ISSN:
Print - 2277 - 0593
Online - 2315 - 7461

Journal of Natural Sciences, Engineering and Technology

A COMPARISON OF FUZZY TIME SERIES AND LEAST-SQUARE METHOD IN FORECASTING STUDENTS' ENROLMENT

[1]A.J. IKUOMOLA AND [2]A.A. ADEBIYI

[1]Department of Computer Science,
[2]Department of General Studies,
Federal University of Agriculture, Abeokuta, Ogun State, Nigeria.
Corresponding author: deronikng@yahoo.com **Tel:** +2347062843043

ABSTRACT

Enrolment forecasting, which provides information for decision making and budget planning, is important in many ways to higher education. Because of its importance, researchers have proposed many forecasting methods to improve accuracy. Different methods such as genetic algorithm, least square that are used to forecast enrolment of student do not give relatively accurate results. However, obtaining accuracy is not an easy task, as many factors have impacts on enrolment numbers. In this work, a fuzzy time series was developed for efficient enrolment forecasting. The model is made up of four steps which are definition of the universe of discourse and intervals, fuzzification of historical data, establishment of fuzzy relationships and enrolment forecast. The max-min operator was used as universe of discourse and we compared our proposed method with the existing linear regression method. The historical enrolment figures of the University of Agriculture, Abeokuta were used as a data set for testing and were implemented using Visual Basic. The forecasting result of the fuzzy time series method is compared with that of the existing least square method, the fuzzy time series method produces the smallest values of the mean square error (MSE) as compared with the least square method. The application was also used to predict students' enrolment for the next five years. The proposed method was found to obtain more accurate forecasting results than the existing method.

Keyword: *Enrolment, Forecast, Fuzzy Time Series, Least Square Method, Universe of Discourse*

INTRODUCTION

Forecasting enrolment of students is necessary for planning and infrastructural development of institutions. From past data, projections can be made about the future and plans made to provide the necessary services that would enhance outputs. Forecasting is prediction of future events based on past/present experiences. It plays a vital role in our world today as it is applied in predicting a lot of events including weather changes, car fatalities and stock price values among others. In order to achieve the above, various approaches (linear regression model, auto-regression model, Neural Network model) have been used.

Enrolment forecasting is of great importance to Institutions of Higher Learning, because it aids efficient planning and administration. Accommodation issues, transportation, provision of lecture rooms and facilities, among others, can be planned for on a long and short term basis with adequate enrolment forecasting.

The enrolment forecasting studies have been presented in many methods and models such as regression models, time series models and artificial intelligences had been applied. Most of them were done for real data or numerical data. It is very important to make reasonably accurate estimates of the future enrolment for a University because many decisions can be taken from this. However, obtaining accuracy is not an easy task, as many variables have impacts on enrolment numbers.

The need for an improved educational standard has been of great importance over the past few decades coupled with this is the desire of virtually every Nigerian youth to achieve as much formal education as possible. Unfortunately, the rate at which demand for learning is increasing is quite alarming compared with facilities available. There arises the need to put measures in place to prevent an imminent breakdown of the educational system thereby forecasting enrolments of student in the University in order to enhance financial budgeting as well as strategic and research planning and to aid efficient utilization of resources in the University.

The technique proposed for this study is based on a fuzzy time series model. The model was used to forecast the enrolments of the Federal University of Agriculture Abeokuta, Ogun State, Nigeria. The accuracy of the method was verified by comparing it with the Linear Regression model vis-à-vis actual enrolment figures. The work is organized as follows: A brief overview of related work is given in section 2. Section 3 provides a design methodology for forecasting enrolment using fuzzy time series. The experiment results and its comparison with the existing linear regression method are given in section 4. Section 5 concludes the research work.

RELATED WORKS

In the last decade, fuzzy time series has been widely used for forecasting data of dynamic and non-linear data in nature. Many contribution have been made on forecasting using fuzzy time series in areas such as enrolment, the stock index, temperature, financial forecasting etc. Damousis and Dokopoulos (2001) presented a fuzzy expert system for forecasting wind speed and power generation in wind farms. Chen and Hwang (2000) developed a model for temperature prediction using fuzzy time series. Park and Lee-Kwang (2001) brought forward a designing method for type-2 fuzzy logic systems using genetic algorithms. Jilani and Burney (2007a) applied M-factor high order fuzzy time series to forecasting road accident data.

