The forecasting of Istanbul stock exchange by using a hybrid fuzzy time series approach

Ozge Cagcag Yolcu*
University of Ondokuz Mayis, Faculty of Arts and Science, Department of Statistics
55139 Samsun, Turkey
E-mail: ozgecagcag@yahoo.com.tr
*Corresponding author

Abstract

Nowadays, forecasting and techniques used to obtain the forecasts are very important. The term of forecast means to make an inference (predict) about the future on the basis of existing information. Especially, forecasting of stock market data are frequently used in time series analysis literature. Moreover, fuzzy time series forecasting methods have been widely used in the analysis of stock market data in recent years. In this study, Istanbul Stock Exchange time series data from four different years were forecasted with a hybrid fuzzy time series approach and evaluated with the results of the other different methods.

Keywords: Fuzzy time series, forecasting, stock exchange, membership degree, artificial neural networks.

1. Introduction

Making an inference about the data or information is one of the major problems encountered in many areas. In addition, the need for forecasting methods to make future plans and to develop strategies has increased in recent years. Forecasting the problems encountered in monetary and capital markets are important research areas. Especially, forecasting exchange rate and stock market may provide a great advantage for investors. Because stock market index is influenced by political, economic, environmental and commercial factors and the index value is dynamic, complex and non-linear, forecasting becomes a difficult and complex process. Time-series analysis methods are often used in the analysis of such data. The main purpose in time series analysis is to introduce the relationships between lagged time series which are composed of past values of related time series and different time series.

Nowadays, the methods proposed for time series forecasting problems in the literature can be evaluated under two headings as stochastic and non-stochastic. Probabilistic methods have several assumptions. The provision of these assumptions is quite difficult for time series encountered in real life. Especially in recent years, non-stochastic
methods not containing these assumptions such as fuzzy time series methods and
artificial neural network methods have been successfully applied for forecasting
problems.

Fuzzy time series analysis methods that attracted the attention of many researchers in
recent years have been widely used for this purpose. Fuzzy time series which was first
proposed by Song and Chissom in 1993a is based on fuzzy set theory proposed by
Zadeh in 1965. Fuzzy time series methods consist of three steps as fuzzification of
observations, identification of fuzzy relations and defuzzification. In fuzzification step
(the first step), universal set fragmentation is often used. Since the length of intervals
used in this method plays an effective role on forecasting performance, there are many
studies related to the determination of interval length in the literature.

In fuzzification step (the first step), universe of discourse is often used. Since the length
of intervals used in this approach plays an effective role on forecasting performance,
there are many studies related to the determination of interval length in the literature. In
most of these studies, interval lengths are determined constantly but Song and Chissom
(1993a,b, 1994) and Chen(1996, 2002) determined equal interval lengths arbitrarily
whereas Huarng (2001) used average and distribution based. Egrioglu et al. (2010,
2011) used optimization based methods. In addition, Huarng and Yu (2006a) and Yolcu
et al. (2009) suggested based on ratio. Kuo et al. (2009, 2010), Davari, et al. (2009),
Park et al(2010) and Hsu et al. (2010) used particle swarm optimization (PSO) whereas
Chen and Chung (2006), Lee et al. (2007, 2008) used genetic algorithms in the
determining of dynamic intervals. In addition, Cheng et al. (2008), Li et al. (2008),
Aladag et al. (2012), Alpaslan et al. (2012a,b), Egrioglu (2012) and Egrioglu et al.(2013)
used fuzzy clustering techniques in fuzzification step.

