Research article

Stock Market Forecasting by Multivariate Higher Order Fuzzy Time Series Model

Jyoti Badge
Assistant Professor, IPER, Bhopal
E-mail: jyoti.badge@gmail.com

Abstract
Present and past behavior of stock market is required when forecasting the trends of a market. Fuzzy Time Series (FTS) has emerged as a noble approach for predicting the future values of a stock when the information is imprecise and vague. In this paper we have considered several technical indicators as input variables. These technical indicators play a very important role to predict the turning points of price and market trends. There are hundreds of technical indicators are available, but all technical indicators are not useful. So we have obtained most effective technical indicators by Principal Component Analysis (PCA). FTS model is applied on significant indicators. Experiment is conducted on taking ten years historical data sample of Indian stock market (S&P CNX NIFTY). On the basis of experimental result it is found that FTS model performed well and it gives Mean Absolute Percentage Error (MAPE) 0.43%. Copyright © www.acascipub.com, all rights reserved.

Keywords: Technical Indicators, Principal Component Analysis (PCA), Fuzzy Time Series (FTS)

1. Introduction
The traditional prediction theories were based on linear time series models, but these models were failed due to the problem arises with linguistic historical data. The linguistic nature of historical data prompted researchers to work on fuzzy time series based model. Fuzzy Time Series (FTS) model is able to forecast the trends or the patterns in
variations of time series. FTS can also be used even when some of the historical values are missing, or are highly uncertain and fluctuating market.

Song and Chissom, (1993) successfully employed the concept of fuzzy sets having linguistic variables presented by Zadeh, (1965, 1975) and the application of fuzzy logic to approximate the reasoning by Mamdani, (1977) to develop the foundation of fuzzy time series forecasting. Song and Chissom, 1993, 1994 implemented time invariant and time variant models on the historical time series data. Chen,(1996) presented a simplified time invariant method for time series forecasting by using the arithmetic operations in place of max-min composition operation used by Song and Chissom (1993). Fuzzy Time Series are not limited to linguistic values but they can also be used for the prediction of numerical values.

In this paper Fuzzy Time Series is used to further improve the forecasting accuracy. We have also tried to find the most probable universe of discourse and fuzzy set that shall have less influence on forecasting error.

1.1. Fuzzy Time Series

\( Y(t) = \ldots, 0, 1, 2, \ldots \) 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 \( f_1(t), f_2(t), \ldots \), then \( F(t) \) is called a fuzzy time-series defined on \( Y(t) \).

Assume that \( F(t) \) is a fuzzy time-series and \( F(t) = F(t - 1) \ast R(t - 1, t) \), where \( R(t - 1, t) \) is a fuzzy relation and \( \ast \) is the max-min composition operator. Then \( F(t) \) is caused by \( F(t - 1) \) and it is denoted as \( F(t - 1) \rightarrow F(t) \), where \( F(t - 1) \) and \( F(t) \) are fuzzy sets.

Song and Chissom, (1993) first proposed a forecasting model called fuzzy time series. It provided a theoretic framework to model a special dynamic process whose observations are linguistic values. Fuzzy time series has ability to cope up with situations of high uncertainty where there is a large fluctuation in consecutive values. It has been widely used for forecasting dynamic and non-linear data.

Yu et al. (2008) proposed a bivariate fuzzy time series model to forecast the TAIEX. They applied neural networks on fuzzy time series and proposed bivariate models in order to improve forecasting. Stock index and its corresponding index futures are taken as the inputs to forecast the stock index for the next day. Teoh et.al. (2008) used fuzzy time series model based on probabilistic approach and rough set rule induction for empirical research in stock markets. They proposed a hybrid fuzzy time series model with two advanced methods, Cumulative Probability Distribution Approach (CPDA) and rough set rule induction for forecasting the stock markets. Cheng et.al. (2008) proposed a new fuzzy time-series model which incorporates the adaptive expectation model into forecasting processes to modify forecasting errors. Singh, (2009) presented a computational method of forecasting based on high-order fuzzy time series. The developed computational method provides a better approach to overcome the drawback of existing high-order fuzzy time series models. The results obtained have been compared in terms of average error of forecast to show superiority of the proposed model. Wong et.al.(2009) proposed traditional time
series method and fuzzy time series method for forecasting problem. They observed that, it is more convenient to use fuzzy time series in case of limited information and when there is a urgency in decision making. Leu et.al. (2009) proposed a new fuzzy time series model termed as distance-based fuzzy time series to predict the exchange rate. The experimental results showed that the distance-based fuzzy time series outperformed random walk model and artificial neural network model in terms of mean square error. In this research paper forecasting accuracy is tested by taking different length of interval of universe of discourse.

