Comparative analysis on Stock Prices Index Prediction

Cvetko J. Andreski[1]

Abstract – **The identification and forecasting values of market index is challenge for many researchers. The main problem in identification of stock prices and indexes is human behaviour in trade markets and uncertainty of it. In this paper we are dealing with two techniques for identification and forecasting of stock market index: time series analysis and fuzzy time series. We made comparative analysis of these techniques, especially for the forecasting feature.**

Keywords – **Identification, Forecast, Analysis.**

I. INTRODUCTION

Many companies (most of them in financial branch) offer interest to their customers. They need to invest their assets in order to cover the expenses. Insurance companies calculate interest for the insured asset. The stock market is one of the most indispensable sources for companies to raise money. The stock prices are always predicted and studied by each day individual investors, stock fund managers and financial analysts. The forecasting accuracy is main target, because more profit will be achieved if more accurate predictions are given [9].

Recently, fuzzy time series has been widely explored for forecasting data of dynamic and non-linear data. Its application varies from forecasting of university student enrollment [1, 2, 4-7] to forecasting stock index [8-14, 20-21], temperature [15] and financial forecasting [16]. This approach is an implementation fuzzy logic in the time series analysis. Many different methods and models have been proposed for stock index forecasting by researcher into fuzzy time series. Huarng[14] initiated a study on heuristic models of fuzzy time series for forecasting by using stock index data. In addition, Huarng *et.al* [21] continued to analyze a multivariate heuristic model for fuzzy time series forecasting. On the other hand, Yu [8] enhanced of weighted fuzzy time series models for Taiwan Stock Index (TAIEX) forecasting in 2005. It is assigned by the recurrent fuzzy logical relationships (FLRs) in fuzzy logical group (FLG). In establishing fuzzy relationship and forecasting are important step to consider the weight. Furthermore, Cheng *et al* [9, 20] presented the trend-weighted fuzzy time series model for forecasting in 2006 and 2009. In 2010, Chen and Wang [21] also presented the optimal weights of the antecedent fuzzy sets in each fuzzy-trend logical relationship groups. Currently, the advancement of fuzzy time series is still going on.

[1]Cvetko J. Andreski is with the Faculty of Tourism and hospitality, kej M. Tito 95, 6000 Ohrid, Macedonia, E-mail: cvetko.andreeski@uklo.edu.mk

In this paper we proceed with time series analysis using ARIMA model for analysis. Starting with unit root test and autoregressive and partial autoregressive analysis we made a model for analysis of Macedonian MBI10 index of share trade. We also made analysis on statistics of the model. At the end we made forecast of future values with analysis of validity.

Next part is basics of fuzzy time series and definitions about fuzzy time series. At the end we made identification of the index and forecast with fuzzy time series model.

II. BOX-JENKINS ARIMA METHODOLOGY FOR TIME SERIES

The ARIMA methodology represents a system-theoretic approach for modelling time series, since the specification of the empirical model does not emerge from the findings and concepts provided by economic theory. Instead, the time series is modelled by a combination of two components - the autoregressive and the moving average term i.e. the current value of a variable is represented as a function of lagged values of that variable and past random shocks that hit the variable.

In this part of the paper, we will illustrate the implementation of this methodology by using it to model the dynamics of inflation, as measured by the changes in the Macedonian index of share trade MBI10. Again, we work with daily data and the sample covers the period starting in 18.03.2009 and ending with 06.06.2009. As can be seen, in modelling the index, we do not use the whole series that is available, but we the sample is limited with 01.06.2009. Of course, we are aware that the restricted sample appears to be a sort of handicap in modelling the series, since it reduces the reliability of the estimates as well as the power of some tests. Yet, although we work with a small sample, its size exceeds the minimum number of 50 observations that is usually recommended by the literature on ARIMA modelling. For the property of stationarity we need to make two differentiations to get stationary data series. We made additional ADF test to check the stationarity and to have starting position for the model we should choose.

