A new approach based on the optimization of the length of intervals in fuzzy time series

5 authors, including:

Erol Egrioglu
Giresun University
123 PUBLICATIONS **1,984** CITATIONS

SEE PROFILE

Cagdas Hakan Aladag
Hacettepe University
95 PUBLICATIONS **1,755** CITATIONS

SEE PROFILE

Murat Alper Basaran
Alanya Alaaddin Keykubat University
43 PUBLICATIONS **720** CITATIONS

SEE PROFILE

Ufuk Yolcu
Giresun University
69 PUBLICATIONS **1,415** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Bidding Strategies in Double Auction Market with Multi-Agent Simulations View project

not related to any project View project

A new approach based on the optimization of the length of intervals in fuzzy time series

Erol Egrioglu[a], Cagdas Hakan Aladag[b,*], Murat A. Basaran[c], Ufuk Yolcu[a] and Vedide R. Uslu[a]
[a]*Department of Statistics, University of Ondokuz Mayıs, Samsun, Turkey*
[b]*Department of Statistics, University of Hacettepe, Ankara, Turkey*
[c]*Department of Mathematics, University of Nigde, Nigde, Turkey*

Abstract. In fuzzy time series analysis, the determination of the interval length is an important issue. In many researches recently done, the length of intervals has been intuitively determined. In order to efficiently determine the length of intervals, two approaches which are based on the average and the distribution have been proposed by Huarng [4]. In this paper, we propose a new method based on the use of a single variable constrained optimization to determine the length of interval. In order to determine optimum length of interval for the best forecasting accuracy, we used a MATLAB function which is employing an algorithm based on golden section search and parabolic interpolation. Mean square error is used as a measure of forecasting accuracy so the objective function value is mean square error value for forecasted observations. The proposed method was employed to forecast the enrollments of the University of Alabama to show the considerable outperforming results.

Keywords: Forecasting, fuzzy sets, fuzzy time series, length of interval, optimization

1. Introduction

Fuzzy time series analysis performs well than the conventional correspond since they improve significantly the forecasting results. There is no need to have at least 50 observations and the linearity assumption as the conventional ones do. Therefore these methods gradually attract researchers' attention since they are simply applicable and there are no any other restrictions.

Fuzzy time series procedure firstly proposed in [10–12] consists of three stages. In the first stage observations must be fuzzified. Secondly fuzzy relationships are evaluated from fuzzy observations. And the final stage is the defuzzification stage. Since then many researches have been done to improve these stages to get more accurate forecasting results. Chen [2] proposed a simple method for discovering fuzzy relationships with respect to the method introduced by Song and Chissom. Chen's method simplified the arithmetic operation process. The two of contributed papers to the identification of fuzzy relationships were done by Huarng and Yu [7, 8]. In the paper done by Huarng [4], he tried to improve the stage of the fuzzification. He pointed out that an affective length of interval could significantly improve forecasting results. Hence he proposed a method based on the average and the distribution to choose a more affective length. In that method, the length of interval is fixed. In preceding study, Huarng [5] proposed another method in which the length of interval is not fixed. Huarng and Yu [6] also used a dynamic approach for adjusting lengths of interval.

In this paper we propose a novel method, for finding an affective length of interval, which minimizes mean square error (MSE) by using a single variable constrained optimization. The objective function value is MSE value obtained from forecasted and actual observations. By minimizing MSE, we try to determine the length of interval, that is, we aim to increase forecasting

*Corresponding author. Cagdas Hakan Aladag, Department of Statistics, University of Hacettepe, Ankara 06800, Turkey. E-mail: chaladag@gmail.com.

accuracy. In order to find the optimal length of interval, a MATLAB function which is employing an algorithm based on golden section search and parabolic interpolation is used. To show what we achieved, we applied the method to a well known data which is the enrollments of the University of Alabama. The results are compared to the results from previous studies.

Section 2 is about the general knowledge of fuzzy time series. In section 3 Chen's method is briefly presented. Section 4 explains our proposed approach and presents the empirical results of an application of the real data. Section 5 finally offers a general conclusion.

