VOL. 2, NO. 1, January 2012 ISSN 2222-9833
ARPN Journal of Systems and Software

http://www.scientific-journals.org

Intervals in Fuzzy Time Series Model Preliminary Investigation for Composite Index Forecasting

Lazim Abdullah [1], Chai Yoke Ling [2]

[1] Department of Mathematics, Faculty of Science and Technology, University Malaysia Terengganu, Malaysia
Email: lazim_m@umt.edu.my
[2] Department of Mathematics, Faculty of Science and Technology, University Malaysia Terengganu, Malaysia
Email: haley_yokeling@yahoo.com

ABSTRACT

Many forecasting models have been proposed to improve forecasting accuracy. Recently, Chen at al. model which incorporates three concepts of Fibonacci sequence, framework of Song and Chissom's model and weighted method of Yu's model has been proposed as a method to improve forecasting accuracy. However, the issue on lengths of intervals has not been investigated by Chen et al. despite Huarng advocated that length of intervals affects forecasting results. The issue of effective intervals in determining efficient forecasting has not been conclusively defined. Therefore this paper tests sixteen intervals using randomly chosen length of interval partitioning into the Chen et al. model to investigate this issue. A two-year weekly period of Kuala Lumpur Composite Index data sets were used to demonstrate the effectiveness of sixteen intervals. Empirical results show that that the sixteen intervals using randomly chosen partitioning method can be applied to improve fuzzy time series forecasting.

Keywords: *Forecasting; Fuzzy sets; Fuzzy time series; Composite Index, Intervals length*

1. INTRODUCTION

In the midst of market volatility and risk, it is imperative to minimize losses by embarking stock price forecasting. Realizing the importance of stock price forecasting, investors have applied number of methods hoping to search for one which could give the most accurate prediction. For many decades, researchers have developed various new approaches and methods to increase the accuracy of stock price forecasting. The most common and widely used method is time series. The most popular model for this method is the Box-Jenkins model introduced by Box and Jenkins [1]. In econometrics, the Box–Jenkins methodology applies autoregressive moving average model to find the best fit of a time series to make better forecasts. However the classical time series methods failed to deal with forecasting problems which the values of time series are linguistic terms [2]. With the advent of fuzzy theory and also challenges to face the uncertainty in stock price forecasting, efforts have been made to incorporate the fuzzy knowledge in stock price and time series. Time-series models have successfully utilized the fuzzy theory to solve several of domains forecasting problems, such as university enrolment forecasting [3] and temperature forecasting [4]. Many researches relating the model's utilization in stock price forecasting [5]-[11]. The early framework of fuzzy times series was laid in Song and Chissom model. The model comes with six steps. The basic fuzzy times series six steps were (1) define and partition universe of discourse, (2) define fuzzy sets for the observations, (3) fuzzify the observations, (4) established the fuzzy relationship, and (5) defuzzify the forecasting results [9].

The development of fuzzy theory in forecasting was further developed. Recently, Chen, et. al [12] developed a new forecasting model to give a more accurate forecasting result to investors. They proposed fuzzy time-series with Fibonacci sequence forecasting and argued that their methods has surpassed the conventional fuzzy time-series models of Chen [3], Yu [10] and Huarng [11]. They have implied the Elliott wave principle into the general fuzzy time-series. Despite the success of Chen, et al [12] in surpassing the three fuzzy time series models, the choice of partitioning intervals in the first step of their algorithm warrants further attentions. Partitioning method determines the lengths of intervals in fuzzy times series.

2. LENGTH OF INTERVALS

Reasonably define the universe of discourse and decide into how many intervals the universe will be partitioned in the first step of the algorithm is an ambiguous statement in which widely open for a new exploratory. Apart from that, it was noticed that length of intervals somehow affects the performance of fuzzy time series. Huarng and Yu [13] explore ways of determining the useful lengths of intervals in fuzzy time series. Huarng [10] also argue that different lengths of intervals lead to different forecasting results and forecasting errors. As to strengthen his argument, Huarng [10] proposed distribution- and average-based length and used sixteen lengths of intervals. Huarng [10] also cautioned that there

VOL. 2, NO. 1, January 2012 ISSN 2222-9833
ARPN Journal of Systems and Software

http://www.scientific-journals.org

will be no fluctuations in the fuzzy time series when an effective length of interval is too large. On the other hand, when the length is too small, the meaning of fuzzy time series will be diminished. Therefore, he suggested that a key point in choosing effective lengths of intervals is that they should not be too large or small. It shows that the length of interval is significant in fuzzy time series.

