Forecasting Unemployment based on Fuzzy Time Series with Different Degree of Confidence

Siti Musleha Ab Mutalib[1], Nazirah Ramli[2], Daud Mohamad[1]
[1]*Faculty of Computer and Mathematical Sciences,*
Universiti Teknologi MARA Pahang, Pahang, Malaysia.
[2]*Faculty of Computer and Mathematical Sciences,*
Universiti Teknologi MARA Shah Alam, Selangor, Malaysia.
musleha78@gmail.com

***Abstract*—Unemployment prediction has attracted much attention to many sectors as it can be a guide for decision making planning. In the last few years, many different forecasting techniques based on fuzzy time series (FTS) have been designed. However, most of the models used the discrete fuzzy set as a base for calculating the predicted values and the discrete fuzzy set cannot provide the forecasted range under different degree of confidence (DDoC). In this paper, FTS with trapezoidal fuzzy numbers and frequency density based partitioning approach is introduced for predicting unemployment. This model is an enhancement of the previous FTS models as it can produce the forecasted ranges under DDoC which can provide more information on the forecasted values.**

***Index Terms*—Fuzzy Forecasting; Unemployment; Trapezoidal Fuzzy Number; Statistical Distribution.**

I. INTRODUCTION

Forecasting enables decision makers to analyze data beforehand in designing policies and rules such as in the case of unemployment, environment quality and tourism. Nowadays, unemployment is increasing dramatically and becomes a critical issue faced by the worldwide government. Many factors contributed to the increasing of unemployment such as incompatibility between the backgrounds of education with the need of the labor market; skills not meet the demand of jobs and economic structure. Thus, forecasting unemployment is important to policy makers as it acts as an indicator to a better understanding of the future economic trend.

At this point, forecasting unemployment has received a great attention to many sectors. To predicts the unemployment, many methods have been proposed by the previous researcher. [1] and [2] proposed neural network method to forecast unemployment in China. Besides that, to forecast unemployment in Nigeria, [3] used univariate time series models. The aforementioned models used traditional time series to forecast unemployment. However, there are some problems in traditional time series which is the historical must follow normal distribution [4]. Furthermore, traditional time series cannot cater for data with linguistic values. Therefore, [5] proposed fuzzy time series (FTS). In order to illustrate the method, enrollment data at the University of Alabama is chosen as a sample set. Based on that FTS, [6, 7, 8, 9, 10] have made some enhancements such as alteration on the interval length. However, the improved models used discrete fuzzy set as a basis for calculating the forecasted values and the models cannot provide the forecasted ranges under different degree of confidence (DDoC).

This paper proposes a FTS technique based on trapezoidal fuzzy numbers (TFNs) approach and frequency density based partitioning method to forecast unemployment in Malaysia. The TFNs is used to represent the linguistic values of unemployment instead of using discrete fuzzy sets as appear in many previous FTS methods'. The TFNs form is selected as it is used most often to represent data compared to other shapes of fuzzy numbers and most common class of fuzzy numbers with linear membership function. In order to design linear uncertainty in scientific and applied engineering problems, TFNs have been used as it is more suitable [11]. This paper is organized as follows. In Section 2, the basic definition on FTS and TFNs are introduced. Section 3 presents the proposed fuzzy forecasting based on TFNs approach. The procedure and result for forecasting unemployment is presented in Section 4. Discussion and conclusion are presented in Sections 5 and 6 respectively.

II. PRELIMINARIES

In this section, the basic concept of FTS and TFNs are presented. FTS was first presented and defined by [5, 12, 6]. The definitions of FTS are given as follows:

Definition 1: Let $X(t)$ $(t = \ldots,0,1,2,\ldots)$ be a subset of P and $X(t)$ be the universe of discourse defined by fuzzy set $\mu_i(t)$ $(i=1,2,\ldots)$. If $A(t)$ consist of $\mu_i(t)$ $(i=1,2,\ldots)$, then $A(t)$ is called a FTS on $X(t)$ $(t = \ldots,0,1,2,\ldots)$.