Identification of fuzzy relation stage which is second step of fuzzy time series methods
is also really effective on the forecasting performance. Song and Chissom (1993a,b,
1994) determined the fuzzy relations by complex matrix operations. Sullivan and
Woodall (1994) used transition matrices based on Markov chain instead of using fuzzy
logic relation matrix. Chen(1996) proposed a new model that involves easier operations
based on fuzzy group relation tables. Although there are many studies in which fuzzy
relations are determined with fuzzy logic relation and group relation tables, studies
determining the fuzzy relations with artificial neural networks (ANN) are also frequent
(2009a,b,c) proposed approaches using feed-forward artificial neural networks in the
identification of fuzzy relations. In addition, Aladag et al.(2012) used artificial neural
network with single multiplicative neuron model (SMNM-ANN) not containing
problem of determining the number of units in hidden layer. In all of these approaches,
when determining the fuzzy relations representing the internal relation of fuzzy time
series, only the fuzzy set having the highest membership value was considered and
membership values were ignored. This leads to loss of information and negatively
affects the performance of the method. In order to overcome this problem, Yu and
Huarng (2008, 2010) used forecasting models which consider membership values their
approach have determined membership values subjectively. Alpaslan et al.(2012a,b),
Yolcu et al.(2013) and Cagcag Yolcu (2013a) used FCM technique instead of
determining the membership values subjectively. The use of ANN in identification of
fuzzy relations has many advantages and disadvantages as well. Determination of unit number in hidden layer (architecture structure) and excessive number of parameters to be used during the analysis are the most prominent ones. Although Aladag (2013) eliminated this problem by using artificial neural network with single multiplicative neuron model (SMNM-ANN) in the determination of fuzzy relations, fuzzy relations were determined subjectively and membership values were not considered. Nevertheless, as the system output of these approaches consists of fuzzy set number or membership values, fuzzification step is necessary. This may be a factor that increases the model error. An approach not requiring defuzzification step would eliminate forecasting error that may occur in this step and improve the performance of the method.

Fuzzy time series approaches proposed in the literature consider these three steps that constitute the solution process as separate processes. Thus, model error is the sum of the errors that may occur in each step. In this regard, synchronous and simultaneous evaluation of the steps constituting the analysis process will produce a single model error and will lead to a reduction in the model error.

Cagcag Yolcu (2013b) has proposed an approach in this manner. In Cagcag Yolcu’s study, the steps constituting the analysis process of fuzzy time series methods were evaluated simultaneously. In this study, we aimed to introduce the approach proposed by Cagcag Yolcu (2013b) and to evaluate its performance in forecasting Istanbul Stock Exchange - 100 Index (IEX) data from four different years. Despite being studied in many recent studies, IEX data have been specifically examined in this study due to its importance.

The rest of the paper is organized as follows: In Section 2, Cagcag Yolcu’s hybrid method is presented in summary. In Section 3, the experimental results are presented. Finally, in the last section, obtained results are discussed.

2. Cagcag Yolcu’s hybrid method

The fuzzy time series was firstly introduced in Song and Chissom (1993a). There are some kind of fuzzy time series definitions which plays an effective role on analysis of fuzzy time series. In this study two definition of fuzzy time series were used in the analysis process.

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

**Definition 2** Suppose \( F(t) \) is caused by \( F(t − 1) \) only, i.e., \( F(t − 1) \rightarrow F(t) \). Then this relation can be expressed as \( F(t) = F(t − 1) \circ R(t, t − 1) \) where \( R(t, t − 1) \) is the fuzzy relationship between \( F(t − 1) \) and \( F(t) \), and \( F(t) = F(t − 1) \circ R(t, t − 1) \) is called the first order model of \( F(t) \). " \( \circ \) " represents max-min composition of fuzzy sets.
Cagcag Yolcu (2013b) aimed to minimize the error. For this purpose, simultaneous evaluation of the steps constituting the analysis process has showed up a single model error and led to a reduction in the model error. In Cagcag Yolcu’s method, cluster centers and parameters (weight and sides) of SMNM-ANN used in the determination of fuzzy relation were obtained in a single optimization process using PSO which is a population based heuristic algorithm, was firstly proposed by Kennedy and Eberhart (1995). Also by taking advantage of obtained cluster centers, membership values constituting the input of artificial neural network were obtained by equation (1) used in fuzzy C-means method.

\[
\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left( \frac{d(X(t), v_i)}{d(X(t), v_k)} \right)^{2(\beta-1)}} \quad i = 1,2,\cdots,c ; \ t = 1,2,\cdots,T
\]  

By means of the formula given in Equation 1, membership values representing the degree of each observation belong to the clusters were calculated. The structure of SMNANN whose inputs and target values consist of membership values and real values of time series is shown in Figure 1.