TABLE I

UNIT ROOT TEST

ADF Test Statistic	-8.288994	1% Critical Value*	-2.6072
		5% Critical Value	-1.9470
		10% Critical Value	-1.6191

*MacKinnon critical values for rejection of hypothesis of a unit root.
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(INDEKS,2)
Method: Least Squares
Date: 06/21/09 Time: 20:17

Sample(adjusted): 3/23/2009 6/02/2009
Included observations: 52 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(INDEKS(-1))	-1.702140	0.205349	-8.288994	0.0000
D(INDEKS(-1),2)	0.358848	0.130303	2.753938	0.0082

R-squared	0.678665	Mean dependent var	-0.957885
Adjusted R-squared	0.672238	S.D. dependent var	54.27995
S.E. of regression	31.07554	Akaike info criterion	9.748421
Sum squared resid	48284.45	Schwarz criterion	9.823469
Log likelihood	-251.4590	F-statistic	105.6008
Durbin-Watson stat	2.211257	Prob(F-statistic)	0.000000

First, although preliminary, information on the possible model can be extracted from the simple visual inspection of the series (See one of the graphs). As can be seen, the graphs clearly show that the series exhibits regular peaks, which over the time appear every fifth day. Thus, this is reflecting the seasonal component in the movement of the inflation rate. Also, it is obvious that a series of upward movements are followed by a series of downward movements and, this suggests the presence of a moving average process. Yet, notwithstanding this obvious information, obtained by the visual inspection of the graph, the most important tool in the stage of model identification is the analysis of the sample autocorrelation (*acf*) and partial autocorrelation (*pacf*) functions along with the accompanying test-statistic (See Table 2). The *acf* cannot determine the precise form of the model, but anyway, it offers a hint that the series can be modelled by a mixed process, which should include both AR and MA components. We are able to gain more information on the possible model, looking at the *pacf*,.

TABLE II
AUTOCORRELATION AND PARTIAL AUTOCORRELATION TEST

	AC	PAC	Q-Stat	Prob
1	-0.485	-0.485	13.172	0.000
2	-0.088	-0.423	13.617	0.001
3	0.019	-0.393	13.638	0.003
4	0.137	-0.186	14.750	0.005
5	-0.248	-0.476	18.479	0.002
6	0.316	-0.142	24.671	0.000
7	-0.146	-0.144	26.024	0.000
8	0.023	0.034	26.060	0.001
9	-0.112	-0.030	26.887	0.001
10	0.105	-0.080	27.640	0.002
11	-0.031	-0.018	27.706	0.004
12	0.059	-0.046	27.956	0.006
13	-0.003	0.150	27.957	0.009
14	-0.093	-0.006	28.602	0.012
15	-0.030	-0.090	28.669	0.018
16	0.132	-0.052	30.047	0.018
17	-0.010	-0.002	30.055	0.026
18	-0.048	0.051	30.250	0.035
19	0.012	0.030	30.262	0.049
20	-0.031	0.041	30.345	0.064

The estimated model along with the accompanying diagnosis tests are given in Table 3. As confirmed by the value of the tests, in all three models, the residuals behave reasonably well i.e. they are distributed normally and do not suffer from serial correlation.

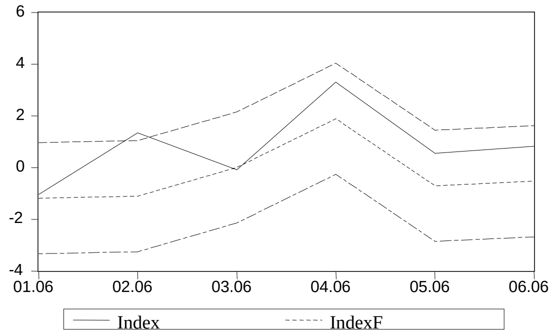
TABLE II
Autocorrelation and partial autocorrelation test

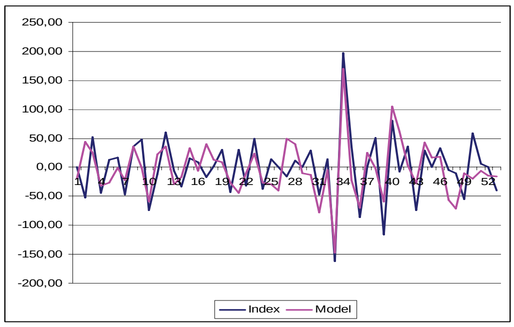
Dependent Variable: INDEKS
Method: Least Squares
Date: 06/21/09 Time: 21:31
Sample(adjusted): 3/18/2009 5/31/2009
Included observations: 53 after adjusting endpoints
Convergence achieved after 37 iterations
Backcast: 3/11/2009 3/17/2009

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-1.219381	0.065879	-18.50930	0.0000
MA(5)	0.228321	0.060001	3.805304	0.0004