1.2. Technical Indicators

Technical indicators help to identify price patterns and trends in financial markets. It is basically a mathematical transformation of price or volume. Technical indicators help to decide which pattern a particular instrument reflects at a given time and what the interpretation of that pattern should be. There are number of technical indicators that have been developed to gain some insight into the market behavior. Some of them are Relative Strength Index (RSI), On Balance Volume (OBV), Accumulation/Distribution Oscillator (AOD), Stochastic Oscillators, Moving average (MA) etc. It is very difficult to find the most effective indicators which are useful for analyzing the current position of market. Hence we have obtained the most effective indicator by applying Principle Component Analysis.

1.3. Principal Component Analysis (PCA)

Principal Component Analysis is a statistical technique that linearly transforms an original set of variables into a substantially smaller set of uncorrelated variables that represents most of the information in the original set of variables. Its goal is to reduce the dimensionality of the original data set. A small set of uncorrelated variables is much easier to understand and can be useful in further analyses than a larger set of correlated variables (Dunteman et.al. 1989).

In Principal Component Analysis, the variance of a matrix (Z) is explained in terms of new latent variables which are called Principal Components (PC). The first Principal Component variable is the linear combination of matrix element that has the greatest variance. The second Principal Component Variable (PCV) is the linear combination with the next greatest variance among coefficient vectors of unit length that are orthogonal to the first coefficient vector. In this manner, one can obtain k possible Principal Component Variables. The calculated Principal Component is given by

\[ t_1 = p_1^t z \quad \text{Subject to} \quad |p_1| = 1 \]  

\[ t_2 = p_2^t z \quad \text{Subject to} \quad |p_2| = 1 \]  

and \[ p_2^t p_1 = 0 \]
The Principal Component loading vectors $p$ are the eigenvectors of the covariance matrix $\sum$ of $Z$ and the corresponding eigen values $\lambda_i$ are the variances of the Principal Components. Using loading vectors the observation can be written as

$$Z = \sum_{i=1}^{k} t_i p_i' + E$$

(4)

Where $k$, is the number of Principal Components obtained and $E$ is the residual matrix (Abraham & Nair, 1998).

2. Forecasting based on Fuzzy Time Series Model

In literature review it was observed that the fuzzy time series based forecasting models were constructed on the basis of one or two factors mainly closing price and volume of the stock. The real world stock market can be affected by many factors. In this research we have taken several technical indicators of market as input parameter to the fuzzy time series model instead of closing price or volume of the stock.

We have considered Accumulation/Distribution Oscillator, Accumulation/Distribution Line, Chaikin Oscillator, Chaikin Volatility, Moving Average Convergence-Divergence (MACD), Stochastic Oscillator %K & %D, Williams %R, Williams Accumulation/Distribution Line, Negative Volume Index, Positive Volume Index, Relative Strength Index (RSI), Bollinger Band(Middle), Bollinger Band(Upper), Bollinger Band(Lower), Highest High, Lowest Low, Median Price, On-Balance Volume (OBV), Price Rate Of Change, Price And Volume Trend (PVT), Typical Price, Volume Rate Of Change, Weighted Close (Edwards et.al. 2007).

It is difficult to find relevant technical indicators. It may happen that some indicators would provide excellent information for stock $A$, but they may not give any insight information for stock $B$. Thus, we needed a tool to choose the right indicators for each stock (Ince et.al. 2004). Therefore we applied principal component analysis for finding relevant technical indicator. SPSS software is used to extract principal component.

Weighted Close (WC), William’s %R, Moving Average Convergence and Divergence (MACD), Chaikin Volatility (CV) are the four principal technical indicators extracted by Principal Component Analysis (PCA). Above four principal indicators are used for building the model. We have then constructed multivariate factors of high-order fuzzy time series model to increase the forecasting accuracy.