Definition 2: Let $A(t)$ is a FTS. $A(t)$ is caused by $A(t) - 1$ if there exists a fuzzy relationship $P(t - 1, t)$, such that $A(t) = A(t - 1) \otimes P(t - 1, t)$ where $\otimes$ represents as fuzzy operator. The relationship can be denoted as $A(t - 1) \rightarrow A(t)$ o$\tilde{B}$r $_i \rightarrow \tilde{B}_j$ if $A(t - 1) = \tilde{B}_i$ and $A(t) = \tilde{B}_j$.

Definition 3: A trapezoidal fuzzy number (TFN) $\tilde{B}$ denoted as $\tilde{B} = (a,b,c,d)$ is defined as:

$$\mu_{\tilde{B}}(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \dfrac{d-x}{d-c}, & c \le x \le d \\ 0, & x > d \end{cases}$$

III. Proposed FTS Forecasting Model Based on TFNs and Frequency Density Based Partitioning Approach

In this section, the proposed fuzzy forecasting method is described as follows:

Step 1: Determine the universe of discourse, *UD* and appropriate length of interval.

i. $H_{\min}$ and $H_{\max}$ can be obtained from the historical data.

ii. The universe of discourse is defined as $UD = [H_{\min} - H_1, H_{\max} + H_2]$ where H_1 and H_2 are proper integer.

iii. The universe of discourse UD is divided into sub-interval based on method proposed by [13] as shown in **Error! Reference source not found.**.

Step 2: Developed new TFNs based on the intervals in Step 1. From [14], assume there are k intervals which are $u_1 = [h_1, h_2], u_2 = [h_2, h_3], ..., u_{k-1} = [h_{k-i}, h_k]$ and $u^k = [h_k, h_{k+1}]$. The linguistic terms of TFNs, $\tilde{B}_1, \tilde{B}_2, \tilde{B}_3, ..., \tilde{B}_{k-1}$ and $\tilde{B}_k$ are defined as follows:

$$\tilde{B}_1 = (h_0, h_1, h_2, h_3),$$
$$\tilde{B}_2 = (h_1, h_2, h_3, h_4),$$
$$\vdots \qquad \vdots$$
$$\tilde{B}_{k-1} = (h_{k-2}, h_{k-1}, h_k, h_{k+1}),$$
$$\tilde{B}_k = (h_{k-1}, h_k, h_{k+1}, h_{k+2}).$$

Step 3: Fuzzify the historical data. If the value of historical data is located in the range of u_j, then it belongs to fuzzy number $\tilde{B}_j$.

Step 4: Create the fuzzy logical relationships (FLR) based on Definition 2. Then, construct the FLR groups.

Step 5: Determine the forecasted values $\tilde{F}_t$ by using the heuristic rules from [15] as follows:

Rule 1: If the FLR group of $\tilde{B}_j$ is empty; $\tilde{B}_j \to \varphi, \tilde{F}_t = \tilde{B}_j$.

Rule 2: If the FLR group of $\tilde{B}_j$ is one to one; $\tilde{B}_j \to \tilde{B}_m, \tilde{F}_t = \tilde{A}_m$.

Rule 3: If the FLR group of $\tilde{B}_j$ is one to many; $\tilde{B}_j \to \tilde{B}_{m1}, \tilde{B}_j \to \tilde{B}_{m2}, ..., \tilde{B}_j \to \tilde{B}_{mp}, \tilde{F}_t = \dfrac{\tilde{B}_{m1} + \tilde{B}_{m2} + ... + \tilde{B}_{mp}}{p}$

Step 6: Defuzzify the forecasted value using centroid method as follows.

$$\tilde{F}_t = \frac{\int x \cdot \mu_{\tilde{B}}(x)dx}{\int \mu_{\tilde{B}}(x)dx}$$

Step 7: Verify the forecasting performance using mean absolute percentage error (MAPE) and root mean square (RMSE) as follows.

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(F_t - R_t)^2}{n}} \text{ and MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{F_t - R_t}{R_t}\right|.$$

where $\tilde{F}_t$ is the forecasted value, R_t is the historical data and n is the number of data.

Table 1
Division of Sub-interval

Level of Frequency Numbers of Interval	Divide into
Highest	4 sub-intervals of equal length
Second highest	3 sub-intervals of equal length
Third highest	2 sub-intervals of equal length
Fourth highest	Length remain unchanged

IV. Forecasting Unemployment

The proposed forecasting approach is applied to the data of unemployment in Malaysia for the year 1982 to 2013 as shown in Figure 1.