The size of intervals was first coined by Song and Chissom [9] and nothing has been critically examined the appropriate length of interval in the first step of the algorithm. Chen et al [12] acknowledged two common flaws exist in fuzzy time series models and one of them was partitioning the universe of discourse. The partitioning method seems very subjective and no conclusive rules inherited it. Therefore, it is essential that we need to decide the length of each interval for the partitioning based on recent development in partitioning methods. In the studies of Song and Chissom [9] for enrolment forecasting, they chose 1000 as the length of intervals without specifying any reason, Since then, 1000 has been used as the length of intervals in subsequent studies. In many previous fuzzy time series models, the lengths of the intervals were all equal and were also arbitrary. This is what we referred as the randomly chosen length of intervals. The randomly chosen length of intervals method has been commonly practiced by Song and Chissom [9] where they randomly choose an interval length for partitioning. Some years later, Huarng [11] discovered that different length of intervals used offer different forecasting results. In his studies, he proposed heuristic time-invariant fuzzy time series forecasting models to cope this issue. Chen et al [12] use seven intervals to test the efficiency of their model to Taiwan Semiconductor Manufacturing Company (TSMC) stock price data. Hence the issue of interval length is undeniable important in ensuring better forecasting results. Based on these premises, we utilize the Chen, et al [12] model by incorporating the effect of interval to composite index forecasting. We take the randomly chosen length of intervals partitioning as the interval lengths determination methods for a testing. Also, we utilise sixteen intervals as part of elucidating the behaviour of interval lengths to forecasting performances. There is no specific reason in justifying the choice of sixteen intervals. It is merely to observe higher intervals may affect forecasting accuracy. In empirical analysis, we employ an experimental datasets of Kuala Lumpur Composite Index (KLCI) to test the interval length into the model.

The rest of the paper is organized as follows. Section III introduces preliminaries that needed in exploring fuzzy time series model. Section IV proposes an experiment of composite index to the model which incorporates randomly chosen length of partitioning method. The forecasting results and performance are also presented in this section. The paper ends with a short conclusion in Section V.

3. PRELIMINARIES

The concept of fuzzy logic and fuzzy set theory were introduced to cope with the ambiguity and uncertainly of most of the real-world problems. Song and Chissom [9] introduced the concept of fuzzy time series and since then a number of variants were published by many authors. The basic concepts of fuzzy set theory and fuzzy time series are given in [9]. Some of the essentials are being reproduced to make the study self-contained. The basic concepts of fuzzy time series are defined by Definition 1 to Definition 4. Definition 5 explains Fibonacci sequence and Definition 6 provides meaning of randomly chosen length of interval partitioning method.

Let U be the universe of discourse, where $U=\{u_1,u_2,\cdots,u_n\}$. A fuzzy set $L_1,L_2,\ldots,L_k$ of U is defined by

$$
\begin{aligned}
L_1 &= a_{11}/u_1 + a_{12}/u_2 + \cdots + a_{1n}/u_n \\
L_2 &= a_{21}/u_1 + a_{22}/u_2 + \cdots + a_{2n}/u_n \\
&\vdots \\
L_k &= a_{k1}/u_1 + a_{k12}/u_2 + \cdots + a_{kn}/u_n
\end{aligned}
$$

where the value of a_{ij} indicates the grade of membership of u_j in fuzzy set L_i, where $a_{ij}\in[0,1]$, $1\le i\le k$ and $1\le j\le n$.

Definition 1. Let $Y(t)(t=\ldots,0,1,2\ldots)$, be the universe of discourse and $Y(t)\subseteq R$.Assume that $f_i(t), i=\ldots,0,1,2,\ldots$ is defined in the universe of discourse $Y(t)$ and $F(t)$ is a collection of $f_i(t),(i=\ldots,0,1,2,\ldots)$, then $F(t)$ is called a fuzzy time series of $Y(t)$, $i=1,2,\ldots$ Using fuzzy relation, we define $F(t)=F(t-1)\circ R(t,t,-1)$ where $R(t,t,-1)$ is a fuzzy relation and "$\circ$" 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.