[Figure 1. The structure of SMNM-ANN]

In Figure 1, \( \mu_{t_k}(X(t-1)) \) is the membership value representing the degree of observation belonging to \( k \)-th fuzzy sets at \( (t-1) \) and constitutes the inputs of network. \( \Omega \) function comprises multiplication of the weighted inputs and is obtained by equation (2). In addition, \( f \) denotes activation function whereas \( \hat{X}(t) \) represents inputs of the model. Output of the model is calculated as in equation (3).

\[
\Omega(\mu, w, b) = \text{net} = \prod_{i=1}^{c} \left[ w_i \times \mu_{t_k}(X(t-1)) + b_i \right] 
\]

\[
\hat{X}(t) = f(\text{net}) = \frac{1}{1 + \exp(-\text{net})}
\]
The structure of a particle that belongs to PSO which is used in the method is given in Figure 2.

Where, \( v_i, i = 1, 2, \cdots, c \) are centers of fuzzy clusters, \( w_i, i = 1, 2, \cdots, c \) weights and biases of SMNM-ANN, respectively.

**Algorithm** The Cagcag Yolcu’s method.

**Step 1** The number of fuzzy sets \( c \) is determined. The number sets should be \( 2 < c < T \) range where \( T \) is the number of observation.

**Step 2** Parameters of PSO algorithm \( (c_1, c_2, w, pn, \text{max} \, t, vm_1, vm_2) \) to be used in fuzzification and the identification of fuzzy relation are determined, where \( c_1 \) is cognitive coefficient, \( c_2 \) is social coefficient, \( w \) is inertia parameter and \( pn \) the number of particle. \( \text{max} \, t \) is the maximum number of iterations, \( t \) effective number of iterations, and \( vm_1 \) and \( vm_2 \) are the velocities of each particle for centers of fuzzy sets and weights and biases of SMNM-ANN, respectively.

**Step 3** Initial positions of the variables to be optimized by PSO are randomly generated. Positions of each \( k^{th} \) \( (k = 1, 2, \cdots, pn) \) particles’ positions and velocities are randomly determined and kept in vectors \( X_{1k}, X_{2k}, V_{1k} \) and \( V_{2k} \) given as follows:

\[
X_{1k} = \{x_{1k,1}, x_{1k,2}, \cdots, x_{1k,c}\} \quad (4)
\]
\[
X_{2k} = \{x_{2k,1}, x_{2k,2}, \cdots, x_{2k,2c}\} \quad (5)
\]
\[
V_{1k} = \{v_{1k,1}, v_{1k,2}, \cdots, v_{1k,c}\} \quad (6)
\]
\[
V_{2k} = \{v_{2k,1}, v_{2k,2}, \cdots, v_{2k,2c}\} \quad (7)
\]

where \( x_{1k,i}, (i = 1, 2, \cdots, c) \) and \( x_{2k,i}, (i = 1, 2, \cdots, 2 \times c) \) represent \( i^{th} \) position of \( k^{th} \) particle for centers of fuzzy set and weights and biases of SMNM-ANN, respectively. \( pn \) and \( d = 3 \times c \) represents the number of particles in a swarm and positions, respectively. The initial positions and velocities of each particle in a swarm are randomly generated from uniform distribution \((\text{min}(X(t)), \text{max}(X(t))), (0,1), (-vm_1, vm_1) \) and \((-vm_2, vm_2)\), respectively.
Step 4 Evaluation function values for each particle are computed. Root mean square error (RMSE) given in below is used as evaluation function.

\[
RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (X(t) - \hat{X}(t))^2}
\]  

(8)

where \( T \) represents the number of learning sample for SMNM-ANN.

