RASHI 1 (1) :54-60 (2016)

FUZZY TIME SERIES FORECASTING OF OIL SEEDS PRODUCTION IN INDIA

[1]G. S. Siva, [2]D. S. Gupta, [3]M. K. Debnath and [4]S. S. Das

[1]Research Scholar,.[2]Associate Professor,.[3]Assistant Professor, [4]Research Scholar.

Department of Agricultural Statistics,UBKV, Pundibari,West Bengal-736165

ABSTRACT

Many approaches have been designed and developed based on the concept of Fuzzy time series forecasting. These approaches have been widely applied to deal with forecasting problems in which historical data are linguistic values. In this paper we applied four fuzzy time series forecasting models namely Chen's type arithmetic model (model-1), chen's weighted average model (model-2), Rajaram and Vamitha model (Model-3) and a combined approach of Model-1 and Model-3 (Model-4) by using the historic time series data of Oil seed production in India over a period of past 26 years. It is shown that the model-3 and model-4 achieve a significant accuracy as compared to other fuzzy time series forecasting models.

Keywords: Fuzzy time series, Fuzzy relationship, Fuzzy logical relationship Groups, forecasts, linguistic value

1. Introduction :

Forecasting plays an important role in our daily life especially in making future decisions such as weather forecasting, university enrollments, production, sales and finance etc., Among them ARIMA models and Box-Jenkins model building approaches are highly rated. But classical time series methods must satisfy proper conditions to be successful and these methods can not deal with forecasting problems in which the values of time series are linguistic terms represented by fuzzy sets [1]. The motivation of applying the fuzzy time series forecasting models is to find ways of modeling the prediction of crop yield, a non real deterministic process, further the area specific crop yield forecasting for lead year may be applied to help the crop planning and agro based business planning of the area and can be used in economics and business analysis [2].

The present work deals with the application of different fuzzy time series forecasting models and its implementation for testing the accuracy in forecasting and comparison of the results of various forecasting models. The historic time series data of Oil seed production in India are used for the present study have been collected from the Ministry of Agriculture, India for the period 1988 to 2013.

2. Review of Fuzzy time series:

Forecasting the size of any phenomenon in future is important and helpful for understanding behavior of phenomenon for a long time. Many studies have interested in fuzzy time series and have been applied in various fields and proved its efficiency in forecasting as a good new method for predicting linguistic values. However, the intention here is not to provide an exhaustive study of every work published. Song and Chissom's [3] first introduced the method of fuzzy time series. Therefore, they are considered as founders of Fuzzy time series science. Next, a more detailed study is provided by Chen's work [4][5] presented in the respective papers are among the most important milestones in this particular field of research. A significant drawback of the models developed by Song and Chissom is that they are associated with unnecessary high computational overheads due to complex matrix operations in step 5 and 6. In order to reduce the computation overhead Chen proposed a simplified model including only simple arithmetic operations [6]. Rajaram et al [7] proposed a modified approach on fuzzy time series forecasting. They used nearest symmetric trapezoidal fuzzy numbers to further enhance the forecasting accuracy.

Fuzzy time series forecasting Algorithm:

The development of fuzzy time series forecasting models and there implementation can be made with the following steps:

1. Define the universe of discourse with given time series data on which fuzzy sets are to be defined.
2. Partition the universe of discourse into seven equal length intervals.
3. Define fuzzy sets (linguistic variables) on the universe of discourse.
4. Fuzzify the historical data.
5. Computing the fuzzy relationships (FLR's) and establish fuzzy relationship groups (FLRG's).
6. Defuzzify the forecasted output.

Email: mkanti1984@gmail.com

3. Forecasting oil seed production

Computational procedures

The implementation of the above algorithm have been carried by the four models: Chen's arithmetic model (model-1), refined arithmetic model (model-2), Rajaram's modified approach model (model-3) and a combined approach of Rajaram's and Chen's arithmetic model (model-4) for the forecasting of the total oil seeds is based on the 26 years (1988-89 to 2013-14) time series production data of India.