R-squared	0.674157	Mean dependent var	-0.934151
Adjusted R-squared	0.667768	S.D. dependent var	53.75577
S.E. of regression	30.98462	Akaike info criterion	9.741864
Sum squared resid	48962.37	Schwarz criterion	9.816215
Log likelihood	-256.1594	F-statistic	105.5169
Durbin-Watson stat	2.548076	Prob(F-statistic)	0.000000

Graph I





We've used the model to generate out-of-sample forecasts for 01-06.06. As shown by the right panel of Graph 1 through Graph 3, from the point of view of the model's ability for forecasting future inflation, then the first model performs best. Namely, for this model, the forecasted inflation moves relatively close to the actual inflation that lies always within the ± two standard errors confidence interval.

III. DEFINITIONS FOR FUZZY TIME SERIES

There are some basic theories have been defined into fuzzy time series as follows:

Definition 1: Let, Y(t) (t=0,1,2...,) a subset of real numbers, be the universe of discourse on which fuzzy sets ,... $f_i(t)$ (i=1,2,…) are defined and *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*) (*t* = …,0, 1, 2, …). [1]

Definition 2: Suppose *F*(*t*) is caused by denoted by *F(t-1)*→ *F*(*t*), then this relationship can be represented by:

$$F(t) = F(t-1)\cdot R(t,t-1)$$

where $R(t,t-1)$ is a fuzzy relationship between *F*(*t*) and and is called the first-order model of *F*(*t*) [1].)1

Definition 3: Assumed that *F*(*t*) is a fuzzy time series, and is a first-order model of *F*(*t*). If $R(t,t-1) = R(t-1,t-2)$ for any time *t*, the *F*(*t*) is called as the time-invariant fuzzy time series. If $R(t,t-1)$ is dependent on time t, that is $R(t,t-1)$ may be different from $R(t-1,t-2)$ for any *t*, then *F*(*t*) is called the time-variant fuzzy time series.

Definition 4: Let *U* be the universe of discourse, where *U* = { u_1, u_2, …, u_b }. A fuzzy set A_i of *U* defined as $A_i = f_{A_i}(u_1)/u_1 + f_{A_i}(u_2)/u_2 + \ldots + f_{A_i}(u_b)/u_b$, where f_{A_i} is the membership function of the fuzzy set A_i; $f_{A_i}: U \to [0, 1]$. Hence, $f_{A_i}(u_1)/u_1$ is the degree of belongingness of u_a to A_i; $f_{A_i}(u_a) \to [0, 1]$ and $1 \le a \le b$ [11].

Definition 5: If there exist a fuzzy relationship), $R(t-1,t)$ such that *F*(*t*) = *F(t-1)xR(t-1,t),* where *x* represent an operation, then *F*(*t*) is said to be caused by F(t-1). [11]

Definition 6: Let $F(t-1) = A_i$ and $F(t) = A_j$. The relationship between two consecutive data (called a fuzzy logical relationship, FLR), i.e., *F*(*t*) and *F*(*t*-1), can be denoted by $A_i \to A_j$, *i*, *j* = 1, 2, …, *p* (where *p* is interval or subinterval number) is called the left-hand side (LHS), and A_j is the right-hand side (RHS) of the FLR. [11]

IV. LEFT AND RIGHT (LAR) METHOD AND DETERMINING OF WEIGHT

Our reviews on previous studies in related area show that the description of recurrence of fuzzy logical relationship in the fuzzy logical group (FLG) is still main issue. Yu[8], Cheng *et al*. [9, 20] and Chen *et al*. [21] suggested that the weight for fuzzy logical relationships should be considered to improve the forecasting accuracy. In addition, determining of weights was assigned based on domain know-how or their chronological order [8].

In this paper, we do not consider the recurrence of fuzzy relation to obtain the weight. Based on our simulation study is that the number of occurrence of the same FLR can be found more then 3 or 4 times for each FLG. To optimize this condition, we propose left and right method as follows

Let $A_3 \to A_5, A_3, A_6, A_6, A_3, A_2, A_3, A_5, A_2, A_4$ be an FLG, then the first value (A_2) that appears to left of the original value (A_3), then first value (A_4) that appears to left of the original value (A_3)and has the same value as original. Thus, this FLG can be written simply as

$A_3 \to A_3, A_2, A_4$ (1)

LAR method is generalized by using naïve assumption which the occurrence of present value A_t is very closed to the past value $A_{(t-1)}$.