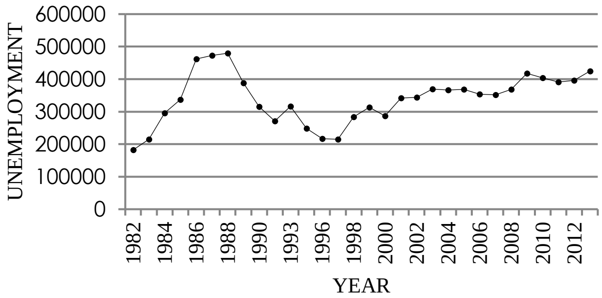

Figure 1: Time Series Data of Unemployment in Malaysia from 1982 to 2013

Step 1: The historical data R_t shows that $H_{\min}$ and $H_{\max}$ of the unemployment are 182400 and 479800 respectively. Let H_1=17400 and H_2= 200, then the universe of discourse, UD is defined as UD= [165000,480000]. The universe of discourse is partitioned into seven even and equal length intervals as in Song and Chissom (1993; 1994) which are

$U_1 = [165000, 210000]$ $U_2 = [210000, 255000]$
$U_3 = [255000, 300000]$ $U_4 = [300000, 345000]$
$U_5 = [345000, 390000]$ $U_6 = [390000, 435000]$
$U_7 = [435000, 480000]$

Table 2
Distribution in Different Interval Length

Equal Length Interval	Frequency	Category	Sub-Interval
u_1=[165000,210000]	1	6	
u_2=[210000,255000]	4	4	
u_3=[255000,300000]	4	4	
u_4=[300000,345000]	6	2	3
u_5=[345000,390000]	7	1	4
u_6=[390000,435000]	5	3	2
u_7=[435000,480000]	3	5	

Therefore, the new sub – intervals are:

u_1 =[165000,210000]	u_2 =[210000,255000]
u_3 =[255000,300000]	u_4 =[300000,315000]
u_5 =[315000,330000]	u_6 =[330000,345000]
u_7 =[345000,356250]	u_8 =[356250,367500]
u_9 =[367500,378750]	u_{10} =[378750,390000]
u_{11} =[390000,412500]	u_{12} =[412500,480000]

Step 2: The TFNs can be defined as follows:

$\tilde{B}_1$ =(120000,165000,210000,255000),
$\tilde{B}_2$ =(165000,210000,255000,300000),
$\tilde{B}_{11}$ =(390000,412500,435000,480000),
$\tilde{B}_{12}$ =(412500,435000,480000,525000).

Step 3: Fuzzify the unemployment data. For instance, R_t of unemployment in year 1982 is 182400 and located at the range of u_1=[165000,210000]. Thus, the corresponding fuzzy number of year 1982 is assigned as $\tilde{B}_1$. Table 3 shows some of the corresponding fuzzy numbers for the unemployment in year 2000 until 2013.

Step 4: After fuzzyfying the R_t, the FLR is generated and further, the FLR group is produced as shown in Table 4.

Step 5: the forecasted output of unemployment can be generated using heuristic rules proposed by [7] and [16]. Table 5 shows the forecasted unemployment in the form of TFNs for the year 2000 to 2014.

V. Discussion

Table 5 shows the forecasted output of unemployment in the form of TFNs and their defuzzified values. For each year, all possible forecasted intervals under DDoC $(0 \le \alpha \le 1)$ can be obtained. The α -cut concept [17] of the fuzzy number was applied to obtain forecasted interval.

Table 3
Corresponding Fuzzy Numbers of the Unemployment

Year	Unemployment	Fuzzy Numbers
2000	286900	$\tilde{B}_3$
2001	342400	$\tilde{B}_6$
2002	343500	$\tilde{B}_6$
2003	369800	$\tilde{B}_9$
2004	366600	$\tilde{B}_8$
2005	368100	$\tilde{B}_9$
2006	353600	$\tilde{B}_7$
2007	351400	$\tilde{B}_7$
2008	368500	$\tilde{B}_9$
2009	418000	$\tilde{B}_{12}$
2010	404400	$\tilde{B}_{11}$
2011	391400	$\tilde{B}_{11}$
2012	396300	$\tilde{B}_{11}$
2013	424600	$\tilde{B}_{12}$