Step 1

Define the universe of discourse U as [D_{min}-D_1, D_{max}+D_2] to accommodate the time series data, where D_{min} and D_{max} are the minimum and maximum historical production respectively. From table 1, we get D_{min}=1484 and D_{max}=3275. The variables D_1 and D_2 are just two positive numbers, properly chosen by the user. If we let D_1 =4 and D_2 =25, we get U= [1480-3300].

Step 2

Partition the universes of discourse into 7 equal length intervals U_1,U_2,……..,U_7 such that

U_1 = [1480-1740],

U_2 = [1740-2000],

U_3 = [2000-2260],

U_4 = [2260-2520],

U_5 = [2520-2780],

U_6 = [2780-3040],

U_7 = [3040-3300].

Step 3

Define 7 fuzzy sets A_1, A_2, ..., A_7 having some linguistic values on the universe of discourse U. The linguistic values to these fuzzy variables are as follows:

A_1: poor production,

A_2: below average production

A_3: average production

A_4: good production

A_5: very good production

A_6: excellent production

A_7: bumper production

Step.4

In this context, fuzzification is the process of identifying associations between the historical values in the dataset and the fuzzy sets defined in the previous step. Each historical value is fuzzified according to its highest degree of membership. If the highest degree of belongingness of a certain historical time variable, say F(t-1), occurs at fuzzy set A_k, then F (t-1) is fuzzified as A_k. A complete overview of fuzzification of historical total oil seed production for four different fuzzy time series models are shown in the table 2.

To exemplify this, let us fuzzify year 1991-92 in model-1 and model-2. According to table 1, the oil seed production in 1991-92 was 1860 (000' tonnes) which lies within the boundaries of interval U_2. Since the highest membership degree of U_2 occurs at A_2, the historical time variable F (1991-92) is fuzzified as A_2.

Step 4.1: By using the following frequency distribution

Intervals	U_1	U_2	U_3	U_4	U_4	U_6	U_7
No. of Historical production data	2	4	7	6	1	3	3

Step.5

The fuzzy logical relations for all the four models have obtained from the table 3a, 3b and table 4.

In case of model-3 and model-4 after executing the step.4 the following steps are executed.

Divide the intervals Ui, i=1to 7 as follows:

V1=[1480-1610], V2=[1610-1740] with interval of 130.

V3=[1740-1805],.........V6=[1935-2000] with an interval of 65.

V7=[2000-2037],..........V13=[2222-2260] with an interval of 37 and 38 alternatively.

V14=[2260-2303],........V19=[2476-2520] with an interval of 43 and 44 alternatively.

V20=[2520-2780] with an interval of 260.

V21=[2780-2866],...........V23=[2953-3040] with an interval of 86 and 87 alternatively.

V24=[3040-3126],.........V26=[3213-3300] with an interval of 86 and 87 alternatively.

After calculating the above intervals the fuzzy trapezoidal numbers are defined for the model-3 are as follows:

A1 = [1350,1480,1610,1740],

A2 = [1480,1610,1740,1805],

A3 = [1610,1740,1805,1870],

.... A25 = [3040, 3126, 3213, 3300],

A26 = [3126, 3213, 3300, 3387].

If the value of the historical data is located in the range of vj, then it belongs to the fuzzy number Aj. All the data must be classified in to the corresponding fuzzy numbers as shown in the table 2.

Step-6

The computational procedure of fuzzy forecast of the oil seed production for each model has been carried as follows:-

Model-1

Assume the fuzzified production of F(t-1) is Aj, then forecasted output of F(t) is determined according to the following principles:

1. If there exists a one-to-one relationship in the relationship group of Aj, say Aj $\rightarrow$ Ak, and the highest degree of belongingness of Ak occurs at interval uk, then the forecasted output of F(t) equals the midpoint of uk.

2. If Aj is empty, *i.e.* Aj $\rightarrow \varnothing$, and the interval where Aj has the highest degree of belongingness is uj, then the forecasted output equals the midpoint of uj.