Step 5 \( P_{best_k} \), \( (k = 1,2,\cdots, pn) \) and \( G_{best} \) are determined due to evaluation function values calculated in the previous step. \( P_{best_k} \) is a vector stores the positions corresponding to the \( k^{th} \) particle’s best individual performance, and \( G_{best} \) is the best particle, which has the best evaluation function value, found so far.

\[
p_{best_k} = \{p_{k,1}, p_{k,2}, \cdots, p_{k,d}\}, \quad k = 1,2,\cdots, pn
\]  

(9)

\[
g_{best} = \{p_{g,1}, p_{g,2}, \cdots, p_{g,d}\}
\]  

(10)

Function values for each particle are computed. Root mean square error (RMSE) given in below is used as evaluation function.

Step 6 New values of positions and velocities are calculated. New values of positions and velocities for each particle are computed by using the formulas given in (11) and (12).

\[
v_{i,d}^{t+1} = [w \times v_{i,d}^{t} + c_1 \times rand_1 \times (p_{i,d} - x_{i,d}) + c_2 \times rand_2 \times (p_{g,d} - x_{i,d})]
\]  

(11)

\[
x_{i,d}^{t+1} = x_{i,d}^{t} + v_{i,d}^{t+1}
\]  

(12)

where, \( rand_1 \) and \( rand_2 \) are randomly generated from uniform distribution (0,1).

Step 3-6 are repeated the number of maximum iteration times. Finally, the elements of \( G_{best} \) are taken as the optimal solution.

3. Applications

In the implementation, four different IEX-100 data of 2009, 2010, 2011 and 2012 year data were used. In the analysis process of the data, new time series which were generated from first-order differences of time series rather than time series were used as in Yu and Huarng’s study and we used observations of last two and three month as the out-of-sample observations (test data 1 and test data 2). Therefore, we carried out eight different analyses for performance evaluation of methods.
As well as Cagcag Yolcu’s hybrid methods (CY-13), eight different methods proposed in the literature were used for the analysis in the study. These studies are:

Song ve Chissom (1993b) method: SC93b  
Chen (1996) method: C96  
Huarng (2001) Method based on average: H01\textsuperscript{1}  
Huarng (2001) Method based on distribution: H01\textsuperscript{2}  
Chen (2002) method: C02  
Huarng ve Yu (2006b) method: HY06b  
Cheng et al. (2008) method: C08  
Aladag et al. (2009) method: A09  
Yolcu et al. (2013) method: Y13

In the first implementation, IEX-2009 data which is given in Figure 3 were analyzed with different fuzzy time series methods. For the time series consisting of 250 observations, observations of the last two months (last 41 observations - test data 1) and the last three months (the last 61 observations - test data 2) were used as test set and two different analyzes were performed.

![Figure 3. The graph of IEX-2009](image)

For all methods, the performance criteria of the results of best conditions for two analyses are summarized in Table 1.

The analysis of Table 1 revealed that the CY-13 has the highest forecasting performance with RMSE and MAPE values of 644.91 and 1.02% for test data 1 and 754.86 and 1.22% for test data 2. The graph of the forecasts obtained from CY-13 which has the highest forecasting performance with actual values are given in Figure 4. and 5. When both graphs were analyzed, it can be concluded that forecasts obtained from the CY-13 method are quite compatible with the observations that belong to the test set
Table 1. Performance of the methods for IEX-2009