The $m$-factor $n^{th}$ order fuzzy logical relationship can be defined as

$$\begin{align*}
(F_1(t-1), F_2(t-1), \ldots, F_m(t-1)), \\
(F_1(t-2), F_2(t-2), \ldots, F_m(t-2)), \ldots, \\
(F_1(t-n), F_2(t-n), \ldots, F_m(t-n))
\end{align*}$$

(1)
3. Forecasting Algorithm based on Fuzzy Time Series Model

Jilani et.al. (2007) proposed multivariate high order fuzzy time series forecasting model for road accidents. In this research it has been modified and is applied on stock market for forecasting the stock market trends. Technical indicators \( WC, \%R, MACD, CV \) and closing price \( CP \) are used as input parameter. The algorithm is as follows:

**Step 1:** Find the differences between every two consecutive observations \( t \) and \( t-1 \) for \( WC, \%R, MACD, CV \) and \( CP \).

\[
egin{align*}
WCd(t-1,t) &= WC(t) - WC(t-1) \\
%Rd(t-1,t) &= %R(t) - %R(t-1) \\
MACDd(t-1,t) &= MACD(t) - MACD(t-1) \\
CVd(t-1,t) &= CV(t) - CV(t-1) \\
CPd(t-1,t) &= CP(t) - CP(t-1)
\end{align*}
\]

Where \( (t) \) and \( (t-1) \) are observations at time \( t \) & \( t-1 \) and \( d(t-1, t) \) is the difference of particular technical indicator and closing price at time \( (t) \) & \( (t-1) \).

**Step 2:** Small positive constant is added to the differences to make the universe of discourse positive, if it is negative.

\[
egin{align*}
WCd'(t-1,t) &= WCd(t-1,t) + \text{Constant} \\
%Rd'(t-1,t) &= %Rd(t-1,t) + \text{Constant} \\
MACDd'(t-1,t) &= MACDd(t-1,t) + \text{Constant} \\
CVd'(t-1,t) &= CVd(t-1,t) + \text{Constant} \\
%CPd'(t-1,t) &= CPd(t-1,t) + \text{Constant}
\end{align*}
\]
Step3: Obtain minimum ($D_{min}$) and maximum ($D_{max}$) from the differences of technical indicators and closing price

$$D_{min} = \min(WCd'(t-1,t)), \forall t, D_{max} = \max(WCd'(t-1,t)), \forall t$$

$$D_{min} = \min(%Rd'(t-1,t)), \forall t, D_{max} = \max(%Rd'(t-1,t)), \forall t$$

$$D_{min} = \min(MACDd'(t-1,t)), \forall t, D_{max} = \max(MACDd'(t-1,t)), \forall t$$

$$D_{min} = \min(CVd'(t-1,t)), \forall t, D_{max} = \max(CVd'(t-1,t)), \forall t$$

$$D_{min} = \min(CPd'(t-1,t)), \forall t, D_{max} = \max(CPd'(t-1,t)), \forall t$$

Step4: Define and partition the universe of discourse in such a way that it covers all observations of technical indicators and closing price in the training data set i.e. $U = [\text{Minimum, Maximum}]$.

Step5: Define the linguistic term $A_i(WC), A_i(%R), A_i(MACD), A_i(CV)$ and $A_i(CP)$ represented by fuzzy sets of the $WC, %R, MACD, CV$ and $CP$

$$A_1(WC) = 1/u_1 + 0.5/u_2 + \ldots + 0/u_l$$

$$A_2(WC) = 0.5/u_1 + 1/u_2 + \ldots + 0/u_l$$

$$\ldots$$

$$A_k(WC) = 0/u_1 + 0/u_2 + \ldots + 1/u_l$$

Similarly define the linguistic terms for $A_i(%R), A_i(MACD), A_i(CV)$ and $A_i(CP)$

Step6: Fuzzify $WC, %R, MACD, CV$ and $CP$.

Step 7: Establish fuzzy logical relationship among present and future state of a time series with the help of fuzzy sets. Define $m$ factor ($WC, %R, MACD, CV & CP$) $n^{th}$ order fuzzy logical relationship

$$\begin{cases}
(WC(t-n), %R(t-n), MACD(t-n), CV(t-n), CP(t-n)), \ldots, \\
(WC(t-2), %R(t-2), MACD(t-2), CV(t-2), CP(t-2)), \\
(WC(t-1), %R(t-1), MACD(t-1), CV(t-1), CP(t-1)) \rightarrow F(t)
\end{cases}$$

Step8: Defuzzified the degree of membership by Center of Area (COA) method. It is defined as

$$WCfd(t-1,t) = \frac{\sum_{k=1}^{k} \mu_{t-1,k} * m_k}{\sum_{k=1}^{k} \mu_{t-1,k}}$$
where $WCfd(t-1, t)$ is the forecasted difference between $t-1$ and $t$ of $WC$; $\mu_{t-1}$ is the forecasted degree of membership of $WC$ and $m_k$ is the corresponding midpoints of the interval, $\mu^k_{t-1}$ of $WC$. Similarly defuzzify $%R, MACD, CV$ and $CP$.