B.

Computation to Derive the Weight

Suppose be an FLG which *i* = *j*, *i*, *j* ≥ 2 and *i*, *j* ϵ *Z*. By using a collection of variations of chronological numbers (*j*-1, *j*, *j*+1) in FLG, then computational weight can be assigned as follows: 11,, jjjiAAAA

W(*t*) = [w_1 w_2 w_3]

$$= \left[\frac{(j-1)}{(j-1)+j+(j+1)}\ \frac{j}{(j-1)+j+(j+1)}\ \frac{j+1}{(j-1)+j+(j+1)}\right]$$

Let *j*-1 = c_1, *j* = c_2, and *j*+1 = c_3, thus

$$= \left[\frac{c_1}{(c_1+c_2+c_3)}\ \frac{c_2}{(c_1+c_2+c_3)}\ \frac{c_3}{(c_1+c_2+c_3)}\right]$$

$$\mathbf{W}(t) = \left[\frac{c_1}{\sum_{h=1}^{3} c_h}\ \frac{c_2}{\sum_{h=1}^{3} c_h}\ \frac{c_3}{\sum_{h=1}^{3} c_h}\right]$$

which **W**(*t*) has satisfied the condition whereby the calculations will equal 1.

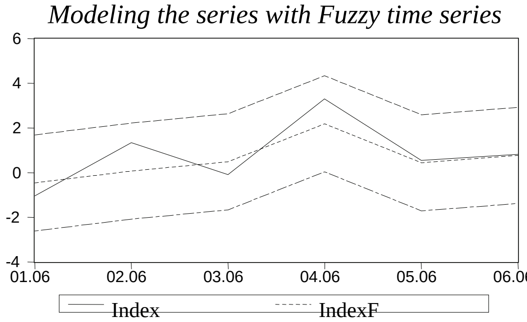
Example 1.

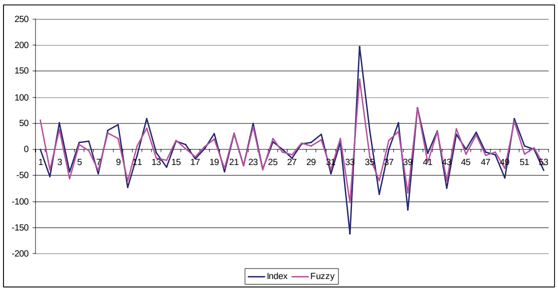
By using (1), the weight can be determined as follows:
Given $A_3 \to A_3, A_2, A_4$ and $c_1 = 3$, $c_2 = 2$, $c_3 = 4$
then
$w_1 = 3/(3 + 2 + 4)$, $w_2 = 2/(3 + 2 + 4)$, $w_3 = 4/(3 + 2 + 4)$
thus
W(A_3) = **W**(*t*) = [w_1 w_2 w_3] = [0.33 0.22 0.45].
Given $A_8 \to A_9, A_8, A_7$ and $c_1 = 9$, $c_2 = 8$, $c_3 = 7$
then
$w_1 = 9/(9 + 8 + 7)$, $w_2 = 8/(9 + 8 + 7)$, $w_3 = 7/(9 + 8 + 7)$
thus
W(A_8) = **W**(*t*) = [w_1 w_2 w_3] = [0.37 0.33 0.30].

Therefore, weight elements do satisfy the condition in both the weight matrices.

Graph II.

Modeling the series with Fuzzy time series





As for ARIMA model we first make the differencing of the time series to get the stationary time series. We made the intervals of the series and each interval is transformed to be a linguistic value time series based on previous equations. We made the fuzzy logical relationships and groups for linguistic value time series. At the end we made forecasting for post processing data based on the linguistic values of previous series. Results are shown in Graph 2.

V. CONCLUSION

We made identification and forecast on MBI10 index using time series analysis and fuzzy time series analysis. These two techniques have common background but different approach on identification and forecasting. Time series analysis are traditional technique for identification and forecasting and proven as trustable one. The fuzzy time series are still in research and innovative model. On the other hand identification and forecast of stock index is not an easy task to forecast the index prices precisely because many affected factor influence the price of stock and index. We can see that we can get better results in forecasting if we use appropriate fuzzy model.

REFERENCES

[1]. Q. Song and B. S. Chissom, “Forecasting enrollments with fuzzy time series” – Part I, *Fuzzy Sets and Systems*, **54**(1993) pp. 1-9.

