) + 0/(u_6) + 0/(u_7) + 0/(u_8) + 0/(u_9) + 0/(u_{10}) + 0/(u_{11}) + 0/(u_{12})
\]

Step 4: Obtain a fuzzy relationship as per the present historical data

\[
(A_i) \rightarrow (A_m)
\]

Table II: Fuzzy length interval

<table>
<thead>
<tr>
<th>(u_{i,j})</th>
<th>(u_{1,2})</th>
<th>(u_{2,3})</th>
<th>(u_{3,4})</th>
<th>(u_{4,5})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(8219, 13219)</td>
<td>(3219, 18219)</td>
<td>(5219, 28219)</td>
<td>(3219, 38219)</td>
<td>(38219, 48219)</td>
</tr>
</tbody>
</table>

Then the universe of discourse \(U\) is correct into following fuzzy length interval.
Step 5: For the purpose of finding if the trend goes up or goes down or is almost equal, divide the fuzzified intervals received in step 2, into four sub-intervals of same length and get first quartile as trend value goes down and the trend value increases in third quartile. The following method is used for forecasting the further value. This procedure is starting with second order difference and it is extended to \( (p^{th}) \) difference with the second order.

The \( t^{th} \) differences of the second order are represented by \( \Delta \).

\[
\Delta = ((f'_{t-1}) - (f'_{t-2})) - ((f'_{t-2}) - (f'_{t-3})) - ((f'_{t-3}) - (f'_{t-4})) - \cdots \cdot ((f'_{t-1}) - (f'_{t-2}))
\]

Condition i: If \( \Delta > 0 \) the trend goes up.
Condition ii: If \( \Delta < 0 \) the trend goes down.
Condition iii: If \( \Delta = 0 \), the trend gets changed in a constant rate, than apply Rule II or Rule III. For forecasting the future value apply Rule I, II and III.

Rule I:
Initially starting with second order difference i.e., \( \Delta = ((f'_{t-1}) - (f'_{t-2})) - ((f'_{t-2}) - (f'_{t-3})) \). Similarly for third difference with second order is 
\[
\Delta = (f'_{t-1}) - (f'_{t-2}) - (f'_{t-3}) - (f'_{t-4})
\]

Proceeding in this way we can predict the \( 4^{th}, 5^{th}, \ldots, p^{th} \) differences with second order.

Now here for predicting value for the year 1982-1983 it is impossible to estimate \( \Delta \), since the year 1979-1980 value is not available, so we can’t calculate the value of \( \Delta \). Therefore if \( |f'_{t-1} - f'_{t-2}| \) > half of length of the interval corresponding to fuzzified capital \( A_1 \) with the membership value 1, then the trend of the forecasting will be going up and forecast data falls at 0.75 point of this interval.

If \( |f'_{t-1} - f'_{t-2}| \) < half of length of the interval corresponding to fuzzy length intervals \( A_1 \) with the membership value (MF) 1, then the trend of the forecasting will be going down and the forecasted value will be 0.25 point of this interval.

If \( |f'_{t-1} - f'_{t-2}| \) = half of length of the interval corresponding to fuzzified data \( A_1 \) with MF value 1, then the forecasted value will be at the mean of this interval.

Rule II:
If \( (((f'_{t-1}) - (f'_{t-2})) - ((f'_{t-2}) - (f'_{t-3})) - ((f'_{t-3}) - (f'_{t-4})) - \cdots \cdot ((f'_{t-1}) - (f'_{t-2}))) (\times m) + (f'_{n-1})) \) or
\[
(((f'_{n-1}) - (f'_{n-2})) - ((f'_{n-2}) - (f'_{n-3})) - ((f'_{n-3}) - (f'_{n-4})) - \cdots \cdot ((f'_{n-1}) - (f'_{n-2}))) (\times m)
\]
where \( m \) represent the number of difference, falls in the interval relating to the fuzzified data \( A_i \) with MF 1, then the forecasting trend will increase with the (0.75) point of that interval.

Rule III:
If the \( (((f'_{t-1}) - (f'_{t-2})) - ((f'_{t-2}) - (f'_{t-3})) - ((f'_{t-3}) - (f'_{t-4})) - \cdots \cdot ((f'_{t-1}) - (f'_{t-2}))) (\times m) \) lies in the fuzzified capital data \( A_1 \) with MF 1 then the forecasting trend will decrease with the (0.25) point of that interval.

Rule IV:
If there is no value occurred in the above conditions II and III, then the prediction value is must be the Arithmetic mean (mean) of the fuzzified interval \( A_i \) with grade value 1.

Forecasting:
Suppose we want to forecast the year 1984-1985, here the present term is \( P_t = 25600 \) and which is in the fuzzy set \( A_2 \). Then \( f'_{t-1} = 21265, f'_{t-2} = 20139, f'_{t-3} = 16986, f'_{t-4} = 12105. \) For given information we need three differences.

Therefore the third differences is \( \Delta = (((f'_{t-1}) - (f'_{t-2})) - ((f'_{t-2}) - (f'_{t-3})) - ((f'_{t-3}) - (f'_{t-4})) = 6908. \) Here the number of difference \( (m = 3) \), since by rule III which is lie in the fuzzy set \( A_4 \). According to the rule, the trend is going downward direction with 0.25 point (third quartile) of the interval will take to forecast. Therefore 0.25 point of \( A_4 \) is 1250. Then the forecast value of the year 1984-1985 is 1250 + 23219 = 24469. Similarly the other values are calculated and presented in the following table IV.
A modified method for forecast gross domestic capital of India using fuzzy time series — 667/667

Table-IV: Determination of error by proposed method and comparing with existing method [8]

From the above table shows that the result the Prediction Error of existing [8] is \(4.99\) \%.
Therefore, Prediction Error (PE) of this present method is \(3.13\) \%.

5. Conclusion

In this paper, presented a method as per the higher-order fuzzy logical relationship for forecasting the Gross Domestic Capital of India using Mean, standard deviation and I and III quartiles. From the result that forecasting error of this method is much smaller than comparing with an already present method [8] due to the fuzzy time series model.

References

[1] C.D. Chen and S.M. Chen, Handling forecasting problems based on high-order fuzzy Logical relation-