Multivariate High Order Fuzzy Time Series Forecasting for Car Road Accidents

Tahseen A. Jilani, S. M. Aqil Burney, and C. Ardil

Abstract—In this paper, we have presented a new multivariate fuzzy time series forecasting method. This method assumes m-factors with one main factor of interest. History of past three years is used for making new forecasts. This new method is applied in forecasting total number of car accidents in Belgium using four secondary factors. We also make comparison of our proposed method with existing methods of fuzzy time series forecasting. Experimentally, it is shown that our proposed method perform better than existing fuzzy time series forecasting methods. Practically, actuaries are interested in analysis of the patterns of causalities in road accidents. Thus using fuzzy time series, actuaries can define fuzzy premium and fuzzy underwriting of car insurance and life insurance for car insurance. National Institute of Statistics, Belgium provides region of risk classification for each road. Thus using this risk classification, we can predict premium rate and underwriting of insurance policy holders.

Keywords—Average forecasting error rate (AFER), fuzziness of fuzzy sets Fuzzy, If-Then rules, Multivariate fuzzy time series.

I. INTRODUCTION

In our daily life, people often use forecasting techniques to model and predict economy, population growth, stocks, insurance/ re-insurance, portfolio analysis and etc. However, in the real world, an event can be affected by many factors. Therefore, if we consider more factors for prediction, with higher complexity then we can get better forecasting results.

During last few decades, various approaches have been developed for time series forecasting. Among them ARMA models and Box-Jenkins model building approaches are highly famous.


The rest of this paper is organized as follows. In section 2, brief review of fuzzy time series is given. In section 3, we present the new method for fuzzy time series modeling. Experimental results are performed in section 4. The conclusions are discussed in section 5.

In this paper, we present a new modified method to predict total number of annual car road accidents based on the m-factors high-order fuzzy time series. This method provides a general framework for forecasting that can be increased by increasing the stochastic fuzzy dependence [3]. For simplicity of computation, we have used triangular membership function. The proposed method constructs m-factor high-order fuzzy logical relationships based on the historical data to increase the forecasting accuracy rate. Our proposed forecasting method for fuzzy time series gives better results as compared to [4], [5] and [10].

II. FUZZY TIME SERIES

Time series analysis plays vital role in most of the actuarial related problems. As most of the actuarial issues are born with uncertainty, therefore, each observation of a fuzzy time series is assumed to be a fuzzy variable along with associated membership function. Based on fuzzy relation, and fuzzy inference rules, efficient modeling and forecasting of fuzzy time series is possible, see [1] and [9]. This field of fuzzy time series analysis is not very mature due to the time and space complexities in most of the actuarial related issue, thus we can extend this concept for many antecedents and single consequent. For example, in designing two-factor kth-order fuzzy time series model with X be the primary and Y be second factor. We assume that there are k antecedent (\\(X_1,Y_1\\),\\(X_2,Y_2\\),...\\(X_k,Y_k\\)) and one consequent \(X_{k+1}\\).

\[ \begin{align*}
&\\text{If } (X_1=y_1,Y_1=y_1),(X_2=y_2,Y_2=y_2), \\
&\ldots(X_k=y_k,Y_k=y_k) \rightarrow (X_{k+1}=y_{k+1}) 
\end{align*} \] (1)

In the similar way, we can define m-factor \(i=1,2,\ldots,m\\) and kth order fuzzy time series as

Manuscript received November 25, 2006. This work was supported in part by the Higher Education Commission of Pakistan.

T. A. Jilani is Lecturer in the Department of Statistics, University of Karachi, Karachi-75100, Pakistan (phone: +923333040963; e-mail: tahseen@jilani@ieee.org)

S. M. Aqil Burney is Professor in the Department of Computer Science, University of Karachi, Karachi-75100, Pakistan (e-mail: burney@computer.org).

C. Ardil is with National Academy of Aviation, AZ1045, Baku, Azerbaijan, Bina, 25th km, NAA (e-mail: cemalardil@gmail.com).
TABLE I