Step 9: Calculate the forecasted value of closing price at time $t$

$$fd'(t-1, t) = Av. [fd(t-1, t) of WC, %R, MACD, CV, CP] - Av. \text{ Constant} \quad (8)$$

$$\text{forecast } (t) = fd'(t-1, t) + obs_{t-1} \quad (9)$$

4. Experimental Result

The experiment was conducted on ten years data of S&P CNX NIFTY i.e from 1-Jan-1999 to 31-Dec-2009. One-third of data was used for testing and rest of the data was used for training purpose. Test data was not used in any way to create the training rules; instead it was used to predict the accuracy or error rate.

The forecasting steps of fuzzy time series model are

Step 1: Minimum and maximum values of principal technical indicators and closing price is shown in Table 1.

Table 1: Minimum and Maximum Value of Close Price and Principal Technical Indicators

<table>
<thead>
<tr>
<th>Weighted Close</th>
<th>William’s %R</th>
<th>Moving Average Convergence-Divergence</th>
<th>Chaikin Volatility</th>
<th>Close Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min. value</td>
<td>25.8</td>
<td>3.4661</td>
<td>0.8098</td>
<td>9.1645</td>
</tr>
<tr>
<td>Max. value</td>
<td>1019.1</td>
<td>145.1936</td>
<td>1155.9281</td>
<td>232.4751</td>
</tr>
</tbody>
</table>

Step 2: Partitioning of universe of discourse into different length of intervals (five, six, seven, eight, nine and ten)

Five linguistic intervals for all the input variables is shown in Table 2.

Table 2: Five Linguistic Intervals for Principal Technical Indicators and Close Price

<table>
<thead>
<tr>
<th>Linguistic Interval</th>
<th>Weighted Close</th>
<th>William’s %R</th>
<th>Moving Average Convergence-Divergence</th>
<th>Chaikin Volatility</th>
<th>Close Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U1$</td>
<td>0-203.82</td>
<td>0-29.03</td>
<td>0-23.18</td>
<td>0-46.49</td>
<td>0-250.36</td>
</tr>
<tr>
<td>$U2$</td>
<td>203.82-407.64</td>
<td>29.03-58.06</td>
<td>23.18-46.36</td>
<td>46.49-92.98</td>
<td>250.36-500.72</td>
</tr>
<tr>
<td>$U3$</td>
<td>407.64-611.46</td>
<td>58.06-87.09</td>
<td>46.36-69.54</td>
<td>92.98-139.47</td>
<td>500.72-751.08</td>
</tr>
<tr>
<td>$U4$</td>
<td>611.46-815.28</td>
<td>87.09-116.12</td>
<td>69.54-93.18</td>
<td>139.47-185.96</td>
<td>751.08-1001.44</td>
</tr>
</tbody>
</table>
Step3: Fuzzy sets for input variables are constructed using eq.5. The five linguistic terms are defined as follows: $A_1$ (Large Decrease), $A_2$ (Decrease), $A_3$ (No change), $A_4$ (Increase), $A_5$ (Large Increase). Fuzzy sets for five linguistic variables are

$$
\begin{align*}
A_1 &= \{1/u_1, 0.5/u_2, 0/u_3, 0/u_4, 0/u_5\} \\
A_2 &= \{0.5/u_1, 1/u_2,0.5/u_3, 0/u_4, 0/u_5\} \\
A_3 &= \{0/u_1, 0.5/u_2, 1/u_3,0.5/u_4, 0/u_5\} \\
A_4 &= \{0/u_1, 0/u_2, 0.5/u_3, 1/u_4, 0.5/u_5\} \\
A_5 &= \{0/u_1, 0/u_2, 0/u_3, 0.5/u_4, 1/u_5\}
\end{align*}
$$

Step4: Fuzzy logical relationship is constructed using step3. Fuzzified and Defuzzified value of $WC$, $%R$, $MACD$, $CV$ & $CP$ is shown in Table 3.