2. Fuzzy time series

The definition of fuzzy time series was firstly introduced by [10–12]. In fuzzy time series approximation, you do not need various theoretical assumptions just as you need in conventional procedures. The most important advantages of fuzzy time series approaches are to be able to work with a few observations and not to require the linearity assumption. The same general definitions of fuzzy time series are given as follows:

Let U be the universe of discourse, where $U = \{u_1, u_2, \ldots, u_b\}$. A fuzzy set A_i of U is defined as $A_i = f_{A_i}(u_1)/u_1 + f_{A_i}(u_2)/u_2 + \cdots + f_{A_i}(u_b)/u_b$, where f_{A_i} is the membership function of the fuzzy set A_i; $f_{A_i} : U \to [0, 1]$. u_a is a generic element of fuzzy set A_i; $f_{A_i}(u_a)$ is the degree of belongingness of u_a to A_i; $f_{A_i}(u_a) \in [0, 1]$ and $1 \le a \le b$.

Definition 1. Fuzzy time series Let $Y(t)(t = \ldots, 0, 1, 2, \ldots)$ a subset of real numbers, be the universe of discourse by which fuzzy sets $f_j(t)$ are defined. If $F(t)$ is a collection of $f_1(t)$, $f_2(t)$, ... then $F(t)$ is called a fuzzy time series defined on $Y(t)$.

Definition 2. Fuzzy time series relationships assume that $F(t)$ is caused only by $F(t-1)$, then the relationship can be expressed as: $F(t) = F(t-1) * R(t, t-1)$, which is the fuzzy relationship between $F(t)$ and $F(t-1)$, where $*$ represents as an operator. To sum up, let $F(t-1) = A_i$ and $F(t) = A_j$. The fuzzy logical relationship between $F(t)$ and $F(t-1)$ can be denoted as $A_i \to A_j$ where A_i refers to the left-hand side and A_j refers to the right-hand side of the fuzzy logical relationship. Furthermore, these fuzzy logical relationships can be grouped to establish different fuzzy relationship.

3. The algorithm of Chen's model

In [2], Chen has improved the approximation given by [10–12]. He proposes a method which uses a simpler operation instead of complex matrix operations, in the establishment step of fuzzy relationships. In [2], the algorithm of Chen's method is given as follows:

Step 1. Define the universe of discourse and intervals for rules abstraction.

Based on the domain issue, The universe of discourse can be defined as: $U = [starting, ending]$. As the length of interval is determined, U can be partitioned into several equally length intervals.

Step 2. Define fuzzy sets based on the universe of discourse and fuzzify the historical data.

Step 3. Fuzzify observed rules.

Step 4. Establish fuzzy logical relationships and group them based on the current states of the data of the fuzzy logical relationships.

For example, $A_1 \to A_2, A_1 \to A_1, A_1 \to A_3$, can be grouped as: $A_1 \to A_2, A_3, A_1$.

Step 5. Forecast.

Let $F(t-1) = A_i$.

Case 1: There is only one fuzzy logical relationship in the fuzzy logical relationship sequence. If $A_i \to A_j$, then $F(t)$, forecast value, is equal to A_j.

Case 2: If $A_i \to A_i, A_j, \ldots, A_k$, then $F(t)$, forecast value, is equal to A_i, $A_j, \ldots, A_k$.

Case 3: If $A_i \to$ empty, then $F(t)$, forecast value, is equal to A_i.

Step 6. Defuzzify.

Apply "Centroid" method to get the results. This procedure (also called center of area, center of gravity) is the most often adopted method of defuzzification.

4. The method based on the optimization of MSE

In fuzzy time series approaches for the aim of forecasting the length of intervals affects the forecasting performance. Hence it is important to choose an affective length of intervals for improving forecasting in fuzzy time series. The method we propose optimizes the length of interval by following with the algorithm of Chen's method [2]. In the optimization process, we

used a MATLAB function called "fminbnd" which min-
imizes MSE. "fminbnd" is used to find minimum of
single-variable function on fixed interval. It finds a min-
imum for a problem specified by

$$\min_x f(x) \text{ such that } x_1 < x < x_2.$$

x, x_1, and x_2 are scalars and $f(x)$ is a function
that returns a scalar. In MATLAB, $\hat{x}$ = fminbnd
$(f(x), x_1, x_2)$ returns a value $\hat{x}$ that is a local mini-
mum of the scalar valued function $f(x)$ in the interval
$x_1 < x < x_2$. In other words, to find the minimum of
the function $f(x)$ on the interval (x_1, x_2),

$$a = \text{fminbnd } (f(x), x_1, x_2)$$

can be used in MATLAB. $f(a)$ gives the local minimum
value in the interval (x_1, x_2).