Definition 2. Let $F(t)$ be a fuzzy time series and let $R(t,t,-1)$ be a first-order model of $F(t)$. If $R(t,t-1)=R(t-1,t-2)$ for any time t, then F(t) is called a 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 a time-variant fuzzy time series.

Definition 3. 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 n th-order fuzzy logical relationship is represented by

VOL. 2, NO. 1, January 2012 ISSN 2222-9833
ARPN Journal of Systems and Software

http://www.scientific-journals.org

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

where $F(t-1), F(t-2), \ldots, F(t-n)$ and $F(t)$ are all fuzzy sets, where $F(t-1), F(t-2), \ldots, F(t-n)$ is called the antecedent and $F(t)$ is called the consequent of the n th-order fuzzy logical relationship. When $F(t-1) = A_i$ and $F(t) = L_j$; the relationship between $F(t-1)$ and $F(t)$ (called a fuzzy logical relationship) is denoted by $L_i \rightarrow L_j$.

Definition 4. Fuzzy logical relationships with the same fuzzy set on the left-hand side can be further grouped into a fuzzy logical relationship group. Suppose there are fuzzy logical relationships such that

$L_i \rightarrow L_{j1}$,

$L_i \rightarrow L_{j2}$,

…

They can be grouped into a fuzzy logical relationship group

$L_i \rightarrow L_{j1}, L_{j2}, \ldots$

Following Chen's model [3], the same fuzzy sets can only show up once on the right-hand side of the fuzzy logical relationship group.

Definition 5. In mathematics, the Fibonacci numbers are the sequence of numbers which is defined by

$$F_n = \begin{cases} 0 \\ 1 \\ F_{n-1} + F_{n-2} \end{cases}$$

$$\left.\begin{matrix} if & n = 0 \\ if & n = 1 \\ if & n > 1 \end{matrix}\right\}_{n=1}^{\infty}$$

The Fibonacci sequence also can be expressed by the general second-order linear recurrence equation (where A and B are constants with arbitrary x_1 and x_2), which is defined as $x_n = Ax_{n-1} + Bx_{n-2}$.

In stock markets, Fibonacci numbers are commonly used in technical analysis with the knowledge of Elliott wave analysis to determine potential support, resistance, and stock price objectives. It was discovered that most patterns are composed of specific waves, matching Fibonacci numbers. These recurring patterns can be employed in forecasting stock market prices, based on the numbers. The application of the Fibonacci sequence, $F_n = F_{n-1} + F_{n-2}$, to the forecasting process.

Definition 6. The randomly chosen length of intervals method is the process where we randomly choose an interval length for partitioning. The length must not be larger than the length of the universe of discourse, U. The universe of discourse is defined as $U = [U_{\min} - U_1, U_{\max} + U_2]$, where $U_{\min}$ and $U_{\max}$ are the minimum and maximum values in U and U_1, U_2 are two real positive numbers. The universe of discourse is then divided equally by the length chosen into n number of intervals.

4. EXPERIMENT

The weekly datasets of 105 KLCI from the period of 2007/01/01 to 2008/12/29 are tested to the model. The full algorithm and the eight-step computation are purposely omitted in this paper as to accommodate with space limitation. In this section, we demonstrate how the algorithm with randomly chosen length of interval with sixteen intervals be used to forecast KLCI datasets.

From the KLCI datasets, the minimum and maximum stock price is 838.28 and 1507.04 respectively, the universe of discourse is hence to be defined as $U = [800, 1600]$. By using the fuzzy method, the low bound can be expanded by 38.28 smaller than 838.28, making Low 800, and up bound can be expanded by 92.96 larger than 1507.04, making Up 1600. As a result, the defined universe of discourse of $U = [800, 1600]$ can cover every occurring stock price in the KLCI dataset. Next, we randomly chose the length intervals to be 50. Hence, there are 16 intervals for this universe of discourse where, $u_1 = [800, 850)$, $u_2 = [850, 900)$, $u_3 = [900, 950)$, $u_4 = [950, 1000)$, $u_5 = [1000, 1050)$, $u_6 = [1050, 1100)$, $u_7 = [1100, 1150)$, $u_8 = [1150, 1200)$, $u_9 = [1200, 1250)$, $u_{10} = [1250, 1300)$, $u_{11} = [1300, 1350)$, $u_{12} = [1350, 1400)$, $u_{13} = [1400, 1450)$, $u_{14} = [1450, 1500)$, $u_{15} = [1500, 1550)$, $u_{16} = [1550, 1600]$.