YEARLY CAR ACCIDENTS MORTALITIES AND VICTIMS FROM 1974 TO 2004

<table>
<thead>
<tr>
<th>Year</th>
<th>Killed (X)</th>
<th>Mortally wounded (Y1)</th>
<th>Died 30 days (Y2)</th>
<th>Severely wounded (Y3)</th>
<th>Light casualties (Y4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>953</td>
<td>141</td>
<td>1,094</td>
<td>5,949</td>
<td>41,627</td>
</tr>
<tr>
<td>2003</td>
<td>1,035</td>
<td>101</td>
<td>1,136</td>
<td>6,898</td>
<td>42,445</td>
</tr>
<tr>
<td>2002</td>
<td>1,145</td>
<td>118</td>
<td>1,263</td>
<td>6,834</td>
<td>39,522</td>
</tr>
<tr>
<td>2001</td>
<td>1,288</td>
<td>90</td>
<td>1,378</td>
<td>7,319</td>
<td>38,747</td>
</tr>
<tr>
<td>2000</td>
<td>1,253</td>
<td>103</td>
<td>1,356</td>
<td>7,990</td>
<td>39,719</td>
</tr>
<tr>
<td>1999</td>
<td>1,173</td>
<td>126</td>
<td>1,299</td>
<td>8,461</td>
<td>41,841</td>
</tr>
<tr>
<td>1998</td>
<td>1,224</td>
<td>121</td>
<td>1,345</td>
<td>8,784</td>
<td>41,038</td>
</tr>
<tr>
<td>1997</td>
<td>1,150</td>
<td>105</td>
<td>1,255</td>
<td>9,229</td>
<td>39,594</td>
</tr>
<tr>
<td>1996</td>
<td>1,122</td>
<td>115</td>
<td>1,237</td>
<td>9,123</td>
<td>38,390</td>
</tr>
<tr>
<td>1995</td>
<td>1,228</td>
<td>109</td>
<td>1,337</td>
<td>10,267</td>
<td>39,140</td>
</tr>
<tr>
<td>1994</td>
<td>1,415</td>
<td>149</td>
<td>1,564</td>
<td>11,160</td>
<td>40,294</td>
</tr>
<tr>
<td>1993</td>
<td>1,346</td>
<td>171</td>
<td>1,517</td>
<td>11,680</td>
<td>41,736</td>
</tr>
<tr>
<td>1992</td>
<td>1,380</td>
<td>173</td>
<td>1,553</td>
<td>12,113</td>
<td>41,772</td>
</tr>
<tr>
<td>1991</td>
<td>1,471</td>
<td>209</td>
<td>1,680</td>
<td>12,965</td>
<td>43,579</td>
</tr>
<tr>
<td>1990</td>
<td>1,574</td>
<td>190</td>
<td>1,764</td>
<td>13,864</td>
<td>46,818</td>
</tr>
<tr>
<td>1989</td>
<td>1,488</td>
<td>312</td>
<td>1,800</td>
<td>14,515</td>
<td>46,667</td>
</tr>
<tr>
<td>1988</td>
<td>1,432</td>
<td>339</td>
<td>1,771</td>
<td>14,029</td>
<td>45,956</td>
</tr>
<tr>
<td>1987</td>
<td>1,390</td>
<td>380</td>
<td>1,770</td>
<td>13,809</td>
<td>44,090</td>
</tr>
<tr>
<td>1986</td>
<td>1,456</td>
<td>330</td>
<td>1,786</td>
<td>13,764</td>
<td>42,965</td>
</tr>
<tr>
<td>1985</td>
<td>1,308</td>
<td>352</td>
<td>1,660</td>
<td>13,287</td>
<td>39,879</td>
</tr>
<tr>
<td>1984</td>
<td>1,369</td>
<td>363</td>
<td>1,732</td>
<td>14,471</td>
<td>45,456</td>
</tr>
<tr>
<td>1983</td>
<td>1,479</td>
<td>412</td>
<td>1,891</td>
<td>14,864</td>
<td>42,023</td>
</tr>
<tr>
<td>1982</td>
<td>1,464</td>
<td>406</td>
<td>1,870</td>
<td>14,601</td>
<td>40,936</td>
</tr>
<tr>
<td>1981</td>
<td>1,564</td>
<td>454</td>
<td>2,018</td>
<td>15,091</td>
<td>41,915</td>
</tr>
<tr>
<td>1980</td>
<td>1,616</td>
<td>557</td>
<td>2,173</td>
<td>15,915</td>
<td>42,670</td>
</tr>
<tr>
<td>1979</td>
<td>1,572</td>
<td>544</td>
<td>2,116</td>
<td>15,750</td>
<td>42,346</td>
</tr>
<tr>
<td>1978</td>
<td>1,644</td>
<td>728</td>
<td>2,372</td>
<td>16,645</td>
<td>44,797</td>
</tr>
<tr>
<td>1977</td>
<td>1,597</td>
<td>701</td>
<td>2,298</td>
<td>15,830</td>
<td>44,995</td>
</tr>
<tr>
<td>1976</td>
<td>1,536</td>
<td>728</td>
<td>2,264</td>
<td>16,057</td>
<td>44,227</td>
</tr>
<tr>
<td>1975</td>
<td>1,460</td>
<td>701</td>
<td>2,161</td>
<td>15,792</td>
<td>42,423</td>
</tr>
<tr>
<td>1974</td>
<td>1,574</td>
<td>819</td>
<td>2,393</td>
<td>16,306</td>
<td>44,640</td>
</tr>
</tbody>
</table>