A number of remarkable contributions have been made by different researchers in relation to enrolment forecasting. Song and Chisson (1991) in their pioneering work introduced the concept of fuzzy time series as an aspect of fuzzy set theory. To evaluate the model, the conventional linear regression model was applied and the predicted values obtained were compared with the fuzzy time series results and actual enrolments. Chen (1996) forecasted enrolments of the University of Alabama using fuzzy time series; which he further improved in Chen (2002) by applying a high order fuzzy time series.

In a research carried out by Chen and Hsu (2004) on the historical enrolments of the University of Alabama using a first order and time invariant fuzzy time series model, it was asserted that their method could get a higher forecasting rate for forecasting enrolments than the methods existing as at the time of their research. This was shown by comparing the mean square error of previous forecasts with the results produced from their model. Chen and Chung (2006) improved on the findings of Song and Chisson (1991) by combining fuzzy time series and genetic algorithms to forecast enrolments.

Jilani and Burney (2007b) asserted that by applying a k_{th} order and time variant method based on frequency density based partitioning of the historical enrolment data of the University of Alabama, a better forecasting accuracy rate can be established. To achieve this, the universe of discourse was initially divided into equal intervals and a weighted aggregation of the historical enrolments was obtained in each interval. The intervals were later sub-divided and fuzzified to get the required forecast.

The work of Stevenson and Porter (2009) was an improvement over the method used by Jilani *et al.* (2007). They used the percentage change in actual enrolments from year to year as universe of discourse, and this formed part of the basis for this research work.

DESIGN METHODOLOGY

The model methodology consists of a series of steps based on the historical enrolments of the Federal University of Agriculture Abeokuta.

The proposed fuzzy time series model for enrolment forecasting procedure can be summarized as follows:

- Definition of the universe of discourse and intervals
- Fuzzification of historical data
- Establishment of fuzzy relationships
- Enrolment forecast

Step1: Definition of the universe of discourse and intervals

D_{min} and D_{max} are defined as the minimum enrolment and the maximum enrolment of known historical data. Based on D_{min} and D_{max}, the universe of discourse U is defined as [Dmin-D1, D_{max}+D2]. D1 and D2 are two positive integers chosen appropriately to suite the historical data to be used. The universe of discourse *U* is then divided into seven intervals u_{-1}, u_2, u_3, u_4, u_5, u_6 and u_7, where each interval is defined with an upper and lower bounds on the universe of discourse. For instance if D_{min}= 3654 and D_{max}= 9568, then D1 = 54 and D2 = 32 making U= [3600, 9600]. The intervals can then be defined thus u_{-1} = [3600, x_1], u_2 = [x_1, x_2], u_3 = [x_2, x_3], u_4 = [x_3, x_4] u_5= [x_4 x_5], u_6= [x_5 x_6]and u_7 =[x_6, 9600] . x_1 x_2 x_3 x_4 x_5 x_6 are integer variables and $x_1 < x_2 < x_3 < x_4 < x_5 < x_6$.

Step 2: Fuzzification of historical data

The intervals X_1, X_2, X_3, X_4, X_5 and X_6 contain integer variables, of which values are initially randomly generated by the system. The historical enrolments are then fuzzified using the randomly generated numbers. To achieve this, the value of each X_i ($1 \le i \le 6$) in each interval is substituted into the intervals of the universe of discourse earlier defined. The historical enrolments are then classified into these intervals to yield fuzzified enrolments. The reason for fuzzifying the historical enrolments is to translate crisp values into fuzzy sets in order to get a fuzzy time series.