<table>
<thead>
<tr>
<th>Methods</th>
<th>Test Data 1</th>
<th>Test Data 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>MAPE (%)</td>
</tr>
<tr>
<td>SC93b</td>
<td>880.90</td>
<td>1.45</td>
</tr>
<tr>
<td>C96</td>
<td>880.90</td>
<td>1.45</td>
</tr>
<tr>
<td>H01(^1)</td>
<td>1035.07</td>
<td>1.46</td>
</tr>
<tr>
<td>H01(^2)</td>
<td>835.66</td>
<td>1.42</td>
</tr>
<tr>
<td>C02</td>
<td>946.50</td>
<td>1.52</td>
</tr>
<tr>
<td>HY06b</td>
<td>890.30</td>
<td>1.43</td>
</tr>
<tr>
<td>C08</td>
<td>2239.15</td>
<td>3.90</td>
</tr>
<tr>
<td>A09</td>
<td>707.87</td>
<td>1.21</td>
</tr>
<tr>
<td>Y13</td>
<td>811.10</td>
<td>1.40</td>
</tr>
<tr>
<td>CY-13</td>
<td>644.91</td>
<td>1.02</td>
</tr>
</tbody>
</table>

Figure 4. The graph of actual values and forecasts for IEX-2009 test data 1

Figure 5. The graph of actual values and forecasts for IEX-2009 test data 2
Similarly, the data of IEX data of 2010, 2011 and 2012 years were analyzed. All IEX data graphs are given in Fig. 6, 7 and 8, respectively.

**Figure 6.** The graph of IEX-2010

**Figure 7.** The graph of IEX-2011

Prediction error for the optimal results obtained from all fuzzy time series methods are presented in Table 2, 3 and 4, respectively.
Figure 8. The graph of IEX-2012

Table 2. Performance of the methods for IEX-2010

<table>
<thead>
<tr>
<th>Methods</th>
<th>RMSE</th>
<th>MAPE (%)</th>
<th>RMSE</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC93b</td>
<td>1064.53</td>
<td>1.27</td>
<td>2995.3</td>
<td>3.42</td>
</tr>
<tr>
<td>C96</td>
<td>1064.53</td>
<td>1.27</td>
<td>886.24</td>
<td>1.04</td>
</tr>
<tr>
<td>H01&lt;sup&gt;1&lt;/sup&gt;</td>
<td>1027.74</td>
<td>1.23</td>
<td>889.53</td>
<td>1.05</td>
</tr>
<tr>
<td>H01&lt;sup&gt;2&lt;/sup&gt;</td>
<td>1348.90</td>
<td>1.66</td>
<td>917.47</td>
<td>1.09</td>
</tr>
<tr>
<td>C02</td>
<td>1143.62</td>
<td>1.41</td>
<td>947.36</td>
<td>1.13</td>
</tr>
<tr>
<td>HY06b</td>
<td>1115.40</td>
<td>1.37</td>
<td>925.70</td>
<td>1.12</td>
</tr>
<tr>
<td>C08</td>
<td>2553.07</td>
<td>3.38</td>
<td>13495.55</td>
<td>19.46</td>
</tr>
<tr>
<td>A09</td>
<td>1096.77</td>
<td>1.32</td>
<td>927.72</td>
<td>1.10</td>
</tr>
<tr>
<td>Y13</td>
<td>988.07</td>
<td>1.20</td>
<td>1136.61</td>
<td>1.41</td>
</tr>
<tr>
<td>CY-13</td>
<td>907.51</td>
<td>1.04</td>
<td>883.84</td>
<td>1.05</td>
</tr>
</tbody>
</table>