[2]. Q. Song and B. S. Chissom, “Forecasting enrollments with fuzzy time series” – Part II, *Fuzzy Sets and* Systems, **64**(1994) pp. 1-8.

[3]. Q. Song and B. S. Chissom, “Fuzzy time series and its models”, *Fuzzy Sets and Systems* **54**(1993) pp. 269-277.

[4]. S. M. Chen, “Forecasting enrollments based on fuzzy time series”. *Fuzzy Sets and Systems*, **81**(1996) pp. 311-319.

[5]. S. M. Chen and C. C. Hsu, “A new method to forecast enrollments using fuzzy time series”, *International Journal of Applied Science and Engineering*, **3**(2004) pp. 234-244.

[6]. M. Sah and Y. D. Konstantin, “Forecasting enrollment model based on first-order fuzzy time series”, *Proceeding of World Academy of Science, Eng and Tech*, **1**(2005) pp. 375-378.

[7]. I. H. Kuo, S. J. Horng, T. W. Kao, T. L. Lin, C. L. Lee and Y. Pan, “An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization”, *Expert Systems with Applications*, **36**(2009) pp. 6108-6117.

[8]. H. K. Yu, “Weighted fuzzy time series models for TAIEX forecasting”, *Physica A* **349**(2005) pp. 609-624.

[9]. C. H. Cheng, T. L. Chen, and C. H. Chiang, “Trend-weighted fuzzy time series model for TAIEX forecasting”, Proceeding in *ICONIP* (2006), Part III, LNNC 4234, pp. 469-477.

[10]. K. Huarng, “Heuristic models of fuzzy time series for forecasting”, *Fuzzy Sets and Systems*, **123** (2001) pp. 369-386.

[11]. K. Huarng, H. K. Tiffany Yu, and W. S. Yu, “A multivariate heuristic model for fuzzy time series forecasting”, *IEEE Transactions on Systems, Man, and Cybernetics*, **37**(2007) pp. 263-275.

[12]. T. A Jilani and S. M. A. Burney, “A refined fuzzy time series model for stock market forecasting”, *Physica A*, **387**(2008) pp. 2857-2862.

[13]. T. H. K. Yu and K. H. Huarng, “A bivariate fuzzy time series model to forecast the TAIEX”, *Expert Systems with Application*, **34**(2008) pp. 2945-2952.

[14]. H. H Chu, T. L. Chen, C. H. Cheng, C. C. Huang, “Fuzzy dual-factor time series for stock index forecasting”, *Expert Systems with Applications*, **36**(2009) pp. 165-171.

[15]. S. M. Chen, “Temperature prediction using fuzzy time series”. *IEEE Transactions on Systems, Man, and Cybernetics*, **30**(2000) pp. 263-275.

[16]. C. H. L. Lee, A. Liu, and W. S. Chen, “Pattern Discovery of Fuzzy time series for financial prediction”, *IEEE Transactions on Knowledge and data Engineering*, **18**(2006) pp. 613-625.

[17]. J.E. Hanke, and D.W. Wichern, “Business forecasting”, *Prentice Hall*, 2009.

[18]. H. T. Nguyen and B. Wu, “Chapter; Fuzzy time series analysis and forecasting”, *Fundamental of Statistics with Fuzzy Data*, StudFuzz, **198**(2006), pp. 145-182.

[19]. S-T. Li and Y-C. Cheng, “An enhanced deterministic fuzzy time series forecasting model”, *Cybernetics and Systems*, **40**(2009), pp. 211-235.

[20]. C.H. Cheng, T.L. Chen, H.J. Teoh and C.H. Chiang, “Fuzzy time series based on adaptive expectation model for TAIEX forecasting”, *Expert Systems with Applications*, **34**(2008), pp. 1126-1132.

[21]. C.H. Chen and N.Y. Wang, “Fuzzy forecasting based on fuzzy trend logical relationship groups”, IEEE trans. Systems, Man, and Cybernatics-Part B. doi 10.1109/TSMCB.2009.2038358.

[22]. Cvetko J. Andreeski, Pandian M. Vasant, Mile J. Stankovski, and Georgi M. Dimirovski, ELMAN NN AND TIME SERIES IN FORECASTING MODELS FOR DECISION MAKING, WAC 24-26 July 2006, ISSCI 175 (ISBN 1-889335-26-6), IEEE Catalog Number: 06EX1486