Step 3: Establishment of fuzzy relationships

Some fuzzy sets A_1, A_2, A_3, A_4, A_5, A_6 and A_7 are defined as linguistic values of the linguistic Variable "enrolments" shown as follows:

$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7$;
$A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7$;
$A_3 = 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + 0/u_5 + 0/u_6 + 0/u_7$;
$A_4 = 0/u_1 + 0/u_2 + 0.5/u_3 + 1/u_4 + 0.5/u_5 + 0/u_6 + 0/u_7$;
$A_5 = 0/u_1 + 0/u_2 + 0/u_3 + 0.5/u_4 + 1/u_5 + 0.5/u_6 + 0/u_7$;
$A_6 = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0.5/u_5 + 1/u_6 + 0.5/u_7$;
$A_7 = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0.5/u_6 + 1/u_7$;

The corresponding notations of these fuzzy sets are:

A_1 = "not many" A_2 ="not too many"
A_3 ="many" A_4 ="many many"
A_5 ="very many" A_6 ="too many"
A_7 ="too many many".

Then, the historical enrolments of the University can be fuzzified using the fuzzy sets defined with the corresponding interval.

Furthermore, fuzzy logical relationship groups can be established among the fuzzified enrolments with the i_{th} fuzzy logical relationship group containing fuzzy relations whose current state is A_i where $1 \le i \le 7$

Step 4: Enrolment Forecast

In order to forecast enrolments any of the following rules can be used when applicable

Rule 1:

Assume that the fuzzified enrolment of the i_{th} year is A_j and assume that there is only one fuzzy logical relationship in the fuzzy logical relationship groups in which the current state of the fuzzy logical relationship is A_j, shown as follows:

$$A_J \to A_K$$

Where A_j and A_k are fuzzy sets and the maximum membership value of occurs at interval u_k, then the forecasted enrolment of the ($i+1_{th}$) year is the midpoint m_k of the interval u_k.

Rule 2:

Assume that the fuzzified enrolment of the i_{th} year is A_j and assume that there are the following fuzzy logical relationships in the fuzzy logical relationship groups in which the current states of the fuzzy logical relationships are A_j, respectively, shown as follows:

$$A_J \rightarrow A_{K1}$$

$$A_J \rightarrow A_{K2}$$

$$\vdots$$

$$A_J \rightarrow A_{KP}$$

Where A_j, A_{K1}, A_{K2}, $\cdots$ A_{KP} are fuzzy sets and the maximum membership values of A_{K1}, A_{K2} $\cdots$ and A_{KP} occur at intervals $u_1, u_2 \cdots$ *and* u_p, respectively, and the midpoints of the interval u_1, u_2, $\cdots$, and u_p are m_1, m_2, $\cdots$, and m_p, respectively; then, the forecasted enrolment the $(i+1_{th})$ year is equal to:

$$\frac{m_1 + m_2 + \ldots + m_p}{p}$$

Rule 3:

Assume that the fuzzified enrolment of the i_{th} year is A_j and assume that there are no fuzzy logical relationship groups whose current state of the fuzzy logical relationship is A_j, where the maximum membership value of A_j occurs at interval u_j; then, the forecasted enrolment of the $(i+1_{th})$ year is the midpoint m_j of the interval u_j.

One of the objectives of this study is to compare the linear regression model for forecasting with that of fuzzy time series. The linear regression model is briefly described thus:

$$X_t = a + bt$$

$$b = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}}$$

$$a = \overline{X_t} - b\bar{t}$$

Where X = the independent variable (year corresponding to each enrolment value)

Y = The dependent variable (Actual enrolment values)