3. If there exists a one-to-many relationship in the relationship group of Aj, say Aj $\rightarrow A_1,A_2,....,A_n$, and the highest degrees of belongingness occurs at set u_1, u_2,......,u_n, then the forecasted output is computed as the average of the midpoints m_1, m_2,......,mn of u_1, u_2,......,u_n. This equation can be expressed as

$$\text{Mean Square error (MSE)} = \frac{\sum_t^n \left(actual\ value_i - forecasted\ value_i\right)^2}{n}$$

Model-2

Similar procedure of defuzzification as in model-1 with additional concept of repeated relations and according weighted mean is computed keeping in view if their frequencies.

Model-3

The forecasted value at time 't' is determined by the following three heuristic rules proposed by Rajaram and Vamitha. Assume the fuzzy number at time t – 1 is A_j.

Rule 1: If the fuzzy logical relationship group of A_j is empty, $A_j \rightarrow \varnothing$ or $A_j \rightarrow A_j$, then the forecasted value is R[NSTFN(A_j)].

Rule 2: If the fuzzy logical relationship group of A_j is one to one, i.e., $A_j \rightarrow A_k$ then the forecasted value is R[NSTFN(A_k)].

Rule 3: If the fuzzy logical relationship group of A_j is one to many *i.e.*, $A_j \rightarrow A_{k1}$, $A_j \rightarrow Ak_2$, …, $A_j \rightarrow A_{kp}$, then the forecasted value is calculated as

$$\frac{R\left[\text{NSTFN}\left(\text{AK1}\right) + \text{NSTFN(Ak2)} + ... + \text{NSTFN(Akp)}\right]}{P}$$

To exemplify this, the forecasted value for the year 2003-04; the fuzzified production of the year 2002-03 is A1 and the corresponding fuzzy logical relationship group is A1-A19.

A19 = [2433, 2476, 2520, 2780]. By note t_2= 2476; t_3= 2520;

T1= 2476-2433 = 43.33; t4= 2780-2520=260

$$\frac{t4 - t1}{4} = 54.25, \qquad \frac{t4 + t1}{2} = 151.50$$

NSTFN(A_{19}) = [2530.25-151.50, 2476+54.25, 2520+54.25, 2574.25+151.50]

= [2378.75, 2530.25, 2574.25, 2725.75]

According to rule 2, the forecast value for 2003-04 is given by R[NSTFN(A_{19})] =2552.25

[1995-96]: the fuzzified production of the year 1994-95 is A10. The fuzzy logical relationship group is A10-A12, A18.

NSTFN(A_{12})=[2147.75, 2185.25, 2222.25, 2259.75]

NSTFN(A_{18})=[2389.75, 2433.25, 2476.25, 2519.75]

According to rule 3, the forecast for the year 1995-96 is given by

$$= \left[\frac{2147.75 + 2389.75}{2}, \frac{2185.25 + 2433.25}{2}, \frac{2222.25 + 2476.25}{2}, \frac{2259.75 + 2519.75}{2}\right] = 2329.25$$

Model-4

The intervals (Vi), which are calculated with the help of frequency distribution as shown in the step.4.1 are used to forecast the Production at time t. All the data must be classified in to the corresponding fuzzy numbers as shown in the table 2 and forecasting the production by following the defuzzification rules given by chen in model-1. To exemplify this, the forecasted value for the year 2003-04; the fuzzified production of the year 2002-03 is A1 and the corresponding fuzzy logical relationship group is A1-A19. The corresponding interval is V19 = [2476, 2520] according to defuzzification rule 1 given in the model-1 the forecasted value for the year 2003-04 $= \frac{2476+2520}{2} = 2498.00$

[1995-96]: the fuzzified production of the year 1994-95 is A10. The fuzzy logical relationship group is A10-A12, A18 and corresponding intervals are $_{V12}$=[2185, 2222] and V18=[2433, 2476] . Then the forecasted value for the year 1995-96 $= \frac{2204+2455}{2} = 2329.00$

Experimental results:

The accuracy of a the four fuzzy time series models adopted are evaluated by using Mean square error (MSE) and average forecasting error (AFE) or Mean Absolute Percentage Error (MAPE). Lower the MSE and AFE or MAPE, better the forecasting method. The formulas are as follows:

$$Mean\ Square\ error(MSE) = \frac{\sum_{t}^{n}\left(actual\ value_i - forecasted\ value_i\right)^2}{n}$$

$$Forecasting\ error = \left|\frac{forecasted\ value - actual\ value}{actual\ value}\right| X\ 100$$

$$MAPE\ or\ AFE = \frac{sun\ of\ forecasting\ error}{number\ of\ errors}$$

Where 'n' is the number years are needed to forecast the production.