The algorithm used by fminbnd is based on golden
section search introduced by Kiefer [9] and parabolic
interpolation. Unless the left endpoint x_1 is very close
to the right endpoint x_2, fminbnd never evaluates $f(x)$ at
the endpoints, so $f(x)$ need only be defined for x in the
interval $x_1 < x < x_2$. If the minimum actually occurs at
x_1 or x_2, fminbnd returns an interior point at a distance
of no more than 2*TolX from x_1 or x_2, where TolX is the
termination tolerance. See [1] or [3] for details about
the algorithm.

For the aim of a comparative study, we used the
proposed method and methods in [4] average and distri-
bution based length to determine the length of interval
in Chen's method [2] when the data of the enroll-
ments of the University of Alabama is being analyzed.
Enrollment data are presented in Table 1 and the results
obtained from the mentioned methods are summarized
in Table 2. MSE for forecasted observations is used as a
measure of forecasting accuracy so the objective func-
tion value is equal to MSE value produced from Chen's
method [2]. Using with the algorithm of a single vari-

Table 1
Enrollment data

Year	Actual	Year	Actual
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 2
The obtained values of MSE for the enrollment data

Length of interval	200	273.2346[a]	300[b]	400[c]	500
MSE	104640	66661	78792	124707	173453
Length of interval	600	700	800	900	1000
MSE	254592	222557	365045	246892	407521

[a]Results of the proposed method; [b]Average based length in
[4]; [c]Distribution based length in [4].

able constrained optimization to minimize MSE value
via MATLAB function called "fminbnd" over the inter-
val between 200 and 500, the optimal length of interval
was obtained. We minimize the MSE value by opti-
mizing the length of interval so that we increase the
forecasting accuracy. We use the function "fminbnd"
as follows:

$$x^* = \text{fminbnd } \left(f_{MSE}(x), \ 200, \ 500 \right)$$

where $f_{MSE}(x)$ gives the MSE value obtained from fore-
casted and actual values for the length of interval x when
Chen's method [2] is used. The left and the right end
points are taken as 200 and 500, respectively. This used
function returns x^* that is a minimum point of the func-
tion $f_{MSE}(x)$ in the interval $200 < x < 500$. We found x^*
as 273.2346. In addition, the running time was lower
than one minute.

When the function "fminbnd" is used in the opti-
mization process, there are two parameters which are
important in terms of fuzzy time series and optimization
processes. These parameters are the left and the right
end points. In fuzzy time series approaches, a key point
in choosing an effective length of interval is that they
should not be too large or small. If a too large right end
point value is taken, a too large length of interval can be
produced by the function "fminbnd". When this value
of length of interval is used in fuzzy time series analysis,
lower MSE values can be obtained but this causes no
fluctuations in the fuzzy time series. On the other hand,
if a too small left end point value is taken, a too small
length of interval can be obtained and this will dimin-
ish the meaning of fuzzy time series [2]. Choosing the
left and the right end points also affects the optimiza-
tion process. If the difference between these parameters
is too small, the possibility of founding a satisfactory
value for the length of interval will be very low. How-
ever, if the difference between these parameters is too
large, the searching time will increase. Therefore, the
constrained optimization technique was employed over
the interval between 200 and 500 in order not to find
a very small or very large interval in the optimization
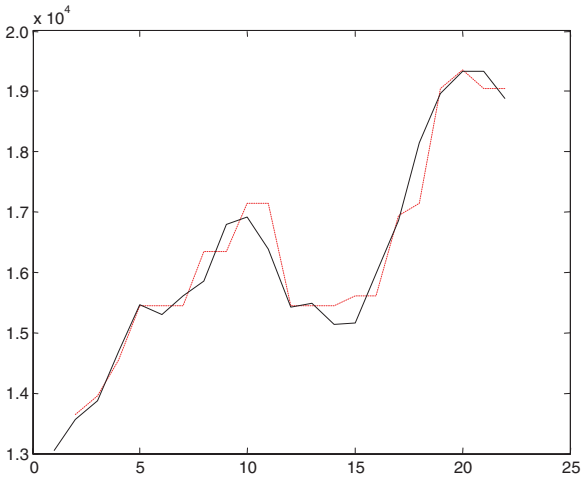process. In other words, we take the left and the right