n =Number of years

a = The Y-intercept

b = The slope of the trend time

t = Time in years

$\bar{t}$ = Mean of time

X_t = Forecasted value for year *t*

$\overline{X_t}$ = Mean of the dependent variable

EXPERIMENTAL RESULTS

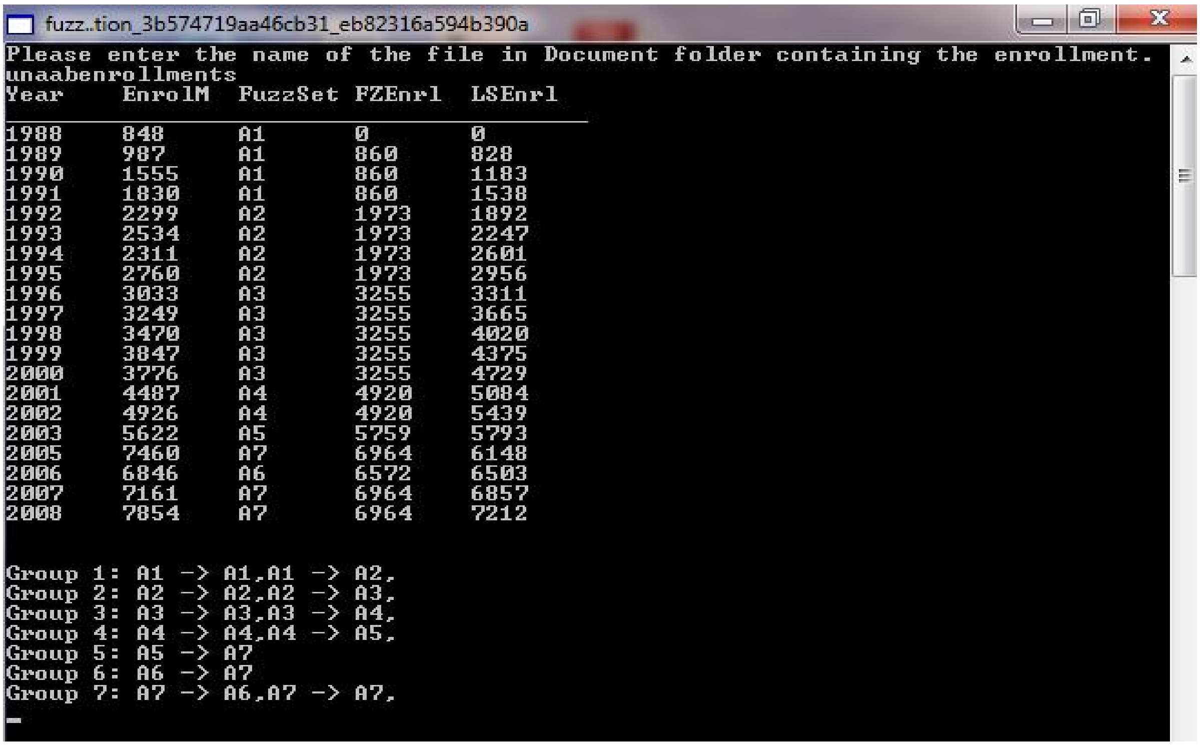
The historical enrolments of The Federal University of Agriculture Abeokuta are presented in Table 1 below:

Table 1: UNAAB Historical Enrolment Figures

Year	Enrolments
1988/89	848
1989/90	987
1990/91	1555
1991/92	1830
1992/93	2299
1993/94	2534
1994/95	2311
1995/96	2760
1996/97	3033
1997/98	3249
1998/99	3470
1999/00	3847
2000/01	3776
2001/02	4487
20002/03	4926
2003/04	5622
2005/06	7460
2006/07	6846
2007/08	7161
2008/09	7854

We have used Visual Basic. Net to implement a program on a Pentium 4 PC to forecast the enrolments of the Federal University of Agriculture Abeokuta.

The screen shot in figure (1) below shows the historical enrolments with their corresponding years, the fuzzified enrolments, the forecasted values, and the results of the least square method



fuzz..tion_3b574719aa46cb31_eb82316a594b390a

Please enter the name of the file in Document folder containing the enrollment.
unaabenrollments

Year	EnrolM	FuzzSet	FZEnrl	LSEnrl
1988	848	A1	0	0
1989	987	A1	860	828
1990	1555	A1	860	1183
1991	1830	A1	860	1538
1992	2299	A2	1973	1892
1993	2534	A2	1973	2247
1994	2311	A2	1973	2601
1995	2760	A2	1973	2956
1996	3033	A3	3255	3311
1997	3249	A3	3255	3665
1998	3470	A3	3255	4020
1999	3847	A3	3255	4375
2000	3776	A3	3255	4729
2001	4487	A4	4920	5084
2002	4926	A4	4920	5439
2003	5622	A5	5759	5793
2005	7460	A7	6964	6148
2006	6846	A6	6572	6503
2007	7161	A7	6964	6857
2008	7854	A7	6964	7212

```
Group 1: A1 -> A1,A1 -> A2,
Group 2: A2 -> A2,A2 -> A3,
Group 3: A3 -> A3,A3 -> A4,
Group 4: A4 -> A4,A4 -> A5,
Group 5: A5 -> A7
Group 6: A6 -> A7
Group 7: A7 -> A6,A7 -> A7,
```

Figure 1: Screen shot showing result generated from fuzzy time series and least square methods

The sub-section in table 2 below shows the fuzzy logical relationship groups established among the fuzzified sets. From the table, the forecasting result of the fuzzy time series method is compared with that of the existing least square method; one could see that the fuzzy time series method produces the smallest values of the AFER and MSE as compared with the least square method.