Table 4
Fuzzy Logical Relationship Group

Group	Fuzzy Logical Relationship Group
1	$\tilde{B}_1 \to \tilde{B}_2$
2	$\tilde{B}_2 \to \tilde{B}_2, \tilde{B}_2 \to \tilde{B}_3$
3	$\tilde{B}_3 \to \tilde{B}_4, \tilde{B}_3 \to \tilde{B}_5, \tilde{B}_3 \to \tilde{B}_6$
4	$\tilde{B}_4 \to \tilde{B}_3$
5	$\tilde{B}_5 \to \tilde{B}_2, \tilde{B}_5 \to \tilde{B}_3$
6	$\tilde{B}_6 \to \tilde{B}_6, \tilde{B}_6 \to \tilde{B}_9, \tilde{B}_6 \to \tilde{B}_{13}$
7	$\tilde{B}_7 \to \tilde{B}_7, \tilde{B}_7 \to \tilde{B}_9$
8	$\tilde{B}_8 \to \tilde{B}_9$
9	$\tilde{B}_9 \to \tilde{B}_7, \tilde{B}_9 \to \tilde{B}_8, \tilde{B}_9 \to \tilde{B}_{12}$
10	$\tilde{B}_{10} \to \tilde{B}_5$
11	$\tilde{B}_{11} \to \tilde{B}_{11}, \tilde{B}_{11} \to \tilde{B}_{12}$
12	$\tilde{B}_{12} \to \tilde{B}_{11}$
13	$\tilde{B}_{13} \to \tilde{B}_{10}, \tilde{B}_{13} \to \tilde{B}_{13}$

Table 5
Forecasted Unemployment in TFNs Form

Year	Forecasted Output	Defuzzified Value
2000	(210000,255000,300000,315000)	269000
2001	(290000,315000,330000,343750)	319519
2002	(361250,377500,401250,423750)	391172
2003	(361250,377500,401250,423750)	391172
2004	(355000,371250,386250,408750)	380606
2005	(356250,367500,378750,390000)	373125
2006	(355000,371250,386250,408750)	380606
2007	(343125,356250,367500,378750)	361325
2008	(343125,356250,367500,378750)	361325
2009	(355000,371250,386250,408750)	380606
2010	(378750,390000,412500,435000)	404464
2011	(384375,401250,423750,457500)	417463
2012	(384375,401250,423750,457500)	417463
2013	(384375,401250,423750,457500)	417463
2014	(378750,390000,412500,435000)	404464

Table 6 shows the forecasted interval for year 2010 for $\alpha = 0$ to 1. For, $\alpha = 0$ forecasted range is between 378750 and 435000, while for $\alpha = 1$, the forecasted range is between 390000 and 412500. The length of forecasted interval for $\alpha = 0$ is the biggest compared to other α. In contrast, the length of forecasted interval for $\alpha = 1$ is the lowest compared to other α. This shows that the higher the values of α, the forecasted output has smaller values of length interval which indicates that the forecasted range is more precise at higher confidence level. For that reason, the possible forecasted intervals under DDoC can be obtained by the decision analyst. Besides that, the estimation of the forecasted ranges can be evaluated by the forecaster while making a future prediction. The previous fuzzy time series model [6, 16, 8, 9, 10] produces forecasted values in term of single point where some information are lost. Thus, the proposed technique can provide different interval estimates with evidence.

Instead of forecasted interval, the RMSE and MAPE values of two methods are computed as a comparison and shown in Table 7. Table 7 indicates that the RMSE and MAPE values of the proposed method are smaller than [5] method. Furthermore, [5] used discrete fuzzy set. Thus, the forecasted interval under DDoC cannot obtain. This shows that the proposed method produces better performance compared to [5].