The fuzzy sets, $L_1, L_2, \ldots, L_{16}$ for the universe of discourse are defined as follows:

$$L_1 = 1/u_1 + 0.5u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7 + 0/u_8 + 0/u_9 + 0/u_{10} + 0/u_{11} + 0/u_{12} + 0/u_{13} + 0/u_{14} + 0/u_{15} + 0/u_{16}$$

$$L_2 = 0.5/u_1 + 1u_2 + 0.5/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7 + 0/u_8 + 0/u_9 + 0/u_{10} + 0/u_{11} + 0/u_{12} + 0/u_{13} + 0/u_{14} + 0/u_{15} + 0/u_{16}$$

$$.... = ...$$
$$.... = ...$$
$$.... = ...$$

$$L_{16} = 0/u_1 + 0u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7 + 0/u_8 + 0/u_9 + 0/u_{10} + 0/u_{11} + 0/u_{12} + 0/u_{13} + 0/u_{14} + 0.5/u_{15} + 1/u_{16}$$

VOL. 2, NO. 1, January 2012 ISSN 2222-9833
ARPN Journal of Systems and Software

http://www.scientific-journals.org

Each observation in the KLCI datasets can be classified by the sixteen partitioned intervals. For each KLCI indices, its values are assigned to its belonging linguistic value based on the linguistic intervals. For example at time $t=1$, the KLCI index value is assigned to L_6. This is because its value, 1096.24 falls under the interval range of $[1050,1100)$.

There are six more computation steps prior to obtain final forecasting results. These steps are

i. Established fuzzy logical relationship (FLR),
ii. Generate FLR Groups and Produce Fluctuation-Weighted Matrix,
iii. Assign fluctuation weight and standardized Weight matrix for each FLR Group,
iv. Compute the Linguistic Spread Center Matrix,
v. Initial Forecasting through Defuzzification, and
vi. Fibonacci Forecasting.

Without showing the details of computational processes using these six steps, we generate the forecasting KLCI index. The observed indexes as well as the forecasted indexes were further investigated.

The trends in the forecasted values of the composite index and their actual index can be visualized in Fig.1

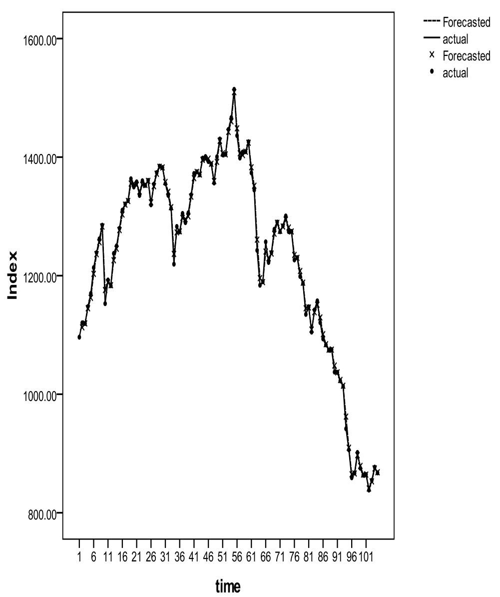

Figure: 1 Trends of actual index versus Forecasted index.

It can be seen that the actual indexes are very close to the forecasted indexes thereby it is anticipated that the model successfully fit with the tested datasets.

Performance of the method was further explored. The model's performance was measured using the mean square error (MSE). The equation for MSE is defined as

$$MSE = \frac{\sum_{t=1}^{n} | actual(t) - forecast(t) |^2}{n}$$

We also use percentage of average forecasting error rate (AFER) to see the forecasting accuracy.

$$AFER = (\frac{\sum |actual(t) - forecast(t)| / actual}{n})100\%$$

By using sixteen linguistic intervals for the partitioning method, we produced performance indicators as shown in Table I.