Table 2: Table showing the results obtained from the two methods being compared

YEAR	ACTUAL	FUZZTS	LSSQR	FUZZTS $(A_i - F_i)^2$	FUZZTS $\frac{\lvert A_i - F_i \rvert}{A_i}$	LSSQR $(A_i - F_i)^2$	LSSQR $\frac{\lvert A_i - F_i \rvert}{A_i}$
1988	848	0	0				
1989	987	860	828	16129	0.12867	25281	0.161094
1990	1555	860	1183	483025	0.44695	138384	0.239228
1991	1830	860	1538	940900	0.53006	85264	0.159563
1992	2299	1973	1892	106276	0.1418	165649	0.177033
1993	2534	1973	2247	314721	0.22139	82369	0.11326
1994	2311	1973	2601	114244	0.14626	84100	0.12549
1995	2760	1973	2956	619369	0.28515	38416	0.07101
1996	3033	3255	3311	49284	0.07319	77284	0.09166
1997	3249	3255	3665	36	0.00185	173056	0.12804
1998	3470	3255	4020	46225	0.06196	302500	0.1585
1999	3847	3255	4375	350464	0.15389	278784	0.13725
2000	3776	3255	4729	271441	0.13798	908209	0.25238
2001	4487	4920	5084	187489	0.0965	356409	0.13305
2002	4926	4920	5439	36	0.00122	263169	0.10414
2003	5622	5759	5793	18769	0.02437	29241	0.03042
2005	7460	6964	6148	246016	0.06649	1721344	0.175871
2006	6846	6572	6503	75076	0.04002	117649	0.050102
2007	7161	6964	6857	38809	0.02751	92416	0.042245
2008	7854	6964	7212	792100	0.11332	412164	0.081742
				MSE = 4670409	AFER = 0.13492775	MSE = 5351688	AFER = 0.1216039

Figure 2 below shows that students' enrolment in FUNAAB in the year 2006 is on the increase.