Table 3. Performance of the methods for IEX-2011

<table>
<thead>
<tr>
<th>Methods</th>
<th>RMSE</th>
<th>MAPE (%)</th>
<th>RMSE</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC93b</td>
<td>925.47</td>
<td>1.35</td>
<td>1022.58</td>
<td>1.48</td>
</tr>
<tr>
<td>C96</td>
<td>925.47</td>
<td>1.35</td>
<td>969.74</td>
<td>1.44</td>
</tr>
<tr>
<td>H01&lt;sup&gt;1&lt;/sup&gt;</td>
<td>1114.43</td>
<td>1.82</td>
<td>1118.63</td>
<td>1.71</td>
</tr>
<tr>
<td>H01&lt;sup&gt;2&lt;/sup&gt;</td>
<td>985.07</td>
<td>1.53</td>
<td>1053.30</td>
<td>1.50</td>
</tr>
<tr>
<td>C02</td>
<td>1105.24</td>
<td>1.76</td>
<td>1076.38</td>
<td>1.53</td>
</tr>
<tr>
<td>HY06b</td>
<td>981.15</td>
<td>1.57</td>
<td>994.54</td>
<td>1.53</td>
</tr>
<tr>
<td>C08</td>
<td>1889.08</td>
<td>3.20</td>
<td>4541.98</td>
<td>7.24</td>
</tr>
<tr>
<td>A09</td>
<td>929.77</td>
<td>1.31</td>
<td>977.45</td>
<td>1.38</td>
</tr>
<tr>
<td>Y13</td>
<td>898.62</td>
<td>1.27</td>
<td>960.05</td>
<td>1.37</td>
</tr>
<tr>
<td>CY-13</td>
<td>891.70</td>
<td>1.34</td>
<td>894.88</td>
<td>1.32</td>
</tr>
</tbody>
</table>
Table 4. Performance of the methods for IEX-2012

<table>
<thead>
<tr>
<th>Methods</th>
<th>Test Data 1</th>
<th>Test Data 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>MAPE (%)</td>
</tr>
<tr>
<td>SC93b</td>
<td>3305.17</td>
<td>3.47</td>
</tr>
<tr>
<td>C96</td>
<td>683.14</td>
<td>0.80</td>
</tr>
<tr>
<td>H01&lt;sup&gt;1&lt;/sup&gt;</td>
<td>669.32</td>
<td>0.78</td>
</tr>
<tr>
<td>H01&lt;sup&gt;2&lt;/sup&gt;</td>
<td>693.73</td>
<td>0.75</td>
</tr>
<tr>
<td>C02</td>
<td>698.49</td>
<td>0.80</td>
</tr>
<tr>
<td>HY06b</td>
<td>694.61</td>
<td>0.80</td>
</tr>
<tr>
<td>C08</td>
<td>17911.70</td>
<td>23.51</td>
</tr>
<tr>
<td>A09</td>
<td>815.56</td>
<td>0.89</td>
</tr>
<tr>
<td>Y13</td>
<td>921.46</td>
<td>0.98</td>
</tr>
<tr>
<td>CY-13</td>
<td>603.27</td>
<td>0.67</td>
</tr>
</tbody>
</table>

When each table was reviewed, it can be clearly seen that forecasts obtained from the CY-13 are best in terms of both RMSE and MAPE criterion for both of test data. The graph of the forecasts obtained from CY-13 which has the highest forecasting performance with actual values are given in Figure 9-14.

![Figure 9. The graph of actual values and forecasts for IEX-2010 test data 1](image)
Figure 10. The graph of actual values and forecasts for IEX-2010 test data 2

Figure 11. The graph of actual values and forecasts for IEX-2011 test data 1

Figure 12. The graph of actual values and forecasts for IEX-2011 test data 2
When all graphs were analyzed, it can be concluded that forecasts obtained from CY-13 method are quite compatible with the observations that belong to the test set.

4. Conclusions and discussion

From the point of economics and personal perspective, forecasting for planning and developing new strategies are very important for both business and establishments. Although forecasts have been obtained using different models and methods, new models and methods which may give better results have also been attempted. In this study, Istanbul Stock Exchange - 100 Index data which have an important place in economic terms and which is susceptible to fluctuations in economy were examined and forecasts for four different years were obtained. Analysis of the studies on IEX data revealed that fuzzy time series forecasting methods give quite good results. But some
problems encountered in fuzzy time series analysis load the dice against these methods. In hybrid fuzzy time series method proposed by Cagcag Yolcu (2013b), this problem was eliminated and it was shown that results were obtained from Cagcag Yolcu’s method were better for IEX data, in this study.

References


25