If \( X_1=x_1, X_2=x_2, \ldots, X_k=x_k \),
\( X_{21}=x_{21}, X_{22}=x_{22}, \ldots, X_{2k}=x_{2k}, \ldots \),
\( X_{m1}=x_{m1}, X_{m2}=x_{m2}, \ldots, X_{mk}=x_{mk} \),
then \( X_{m+1,k+1}=x_{m+1,k+1} \)
for \( i=1, 2, \ldots, m \), \( j=1, 2, \ldots, k \)

Now, we formally give details of proposed method in section 3.

III. NEW METHOD OF FORECASTING USING FUZZY TIME SERIES

Let \( Y(t), t=\ldots,-0,1,2,\ldots \) be the universe of discourse and \( Y(t) \subseteq R \). Assume that \( f(t), t=1,2,\ldots \) is defined in the universe of discourse \( Y(t) \) and \( F(t) \) is a collection of \( f(t), t=1,2,\ldots \), then \( F(t) \) is called a fuzzy time series of \( Y(t), t=1,2,\ldots \). Using fuzzy relation, we define \( F(t)=F(t-1)+R(t,t-1) \), where \( R(t,t-1) \) is a fuzzy relation and \( "o" \) is the max–min composition operator, then \( F(t) \) is caused by \( F(t-1) \) where \( F(t) \) and \( F(t-1) \) are fuzzy sets. For forecasting purpose, we can define relationship among present and future state of a time series with the help of fuzzy sets. Assume the fuzzified data of the \( i \text{th} \) and \( (i+1)\text{th} \) day are \( A_j \) and \( A_k \), respectively, where \( A_j, A_k \in U \), then \( A_j \rightarrow A_k \) represented the fuzzy logical relationship between \( A_j \) and \( A_k \).

Let \( F(t) \) be a fuzzy time series. If \( F(t) \) is caused by \( F(t-1), F(t-2), \ldots, F(t-n) \), then the fuzzy logical relationship is represented by

\[
F(t-n), \ldots, F(t-2), F(t-1) \rightarrow F(t)
\]

is called the one-factor nth order fuzzy time series forecasting model. Let \( F(t) \) be a fuzzy time series. If \( F(t) \) is caused by \( \{F(t-1),F(t-2)\}, \{F(t-2),F(t-2)\} \ldots \{F(t-n),F(t-n)\} \), then this fuzzy logical relationship is represented by

\[
\{F(t-n),F(t-n)\}, \ldots, \{F(t-2),F(t-2)\}, \{F(t-1),F(t-1)\} \rightarrow F(t)
\]

is called the two-factors nth order fuzzy time series forecasting model, where \( F_1(t) \) and \( F_2(t) \) are called the main factor and the secondary factor fuzzy time series respectively. In the similar way, we can define m-factor nth order fuzzy logical relationship as

\[
\{F_1(t-n),F_2(t-n)\}, \ldots, \{F_m(t-m)\} \rightarrow F(t)
\]

Here \( F_1(t) \) is called the main factor and \( F_2(t),F_3(t)\ldots F_m(t) \) are called secondary factor fuzzy time series'. Here we can implement any of the fuzzy membership function to define the fuzzy time series in above equations. Comparative study by using different membership functions is also possible. We have used triangular membership function due to low computational cost. Using fuzzy composition rules, we establish a fuzzy inference system for fuzzy time series forecasting with higher accuracy. The accuracy of forecast can be improved by considering higher number of factors and higher dependence on history.

Now we present an extended method for handling forecasting problems based on m-factors high-order fuzzy time series. The proposed method is now presented as follows.