Table 1: Model's Performance

Error Measures	Model Fuzzy time series based on Fibonacci Sequence with Randomly Chosen Length of Intervals
RMSE	6.65
MSE	44.21
AFER	0.3995%

From the Table I, it is obvious that the model of fuzzy time series based on Fibonacci sequence with randomly chosen length of interval fit to the composite index. The average errors with less than one percent indicate the efficiency of the model with sixteen intervals. Relatively this performance is not meant to be portrayed that the sixteen intervals is an ideal number of intervals. The same data sets are neither tested with other different number of intervals nor other methods of partitioning with the same number of intervals. However the same model with seven intervals was tested to Taiwan Semiconductor Manufacturing Company stock price data [12] and bears RMSE at 1.32. Under two different platforms, it is not appropriate to postulate that the seven intervals is perform better that the sixteen intervals. Therefore this paper merely shows that the sixteen intervals of randomly chosen length of interval are feasibly at least in KLCI forecasting. Further research need to undertake to test the same model with the same datasets with other number of intervals. The future research would provide a new insight to deal with the number of intervals for the fuzzy time series forecasting model.

5. CONCLUSIONS

In this paper we chose fuzzy time series proposed by Chen, et al. [12] and applied the method of randomly chosen partitioning in defining interval length. Experimental results on the KLCI stock index demonstrate the partitioning method affect forecasting. Through analysis, we conclude that the proposed model of fuzzy time series based on Fibonacci sequence with the employment of randomly chosen length interval of partitioning have generated a very efficient predicting model specifically for KLCI. Higher number of interval may contribute to better forecasting accuracy but it sometimes creates computational complexities. This issue perhaps prompts a new venture into unravelling the computational risk in forecasting studies. The application of the method of randomly chosen length of intervals in calculating the number of intervals is seemed as a limitation and scope of this paper. Perhaps the same data sets and lengths of intervals could be tested with newly discovered length of intervals methods. A forecasting performance comparison could be discussed at this juncture. However it should be noted that performance of composite index forecasting and profits are depend with many external factors. The external factors that affecting the behaviour of stock market such as trading volumes, news and financial reports are also need to be accounted in forecasting.

REFERENCES

[1] G.E.P. Box and G. Jenkins, *Time Series Analysis*: Forecasting and Control, Holden-Day. 1976.

[2] L.A. Zadeh, Fuzzy Sets, *Information Control*, 1965, 8, 338-358.

[3] S.M. Chen,. Forecasting enrolments based on fuzzy time series, *Fuzz Set and Systems*, 1996, 81, 311-319.

[4] S.M. Chen and J.R. Hwang, Temperature prediction using fuzzy time series, *IEEE Trans. Systems, Man, and Cybernetics-part B:Cybernetics*, 30, 2000, 263-275.

[5] R.W. Faff, R.D. Brooks and Y.K. Ho, New evidence on the impact of financial everage on beta risk: a time-series approach, *North Am. J. Econ. Finance* , 2002, 13, 1–20.

[6] H.K. Yu, Weighted fuzzy time-series models for TAIEX forecasting, *Physica A* , 2004, 349, 609–624.

[7] C.H. Leon Lee, A. Liu and W.S. Chen, Pattern discovery of fuzzy time series for financial prediction, *IEEE Trans. Knowl. Data Eng*, 18, 2006, 613–625.

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

[9] Q. Song, and B.S. Chissom,. Fuzzy time-series and its models, *Fuzzy Sets and Systems* 1993,54, 269-277.

[10] K.H. Huarng, Heuristic models of fuzzy time-series for forecasting, *Fuzzy Sets Systems*. 2001, 123, 137–154.

[11] H.K. Yu, Weighted fuzzy time-series models for TAIEX forecasting, *Physica A*, 2004, 349, 609–624.

[12] T.L. Chen, C.H. Cheng and H.J. Teoh, Fuzzy time- series based on Fibonacci sequence for stock price forecasting, *Physica A*, 2007, 380, 377–390.

[13] K.H. Huarng and H.K. Yu. Ratio-based lengths of intervals to improve fuzzy time series forecasting. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Trans*. 2006, 36, 328-340.