In case of oil seed production, the forecasted values from model-3 and model-4 are in close agreement with each other, where as model-1 and model-2 exhibits some variation with the other two models and which can be visualized in figure.1.

The comparison of the accuracy of four fuzzy time series forecasting models is shown in the table 5. The MSE is smaller for model-4 compared to other forecasting models but the AFE (%) or MAPE (%) is smaller for Model-3 compared with other models as shown in table-5 and figure-2 and there is a negligible difference of about 0.26% between model-3 and model-4. Overally, the combined approach of Model-1 and Model-3 as a model-4 is performed well in forecasting oil seed production in India.

Table 1 Historical total oil seed production from 1988-89 to 2013-14 in India

Year	Oil Seed Production (000' tonnes)	Year	Oil Seed Production (000' tonnes)
1988-89	1803	2001-02	2066
1989-90	1692	2002-03	1484
1990-91	1861	2003-04	2519
1991-92	1860	2004-05	2435
1992-93	2011	2005-06	2798
1993-94	2150	2006-07	2429
1994-95	2134	2007-08	2976
1995-96	2211	2008-09	2772
1996-97	2438	2009-10	2488
1997-98	2132	2010-11	3248
1998-99	2475	2011-12	2980
1999-00	2072	2012-13	3094
2000-01	1844	2013-14	3275

4. Conclusion

The study provides a foundation for the application of fuzzy time series models for short term agricultural production forecasting. In this paper we have applied three existing fuzzy time series forecasting models proposed by Chen (model-1), Chen's refined model (model-2), Rajaram and vamitha (model-3) and developed a combined approach model by using model-1 and model-3 (model-4) for the forecasting of the total oil seed production in India. The forecasting accuracy is improved by using the frequency distribution to divide the intervals with different length and these intervals are used for forecasting the production by following the defuzzification rules given by chen in model-1. Results obtained demonstrate the effectiveness of developed model comparing to other three methods in accuracy and simplicity.

ACKNOWLEDGEMENT

The authors are grateful to the Department of Science and Technology- Innovative Science Pursuit for Inspired Research (DST-INSPIRE), India for encouraging and providing financial support for the research.

Table 2 Fuzzified historical total oil seed production from 1988-89 to 2013-14 in India

Year	Oil Seed Production (000' tonnes)	Fuzzified Production		Year	Oil Seed Production (000' tonnes)	Fuzzified Production	
		Model -1 & 2	Model -3 & 4			Model -1 & 2	Model -3 & 4
1988-89	1803	A_2	A_3	2001-02	2066	A_3	A_8
1989-90	1692	A_1	A_2	2002-03	1484	A_1	A_1
1990-91	1861	A_2	A_4	2003-04	2519	A_4	A_{19}
1991-92	1860	A_2	A_4	2004-05	2435	A_4	A_{18}
1992-93	2011	A_3	A_7	2005-06	2798	A_6	A_{21}
1993-94	2150	A_3	A_{11}	2006-07	2429	A_4	A_{17}
1994-95	2134	A_3	A_{10}	2007-08	2976	A_6	A_{23}
1995-96	2211	A_3	A_{12}	2008-09	2772	A_5	A_{20}
1996-97	2438	A_4	A_{18}	2009-10	2488	A_4	A_{19}
1997-98	2132	A_3	A_{10}	2010-11	3248	A_7	A_{26}
1998-99	2475	A_4	A_{18}	2011-12	2980	A_6	A_{23}
1999-00	2072	A_3	A_8	2012-13	3094	A_7	A_{24}
2000-01	1844	A_2	A_4	2013-14	3275	A_7	A_{26}