Table 6
Forecasted Interval for Year 2010 (from =0 to 1)

Degree of confidence	Interval estimate
α = 0	[378750,435000]
α = 0.1	[379875,432750]
α = 0.2	[381000,430500]
α = 0.3	[382125,428250]
α = 0.4	[383250,426000]
α = 0.5	[384375,423750]
α = 0.6	[385500,421500]
α = 0.7	[386625,419250]
α = 0.8	[387750,417000]
α = 0.9	[388875,414750]
α = 1.0	[390000,412500]

Table 7
Forecasting Performance

Index	Song and Chissom's method	Proposed method
Type of fuzzy number	Discrete	Trapezoidal
RMSE	42832	29642
MAPE (%)	0.097	0.071

VI. CONCLUSIONS

Unemployment forecast is important to many sectors as it can be a guide of the economic health. In this study, FTS forecasting model based on TFNs approach and frequency density based partitioning method is presented. The proposed forecasting model can provide various forecasted intervals under DDoC $\left(0 \leq \alpha \leq 1\right)$. The higher the values of α, the forecasted output has smaller values of length interval which indicates that the forecasted range is more precise at higher confidence level. The RMSE of the proposed method is 29642 which is smaller than the method of [5] and has better performance. As compared to the previous model, this study can provide a better decision tool for decision analyst. For future work, the forecasted performance can be compared by using different type of interval length and different type of FLR.

ACKNOWLEDGMENT

This research is supported by Ministry of Education Malaysia (MOE) and Universiti Teknologi MARA Malaysia (UiTM) under the Research Grant No. 600-RMI/RAGS 5/3 (149/2014).

REFERENCES

[1] Xu W., Zheng T., and Li Z., "A neural network based forecasting method for the unemployment rate prediction using the search engine query data," *in e-Business Engineering (ICEBE)*, pp. 9–15, 2011.

[2] Wang G. and Zheng X., "The Unemployment Rate Forecast Model Based on Neural Network," *Syst. Appl*. pp. 9–12, 2009.

[3] Nkwatoh L., "Forecasting Unemployment Rates in Nigeria Using Univariate time series Models," *Int. J. Bus. Commer*. vol. 1, no. 12, pp. 33–46, 2012.

[4] Hwang J., Chen S., and Lee C., "Handling forecasting problems using fuzzy time series," *Fuzzy Sets Syst*., vol. 100, pp. 217–228, 1998.

[5] Song Q. and Chissom B. S., "Forecasting enrollments with fuzzy time series — Part I," *Fuzzy Sets Syst*., vol. 54, pp. 1–9, 1993.

[6] Song Q. and Chissom B. S., "Forecasting enrollments with fuzzy time series — part II," *Fuzzy Sets Syst*., vol. 62, pp. 1–8, 1994.

[7] Chen S.-M., "Forecasting enrollments based on fuzzy time series," *Fuzzy sets Syst*., vol. 81, pp. 311–319, 1996.

[8] M. Lee, R. Efendi, and Z. Ismail, "Modified weighted for enrollment forecasting based on fuzzy time series," *Matematika*, vol. 25, no. 1, pp. 67–78, 2009.

[9] Huarng K., "Effective lengths of intervals to improve forecasting in fuzzy time series," *Fuzzy sets Syst*., vol. 123, pp.387–394, 2001.

[10] Li S. and Chen Y., "Natural partitioning-based forecasting model for fuzzy time series," *IEEE International Conference on Fuzzy System*, pp. 1355–1359, 2004.

[11] Bansal A., "Trapezoidal Fuzzy Numbers (a , b , c , d): Arithmetic Behavior," *Int. J. Phys. Math. Sci*. pp. 39–44, 2011.

[12] Song Q. and Chissom B. S., "Fuzzy time series and its models," *Fuzzy Sets and Systems*, vol. 54, pp. 269–277, 1993.

[13] Chen S. and Hsu C., "A new method to forecast enrollments using fuzzy time series," *Int. J. Appl. Sci. Eng*. vol. 2, pp. 234–244, 2004.

[14] Liu H.-T., "An integrated fuzzy time series forecasting system," *Expert Syst. Appl*, vol. 36, no. 6, pp. 10045–10053, Aug. 2009.

[15] Cheng C., Wang J., and Li C., "Forecasting the number of outpatient visits using a new fuzzy time series based on weighted-transitional matrix," *Expert Syst. Appl*, vol. 34, no. 4, pp. 2568–2575, May 2008.

[16] Kuo I.-H., Horng S.-J., Kao T.-W., Lin T.-L., Lee C.-L., and Pan Y., "An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization," *Expert Syst. Appl*, vol. 36, no. 3, pp. 6108–6117, Apr 2009.

[17] Liu H., "An improved fuzzy time series forecasting method using trapezoidal fuzzy numbers," *Fuzzy Optim Decis Mak*. vol. 6, no.1, pp. 63–80, 2007.