Fig. 1. The graph of actual values and forecasts based on the 300 length of interval.
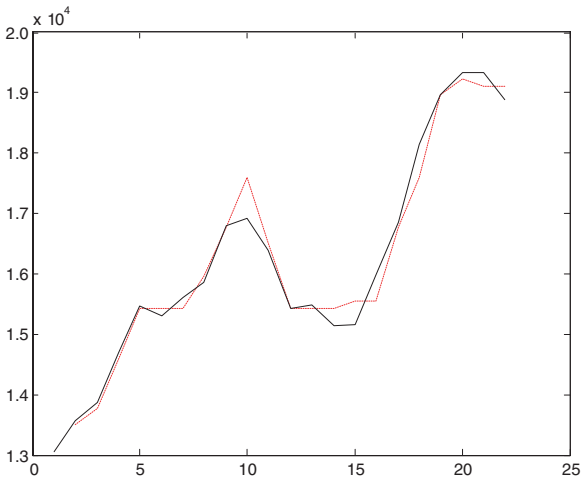

Fig. 2. The graph of actual values and forecasts based on our proposed method.

end points as 200 and 500, respectively to avoid getting very small or very large intervals in fuzzy time series analysis.

The optimal length of the intervals obtained from our proposed method is 273.2346 and the corresponding MSE value is 66661. As seen from Table 2 the results are superior to those obtained from [4].

In Fig. 1 the graph of the forecasts for the length of the interval 300 which gave the MSE value obtained by Huarng [4] and the actual values is presented. The graph of the forecasts obtained from our proposed method and the actual values are given in Fig. 2. In the graphs, solid lines represent the forecasts and dashed lines denote the actual values.

5. Discussion

In this paper, we use a single variable constrained optimization to obtain an effective length of interval for Chen's model [2]. In the optimization process, we used a MATLAB function called "fminbnd" which is based on golden section search and parabolic interpolation. We constrained the length of the interval not to get a very small or a very large length in fuzzy time series analysis so the left and the right end points are taken as 200 and 500. We applied the proposed method and methods in [4] average and distribution based length to determine the length of interval in Chen's method [2] to analyze the data of the enrollments of the University of Alabama for comparison. Huarng [4] found that the optimal length as 300 for the average based approach and MSE as 78792. Here, we investigated that optimizing the length of the interval over between 200 and 500 decreased the value of MSE. Due to our proposed method the optimal length of the interval is 273.2346 and the corresponding MSE value is 66661. Hence, it is clear that the implementation of the constrained optimization increases the forecasting performance.

In the optimization process, the left and the right points are parameters which can affect the results. We explained their effects and why we took their values as 200 and 500 in the previous section. These parameters can be changed for obtaining different results by taking into account the explanations in Section 4.

In this paper, a single variable constrained optimization is used to optimize the length of interval for obtaining more accurate forecasts. Another optimization technique can also be used and different results can be obtained. In addition, we employed this optimization method in Chen's model [2]. This optimization method, however, can be used for other fuzzy time series approaches in determining the length of interval. For future studies, these can be performed to obtain more effective lengths of interval and more accurate forecasts.

References

[1] R.P. Brent, *Algorithms for Minimization without Derivatives*, Prentice-Hall, Englewood Cliffs, New Jersey, 1973.
[2] S.M. Chen, Forecasting enrollments based on fuzzy time-series, *Fuzzy Sets and Systems* **81** (1996), 311–319.

[3] G.E. Forsythe, M.A. Malcolm and C.B. Moler, *Computer Methods for Mathematical Computations*, Prentice Hall, New Jersey, 1976.

[4] K. Huarng, Effective length of intervals to improve forecasting in fuzzy time-series, *Fuzzy Sets and Systems* **123** (2001), 387–394.

[5] K. Huarng, Ratio-based lengths of intervals to improve fuzzy time series forecasting, *IEEE Transactions on Systems, Man., and Cybernetics-Part B: Cybernetics* **36** (2006), 328–340.

[6] K. Huarng and H. Yu, A dynamic approach to adjusting lengths of intervals in fuzzy time series forecasting, *Intell Data Anal* **8**(1) (2004), 3–27.

[7] K. Huarng and H. Yu, A type 2 fuzzy time series model for stock index forecasting, *Physica A* **353** (2005), 445–462.

[8] K. Huarng and H. Yu, The application of neural networks to forecast fuzzy time series, *Physica A* **363** (2006), 481–491.

[9] J. Kiefer, Sequential minimax search for a maximum, *Proceedings of the American Mathematical Society* **4** (1953), 502–506, MR0055639. doi:10.2307/2032161.

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

[11] Q. Song and B.S. Chissom, Forecasting enrollments with fuzzy time series – Part I, *Fuzzy Sets and Systems* **54** (1993), 1–10.

[12] Q. Song and B.S. Chissom, Forecasting enrollments with fuzzy time series – Part II, *Fuzzy Sets and Systems* **62** (1994), 1–8.