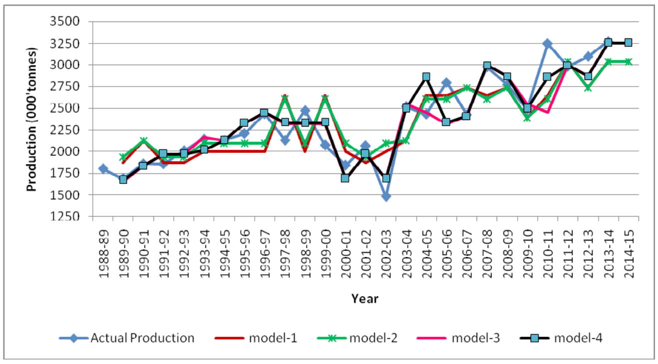

Fig 1: Fuzzy time series forecasting of oil seed production

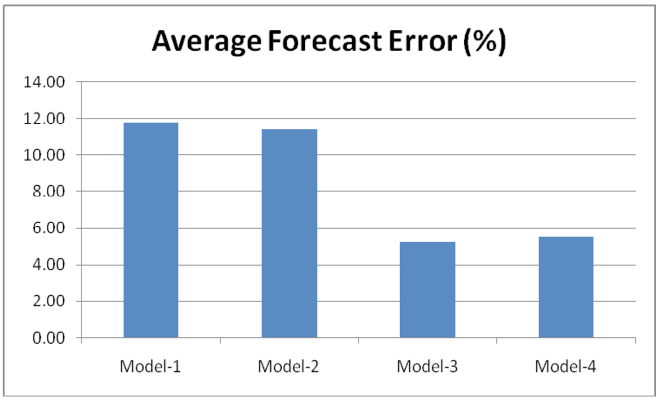

Fig 2: Comparison of average forecast error of oil seed production forecasted values

Table 3 a Fuzzy logical relationship groups for model-1 and model-2.

1.	$A_1 \rightarrow A_2$	$A_1 \rightarrow A_4$					
2.	$A_2 \rightarrow A_1$	$A_2 \rightarrow A_2$	$A_2 \rightarrow A_3$	$A_2 \rightarrow A_3$			
3.	$A_3 \rightarrow A_1$	$A_3 \rightarrow A_2$	$A_3 \rightarrow A_3$	$A_3 \rightarrow A_3$	$A_3 \rightarrow A_3$	$A_3 \rightarrow A_4$	$A_3 \rightarrow A_4$
4.	$A_4 \rightarrow A_3$	$A_4 \rightarrow A_3$	$A_4 \rightarrow A_4$	$A_4 \rightarrow A_6$	$A_4 \rightarrow A_6$	$A_4 \rightarrow A_7$	
5.	$A_5 \rightarrow A_4$						
6.	$A_6 \rightarrow A_4$	$A_6 \rightarrow A_5$	$A_6 \rightarrow A_7$				
7.	$A_7 \rightarrow A_6$	$A_7 \rightarrow A_7$					

Table 3 b Fuzzy logical relationship groups for model-1 and model-2 without repetition.

1.	$A_1 \rightarrow A_2$	$A_1 \rightarrow A_4$		
2.	$A_2 \rightarrow A_1$	$A_2 \rightarrow A_2$	$A_2 \rightarrow A_3$	
3.	$A_3 \rightarrow A_1$	$A_3 \rightarrow A_2$	$A_3 \rightarrow A_3$	$A_3 \rightarrow A_4$
4.	$A_4 \rightarrow A_3$	$A_4 \rightarrow A_4$	$A_4 \rightarrow A_6$	$A_4 \rightarrow A_7$
5.	$A_5 \rightarrow A_4$			
6.	$A_6 \rightarrow A_4$	$A_6 \rightarrow A_5$	$A_6 \rightarrow A_7$	
7.	$A_7 \rightarrow A_6$	$A_7 \rightarrow A_7$		

Table 4 Fuzzy logical relationship groups for model-3 and model-4.