**Step 1** Define the universe of discourse \( U \) of the main factor \( U=[D_{\min}-D_1, D_\max-D_2] \), where \( D_{\min} \) and \( D_\max \) are the minimum and the maximum values of the main factor of the known historical data, respectively, and \( D_1, D_2 \) are two...
proper positive real numbers to divide the universe of discourse into \( n \) equal length intervals \( u_1, u_2, \ldots, u_n \). Define the universes of discourse \( V_i, \, i=1,2,\ldots,m-1 \) of the secondary-factors

\[
V_i = \left[ \left( E_i \right)_{\min}^E \left( E_i \right)_{\max}^E \right],
\]

where

\[
\left( E_i \right)_{\min}^E = \left( E_i \right)_{\min}^E, \quad \left( E_i \right)_{\max}^E = \left( E_i \right)_{\max}^E.
\]

Step 2) Define the linguistic term \( A_i \) represented by fuzzy sets of the main factor shown as follows:

\[
A_i = \left[ \begin{array}{c} 0.5 / v_{i1} + 0.5 / v_{i1} + 0.5 / v_{i1} + \ldots + 0.5 / v_{i1} + 0.5 / v_{i1} \end{array} \right]
\]

Similarly, for \( i \)th secondary fuzzy time series, we define the linguistic term \( B_{i,j}, \, i=1,2,\ldots,m-1, \, j=1,2,\ldots,n \) represented by fuzzy sets of the secondary-factors,

\[
B_{i,j} = \left[ \begin{array}{c} 0.5 / v_{i1} + 0.5 / v_{i1} + 0.5 / v_{i1} + \ldots + 0.5 / v_{i1} + 0.5 / v_{i1} \end{array} \right]
\]

Step 3) Fuzzify the historical data described as follows. Find the interval \( \left[ V_i, j \right], \, j=1,2,\ldots,p \) to which the value of the main factor belongs

Case 1) If the value of the main factor belongs to \( u_1, l=2,3,\ldots,p-1 \) then the value of the main factor is fuzzified into \( 0.5 / A_{i-1} + 1 / A_{i} + 0.5 / A_{i+1} \), denoted by \( Y_j \).

Case 3) If the value of the main factor belongs to \( u_{p-1}, l=2,3,\ldots,p \) then the value of the main factor is fuzzified into \( 0.5 / A_{n-2} + 0.5 / A_{n-1} + 1 / A_{n} \), denoted by \( Y_n \).

Now, for \( p \)th secondary-factor, find out the interval \( \left[ V_{i,j} \right] \) to which the value of the secondary-factor belongs.

Case 1) If the value of the \( i \)th secondary-factor belongs to \( v_{i,j}, l=2,3,\ldots,p-1 \) then the value of the \( i \)th secondary-factor is fuzzified into \( 0.5 / B_{i,j-1} + 0.5 / B_{i,j} + 1 / B_{i,j+1} \), denoted by \( Y_{i,j} \).

Case 2) If the value of the \( i \)th secondary-factor belongs to \( v_{i,j}, l=2,3,\ldots,n-1 \) then the value of the \( i \)th secondary-factor is fuzzified into \( 0.5 / B_{i,j-2} + 0.5 / B_{i,j-1} + j=2,3,\ldots,n-1 \), denoted by \( Y_{i,j} \), where \( j=2,3,\ldots,n-1 \).
Case 3) If the value of the ith secondary-factor belongs to $v_{i,p}$, then the value of the secondary-factor is fuzzified into $0/_{B_{i,1}}+0.5/_{B_{i,2}}+1/_{B_{i,1}}$, denoted by $y_{i,t}$.

Step 4) Get the m-factors kth-order fuzzy logical relationships based on the fuzzified main and secondary factors from the fuzzified historical data obtained in Step 3). If the fuzzified historical data of the main-factor of ith day is $X_j$, then construct the m-factors kth-order fuzzy logical relationships, 

$$
\left\{X_{j-k}, X_{j-k-1}, ..., X_{j-m}, X_{j-m-k+1}, ..., X_{j-m-k+1}, X_{j-m-k+2}, ..., X_{j-m-k+2}\right\} = \left\{X_{j-2}, X_{j-2}, ..., X_{j-m-1}\right\}, \rightarrow X_j
$$

where $j > k$. $X_{j-k}$ shows the k-step dependence of jth value of main factor $X_j$, $y_{i,j-k}, i=1, ..., m-1$, $j=1, ..., k$. Then, divide the derived fuzzy logical relationships into fuzzy logical relationship groups based on the current states of the fuzzy logical relationships. The secondary factors acts like a secondary component to the m-dimensional state vector and is used in Step 5).