Table 3: Fuzzified and Defuzzified Value of $WC$, $%R$, $MACD$, $CV$ & $CP$ by using Five Linguistic Variables

<table>
<thead>
<tr>
<th>Date</th>
<th>Fuzzified value of Weighted close</th>
<th>Defuzzified value of Weighted close</th>
<th>Fuzzified value of William’s %R</th>
<th>Defuzzified value of William’s %R</th>
<th>Fuzzified value of Moving Average Convergence and Divergence</th>
<th>Defuzzified value of Moving Average Convergence and Divergence</th>
<th>Fuzzified value of Chaikin Volatilit</th>
<th>Defuzzified value of Chaikin Volatilit</th>
<th>Fuzzified value of Close Price</th>
<th>Defuzzified value of Close Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-Jan-2010</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>509.5</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>72.5</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>116.225</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>625.9</td>
<td></td>
</tr>
<tr>
<td>5-Jan-2010</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>509.5</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>72.5</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>116.225</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>625.9</td>
<td></td>
</tr>
<tr>
<td>6-Jan-2010</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>509.5</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>72.5</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>116.225</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>625.9</td>
<td></td>
</tr>
<tr>
<td>7-Jan-2010</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>509.5</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>72.5</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>116.225</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>625.9</td>
<td></td>
</tr>
<tr>
<td>8-Jan-2010</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>509.5</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>72.5</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>116.225</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>625.9</td>
<td></td>
</tr>
<tr>
<td>11-Jan-2010</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>509.5</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>72.5</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>116.225</td>
<td>0.5/u2+1/u3+0.5/u4</td>
<td>625.9</td>
<td></td>
</tr>
</tbody>
</table>
Similarly for six, seven, eight, nine and ten linguistic variable

**Step 5:** The actual and forecasted value of closing price is shown in Table 4.

**Table 4:** Ten Days Forecast for S&P CNX NIFTY (4-01-10 to 15-01-10)

<table>
<thead>
<tr>
<th>Date</th>
<th>Forecasted Value</th>
<th>Actual Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 interval</td>
<td>6 interval</td>
</tr>
<tr>
<td>4-Jan-10</td>
<td>5204.85</td>
<td>5205.12</td>
</tr>
<tr>
<td>5-Jan-10</td>
<td>5236</td>
<td>5278</td>
</tr>
<tr>
<td>6-Jan-10</td>
<td>5286.34</td>
<td>5323.7</td>
</tr>
<tr>
<td>7-Jan-10</td>
<td>5290.24</td>
<td>5290.71</td>
</tr>
<tr>
<td>8-Jan-10</td>
<td>5266.9</td>
<td>5238.05</td>
</tr>
<tr>
<td>11-Jan-10</td>
<td>5257.85</td>
<td>5219.7</td>
</tr>
<tr>
<td>12-Jan-10</td>
<td>5253.2</td>
<td>5219.51</td>
</tr>
<tr>
<td>13-Jan-10</td>
<td>5214.2</td>
<td>5185.35</td>
</tr>
<tr>
<td>14-Jan-10</td>
<td>5237.75</td>
<td>5204.06</td>
</tr>
<tr>
<td>15-Jan-10</td>
<td>5263.7</td>
<td>5305.7</td>
</tr>
</tbody>
</table>

To inspect the forecasting performance of the proposed model, we used mean absolute percentage error and root mean square error as performance indicators. Performance measure for different length of linguistic intervals of the universe of discourse is shown in Table 5.

**Table 5:** Ten days forecasting performance measure of Fuzzy Time Series Model

<table>
<thead>
<tr>
<th>Method Error</th>
<th>5 linguistic intervals</th>
<th>6 linguistic intervals</th>
<th>7 linguistic intervals</th>
<th>8 linguistic intervals</th>
<th>9 linguistic intervals</th>
<th>10 linguistic intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>666.738</td>
<td>1260.408</td>
<td>684.5824</td>
<td>974.487</td>
<td>658.1291</td>
<td>816.7964</td>
</tr>
</tbody>
</table>
5. Conclusion

Time series model using multiple factors has been proposed to improve the forecasting accuracy. We have used technical indicators and closing price as input parameter for modeling the stock market where as researchers have used stock prices, volumes etc as input parameters. Technical indicators play an important role in analysis of stock market and they measure different dimensions of financial performance. It was observed at the time of experiment that the combination of weighted close, %R, moving average convergence and divergence, chainkin volatility and close price produced the best results. It was also observed in the experiment that prediction can be improved by taking appropriate length of linguistic intervals of universe of discourse.

References