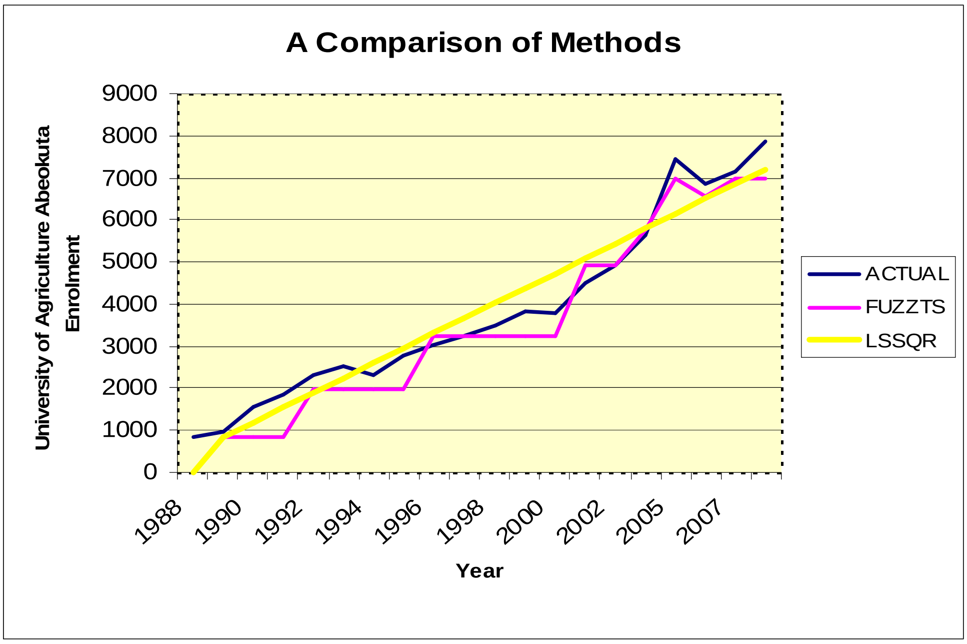


Figure 2: A comparison between the proposed method and Least Square Method

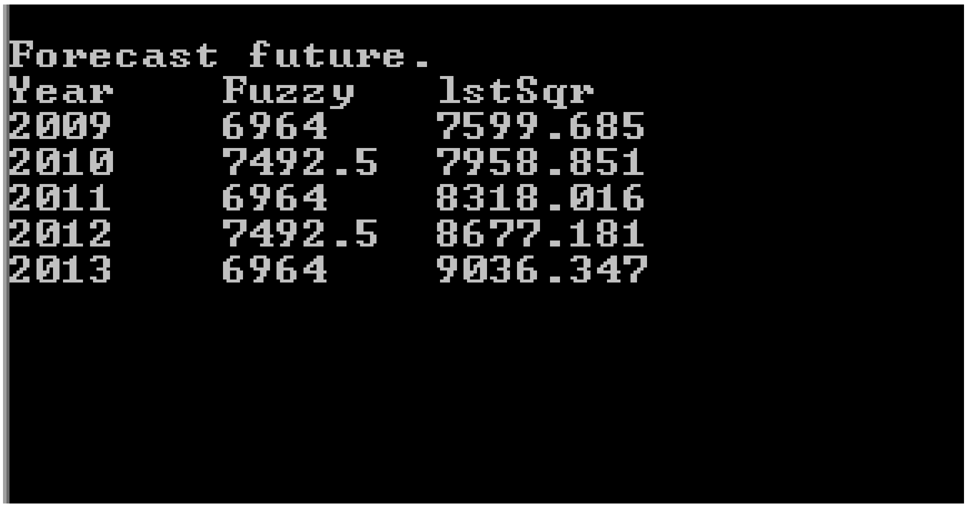
Furthermore, the chart shows that the result obtained from the fuzzy time series (FUZZTS) is approximately closer to the actual enrolment (ACTUAL) than the least square method (LSSQR).
The average forecasting error rate (AFER) and mean square error (MSE) were used to compare the forecasting results of fuzzy and least square methods:

$$\text{AFER} = \frac{\frac{|A_i - F_i|}{A_i} \times 100\%}{n}$$

$$\text{MSE} = \frac{\sum_{i=1}^{n} (A_i - F_i)^2}{n}$$

where A_i denotes the actual enrolment and F_i denotes the forecasting enrolment of year i, respectively.

The fuzzy time series and least square applications were also applied to predict students' enrolment in the Federal University of Agriculture Abeokuta for the next five years. The output generated was shown in figure 3 while figure 4 is a graphical representation of the fuzzy time series and least square method used for forecasting enrolment of students.


Forecast future.

Year	Fuzzy	lstSqr
2009	6964	7599.685
2010	7492.5	7958.851
2011	6964	8318.016
2012	7492.5	8677.181
2013	6964	9036.347