Step 5) For m-factor kth order fuzzy logical relationship, the forecasted value of day j based on history of third order is calculated as follows,

$$
I_j=\sum_{j=1}^{j+1} w_i \frac{w_j}{j} + \frac{w_j}{j+1}
$$

(9)

Where $u_{j-1}, u_j$ and $u_{j+1}$ are the midpoints of the intervals $u_{j-1}, u_j$ respectively. Above forecasting formula fulfills the axioms of fuzzy sets like monotonicity, boundary conditions, continuity and idempotency. For measurement of accuracy of forecasting for fuzzy time series forecasting, we use average forecasting error rate (AFER) as the performance criteria, defined as

$$
AFER = \frac{1}{n} \sum_{j=1}^{n} \left| \frac{\text{Forecasted Value of Day } j - \text{Actual Value of Day } j}{\text{Actual Value of Day } j} \right| \times 100\% \tag{10}
$$

IV. EXPERIMENT

In this experiment, our goal is to extend the work of [6]. We have applied this new technique on car road accident data taken from National Institute of Statistics, Belgium for the period of 1974-2005. In this data, the main factor of interest is the yearly road accident causalities and secondary factors are mortally wounded, died within one month, severely wounded and light casualties.

We assumed eight intervals of equal length for the main and secondary fuzzy time series. For main factor, we assume $D_{min}$=953 and $D_{max}$=1644, thus for main factor time series we get $U = [850, 1650]$. Similarly for secondary factors $y_1, y_2, y_3$ and $y_4$, we assumed that $E_{min}=[90, 1094, 5949, 38390]$ and $E_{max}=[819, 2393, 16645, 46818]$ to determine $v_1, v_2, v_3, v_4$. Selection of $D_{min}$, $D_{max}$, $E_{min}$ and $E_{max}$ have significant effects on the accuracy of this new method. We can introduce learning to stabilize the heuristic selection of these constants.

Using (6) and (7), we formed fuzzy times series from main and secondary factors. Therefore, each observation of a time series is now represented by a combination of fuzzy sets. Using (10), we calculated the forecasted values corresponding to each actual value of the main factor time series in III. Using equation (11) for AFER, we formed table IV, showing a comparison of actual and forecasted values. Finally, we have compared proposed method with [6].
From Table IV, we can see that our proposed method is better than [6]. As the work of Lee et. al. [6] outperformed the work of [4], [5] and [10], so, indirectly we can conclude that our general class of methods for fuzzy time series modeling and forecasting.

Furthermore, we have shown fuzziness of fuzzy observations by presenting each datum of the main series as composed of many fuzzy sets. Thus, fuzzy time series modeling extends to type-II fuzzy time series modeling. The type-II defuzzified forecasted values ($F_t^*$) may also be calculated using some other method, e.g. learning rules from fuzzy time series.

**ACKNOWLEDGMENT**

The authors are very thankful for the kind support of National Institute of Statistics, Belgium.

**REFERENCES**


S. M. Aqil Burney received the B.Sc.(Physics, Mathematics, Statistics) from D. J. College affiliated with the University of Karachi in 1970; First class first M.Sc. (Statistics,) from Karachi University in 1972 with M.Phil. in 1983 with specialization in Computing, Simulation and stochastic modeling in from risk management. Dr. Burney received Ph.D. degree in Mathematics from Strathclyde University, Glasgow with specialization in estimation, modeling and simulation of multivariate Time series models using algorithmic approach with software development.

He is currently professor and approved supervisor in Computer Science and Statistics by the High education Commission Government of Pakistan. He is also the project director of Umair Basha Institute of Information technology (UBIT). He is also member of various higher academic boards of different universities of Pakistan. His research interest includes AI, Soft computing neural networks, fuzzy logic, data mining, statistics, simulation and stochastic modeling of mobile communication system and networks and network security. He is author of three books, various technical reports and supervised more than 100 software/Information technology projects of Masters level degree programs and project director of various projects funded by Government of Pakistan.

He is member of IEEE (USA), ACM (USA) and fellow of Royal Statistical Society United Kingdom and also a member of Islamic society of Statistical Sciences. He is teaching since 1973 in various universities in the field of Econometric, Bio-Statistics, Statistics, Mathematic and Computer Science. He has vast education management experience at the University level. Dr. Burney has received appreciations and awards for his research and as educationist.

Tahseen A. Jilani received the B.Sc.(Computer Science, Mathematics, Statistics), First class second M.Sc. (Statistics) and MA(Economics) in 1998, 2001 and 2003 respectively. Since 2003, he is Ph.D. research fellow in the Department of Computer Science, University of Karachi. His research interest includes AI, neural networks, soft computing, fuzzy logic, Statistical data mining and simulation. He is teaching since 2002 in the fields of Statistics, Mathematic and Computer Science.