1.	$A_1 \rightarrow A_{19}$		
2.	$A_2 \rightarrow A_4$		
3.	$A_3 \rightarrow A_2$		
4.	$A_4 \rightarrow A_4$	$A_4 \rightarrow _{A7}$	$A_4 \rightarrow A_8$
5.	$A_7 \rightarrow A_{11}$		
6.	$A_8 \rightarrow A_1$	$A_8 \rightarrow A_4$	
7.	$A_{10}A_{12}$	$A_{10} \rightarrow A_{18}$	
8.	$A_{11}A_{10}$		
9.	$A_{12}A_{18}$		
10.	$A_{17}A_{23}$		
11.	$A_{18}A_8$	$A_{18} \rightarrow A_{10}$	$A_{18} \rightarrow A_{21}$
12.	$A_{19}A_{18}$	$A_{19} \rightarrow A_{26}$	
13.	$A_{20}A_{19}$		
14.	$A_{21}A_{17}$		
15.	$A_{23}A_{20}$	$A_{23} \rightarrow A_{24}$	
16.	$A_{24}A_{26}$		
17.	$A_{26} \rightarrow A_{23}$		

Table 5 Comparison of the four fuzzy time series forecasting models with forecasted values and accuracy indices.

Year	Actual Production	Forecasted production			
		model-1	model-2	model-3	model-4
1988-89	1803	-	-	-	-
1989-90	1692	1870.00	1935.00	1658.75	1675.00
1990-91	1861	2130.00	2130.00	1837.50	1837.50
1991-92	1860	1870.00	1935.00	1968.17	1970.50
1992-93	2011	1870.00	1935.00	1968.17	1970.50
1993-94	2150	2000.00	2092.86	2166.50	2018.50
1994-95	2134	2000.00	2092.86	2129.50	2129.50
1995-96	2211	2000.00	2092.86	2329.25	2329.00
1996-97	2438	2000.00	2092.86	2454.75	2454.50
1997-98	2132	2650.00	2606.67	2321.58	2336.00
1998-99	2475	2000.00	2092.86	2329.25	2329.00
1999-00	2072	2650.00	2606.67	2321.58	2336.00
2000-01	1844	2000.00	2092.86	1691.25	1691.25
2001-02	2066	1870.00	1935.00	1968.25	1970.50
2002-03	1484	2000.00	2092.86	1691.25	1691.25
2003-04	2519	2130.00	2130.00	2552.25	2498.00
2004-05	2435	2650.00	2606.67	2454.75	2855.50
2005-06	2798	2650.00	2606.67	2321.58	2336.00
2006-07	2429	2736.67	2736.67	2411.25	2411.50
2007-08	2976	2650.00	2606.67	2996.25	2996.50
2008-09	2772	2736.67	2736.67	2871.75	2866.50
2009-10	2488	2390.00	2390.00	2552.25	2498.00
2010-11	3248	2650.00	2606.67	2454.75	2855.50
2011-12	2980	3040.00	3040.00	2996.25	2996.50
2012-13	3094	2736.67	2736.67	2871.75	2866.50
2013-14	3275	3040.00	3040.00	3256.50	3256.50
2014-15	-	3040.00	3040.00	3256.50	3256.50
MSE		101329.65	97799.84	45903.40	**34444.92**
AFE or					
MAPE(%)		11.80	11.41	**5.26**	5.52

REFERENCES

Zadeh L (1965) [1]: A, "Fuzzy sets". Information and Control, **8 (1965):** 338-53.

Narendra Kumar et al.[2]: "Fuzzy time series forecasting of wheat production", International Journal on Computer Science and Engineering, Vol. 02(03) 2010.

Song Q. and Brad S Chissom (1993)[3]: Fuzzy time series and its models, *Fuzzy Sets and Systems.* **54:** 269-77.

S.M. Chen (1995) [4]: "A new method for forecasting enrollments using Fuzzy time series," in Proc. 1995 National computer Symp. Chungli, Taoyuan, Taiwan, R.O.C., PP. 603-09.

S. M. Chen (1996)[5]: "forecasting enrollments based on Fuzzy time series", *Fuzzy Sets and Systems,* 81 311-19.

Poulsen J.R. and Arroyo D. O (2009) [6]: "Fuzzy Time Series Forecasting- Developing a new forecasting model based on high order fuzzy time series", Aalborg University Esbjerg (AAUE):

Rajaram S. and Vamitha V (2012) [7]: "A Modified Approach on Fuzzy Time Series Forecasting", Annals of Pure and Applied Mathematics. Vol.2 (1), 96-106.