Figure 3: Output generated for forecasting for the next five years

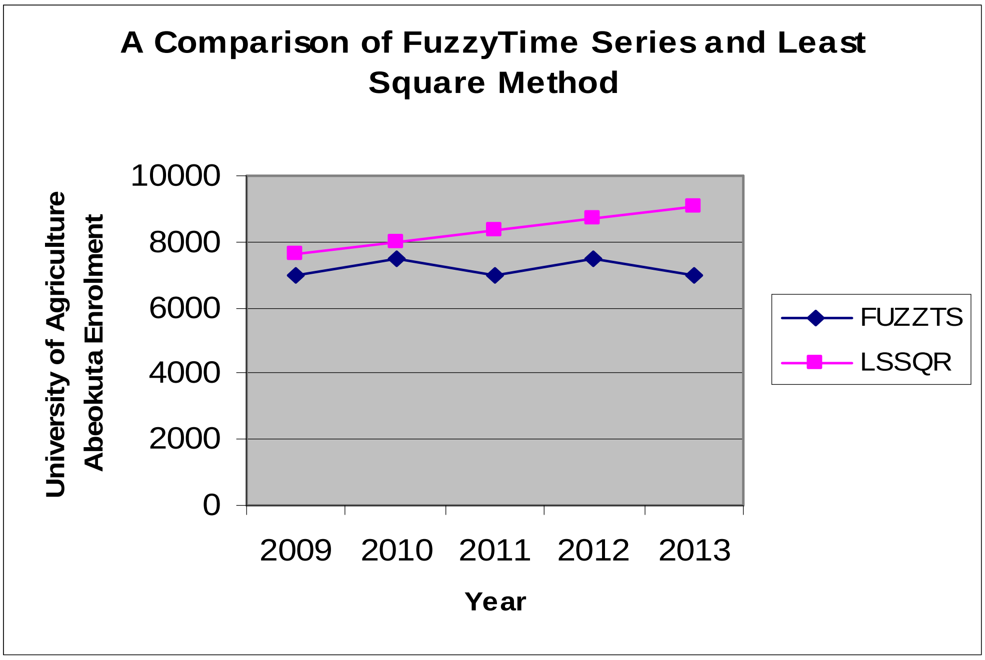


Figure 4: Forecasting Enrolment of Students for the next five years using fuzzy time series and least square method

The output generated on forecasting for the next five years as shown in figures 3 and 4 above shows a weakness of fuzzy time series in predicting values beyond two years which the least square method would do more conveniently.

CONCLUSION

Enrolment forecast is of great importance in institutions of learning because it aids proper administrative planning. A model based on fuzzy time series was used to predict the enrolment of students in the Federal University of Agriculture Abeokuta. The method uses the max-min operator as the universe of discourse which denotes the maximum and minimum enrolments respectively; this was used to form intervals for fuzzifying the enrolment figures and hence make predictions. The results were compared with the least square method of forecasting. An attempt was also made to predict the enrolments of the institution for the next five years. The results obtained shows that least square method will predict values beyond two year better than fuzzy time series but one could see that the fuzzy time series method produces the smallest values of the mean square error as compared with the least square method.

REFERENCES

Chen, S. M. 1996. Forecasting Enrolments based on Fuzzy Time Series. *Fuzzy Sets and Systems*, 81: 311-319.

Chen, S., Hsu, C. 2004. A new method to forecast enrolments using fuzzy time series, *International Journal of Applied Science and Engineering*, 2(3): 234-244.

Chen, S., Chung, N. 2006. Forecasting Enrolments of Students by Using Fuzzy Time Series and Genetic Algorithms, *Information and Management Sciences,* 17(3): 1-17.

Chen, S.M. Hwang J.R. 2000. Temperature prediction using fuzzy time series, *IEEE Transactions on Systems, man and Cybernetics-part B: Cybernetics,* 30(2): 263-275.

Chen S.M. 2002. Forecasting enrolments based on high-order fuzzy time series, *Cybernetics and Systems: an International Journal,* 33(1): 1-16.

Damousis, I.G., Dokopoulos, P. 2001. A fuzzy expert system for the forecasting of wind speed and power generation in wind farms, *proceedings of the second IEEE International conference on power industry computer applications, Sydney Australia,* 63-69.

Jilani, T.A., Burney, S.M.A. 2007a. M- factor high order fuzzy time series forecasting for road accident data, *IEEE-IFSA, 2007, World congress*. Cancun, Mexico. 18-21.

Jilani, T.A., Burney, S.M.A 2007b. Fuzzy metric approach for fuzzy time series forecasting based on frequency density based partitioning, *World Academy of Science, Engineering and Technology*, 34.

Park, S., Lee-Kwang, H. 2001. A designing method for type-2 fuzzy logic systems using genetic algorithms, *Proceedings of the joint 9th IFSA World Congress and 20th NAFIPS International Congress,* Vancouver, Canada, 5: 2567-2572.

Song, Q., Chisson, B.S. 1991. Forecasting enrolments with fuzzy time series, *Paper Presented at the Annual Meeting of the Mid-South Educational Research Association, 20th, Lexington, KY*, 12-15.

Stevenson, M., Porter, J.E. 2009. Fuzzy time series forecasting using percentage change as the universe of discourse, *World Academy of Science, Engineering and Technology*, 55.

(Manuscript received: 6th January, 2011; accepted: 15th May, 2012).