DOCUMENT RESUME

ED 340 733 TM 017 784

AUTHOR	Song, Qiang; Chissom, Brad S.
TITLE	Forecasting Enrollments with Fuzzy Time Series.
PUB DATE	Nov 91
NOTE	16p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (20th, Lexington, KY, November 12-15, 1991).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Educational Trends; *Enrollment Projections; *Estimation (Mathematics); Higher Education; Long Range Planning; Mathematical Models; *Regression (Statistics); *Trend Analysis; *Universities
IDENTIFIERS	Fuzzy Set Theory; *Fuzzy Time Series

ABSTRACT

The concept of fuzzy time series is introduced and used to forecast the enrollment of a university. Fuzzy time series, an aspect of fuzzy set theory, forecasts enrollment using a first-order time-invariant model. To evaluate the model, the conventional linear regression technique is applied and the predicted values obtained are compared to the fuzzy time series results and actual enrollments. The forecasting procedure begins with "fuzzylising" the universe on which the historical data are based, and then interprets output results (actually fuzzy data sets). Comparison with linear regression shows the superiority of the fuzzy time series for forecasting enrollment. Three tables give comparison figures for fuzzy time series, linear regression, and actual values largely for 1972 through 1990. Three references are listed. (SLD)

Forecasting Enrollments with Fuzzy Time Series

Part I

Qiang Song and Brad S. Chissom

College of Education

The University of Alabama

Paper Presented at The Annual Meeting

Of the Mid-South Educational Research

Association, Lexington, KY,

November 13-15, 1991

Abstract

In this paper, the authors use fuzzy time series and its model to forecast the enrollments of a university. A method is applied and the procedures are described. The calculated results show that the method using fuzzy time series is feasible and has many advantages.

Key words: Fuzzy Time Series, Enrollments, Memberships, Forecasting.

I. Introduction

Fuzzy time series is a new concept proposed by the authors. In a recent study, Song and Chissom (1991) put forth the definition of fuzzy time series and the outline of its modeling using fuzzy logical reasoning for the first time. Although in the literature, one can easily find research papers introducing successful applications of fuzzy logical reasoning, fuzzy control, fuzzy linear regression and many other aspects of fuzzy set theory and applications, the concept of fuzzy time series has not been delineated. This is not surprising because we are just at the point where fuzzy time series, both theory and application, can be studied systematically. Bintley (1987) has successfully applied fuzzy logical reasoning to a practical case of forecasting, but he did not use the concept of fuzzy time series or the order of the model. In this paper, the concept of fuzzy time series is used to forecast the enrollments of a university in the United States using real data. A first-order, time-invariant model will be used. To evaluate the model, the conventional linear regression technique is applied to the same data and different predicted values are obtained. The results indicate that there are many advantages to using fuzzy time series.

II. Some Concepts of Fuzzy Time Series and Its Models

Definition 1. Let $Y(t) \in R^1$ $(t=0,1,2, \ldots)$ be a time series. If $f_i(t)$ is a fuzzy set in $Y(t)$ and $F(t)=\{f_1(t), f_2(t), \ldots\}$, then $F(t)$ is called a fuzzy time series in $Y(t)$.

For simplicity, only the first-order model will be used. Therefore, just the definition of the first-order model is given below.

Definition 2. Suppose $F(t)$ is caused by $F(t-1)$ only, i.e.,$F(t-1) \dashrightarrow F(t)$. Then this relation can be expressed as $F(t)=F(t-1) \circ R(t,t-1)$ where $R(t,t-1)$ is a fuzzy relationship and is called the first-order model of $F(t)$.

Definition 3. Suppose $R(t,t-1)$ is a first-order model of $F(t)$. If for any t $R(t,t-1)$ is independent on t, i.e., for any t, $R(t,t-1) = R(t-1,t-2)$, then $F(t)$ is called a time-invariant fuzzy time series or else it is called a time-variant fuzzy time series.

To appreciate the method used below, the following theorems are useful.

Theorem 1. Let $F(t)$ be a fuzzy time series. If for any t, $F(t)=F(t-1)$, then $F(t)$ is a time-invariant fuzzy time series.

We can see that just as the definition of $f_i(t)$ $(i=1,2,.....)$ is subjective, so is $F(t)$, i.e., we can define either a time-invariant or a time-variant fuzzy time series on one universe. But since the definition of $f_i(t)$ is not arbitrary, the concepts of time-invariant and time-variant fuzzy times series are meaningful, which can be seen from the following theorem.

Theorem 2. If $F(t)$ is a fuzzy time series, $F(t)=F(t-1)$ for any t; and $F(t)$ has only finite elements $f_i(t)$, then

$$R(t,t-1)=f_{i1}(t-1) x f_{j1}(t) U f_{i2}(t-2) x f_{j2}(t-1) U \ldots\ldots\ldots$$
$$\ldots\ldots\ldots U f_{im}(t-m) x f_{jm}(t-m+1) \qquad (1)$$

where $m>0$.

This theorem implies that in the case of time-invariant fuzzy time series, it is very easy and convenient to calculate the first-order model. As a matter of fact, since we can hardly define infinite fuzzy sets on any universe, once $F(t)=F(t-1)$, i.e., at any two successive time points t_1 and t_2, we have the same fuzzy sets, then we can get a time-invariant fuzzy time series. Therefore, Theorem 2 is very useful.

III. Forecasting Enrollments with Fuzzy Time Series

In the application of fuzzy time series to any practical case, there are two main steps. The first one is modeling, i.e., according to the historical data or experience knowledge setting up the fuzzy logical model. Here, we should point out the particular importance of historical experience knowledge in modeling. This is because, in some cases, we might not be able to get accurate data to use but we have or can collect some practical experiences from the persons who are acquainted with what we are to forecast. Using the experience, we can develop a kind of model for use. Thus, we can perceive the features of this kind of forecasting method.

The second step is to interpret the results from the forecasting model. From the following section we could see that human subjective experiences has a very important role in the interpretation of the results. With the above descriptions, the procedures used to forecast the enrollments follow. The procedures are made up of two steps:

A. Fuzzylising the universe on which the historical data are based. We have on hand the enrollments of the university from 1971 to 1990. The maximum enrollment is 19328 while the minimum is 13055. Usually the universe should be larger than the interval formed by the maximum and the minimum of the data. We select (13000, 20000) as the universe.

To fuzzylise the universe, first, departmentalize the overall interval (13000, 20000) into seven even lengthy intervals. We use u1, u2, u3, u4, u5, u6 and u7 for each interval, i.e., u1=(13000, 14000), u2=(14000, 15000), u3=(15000, 16000), u4=(16000, 17000), u5=(17000, 18000), u6=(18000, 19000) and u7=(19000, 20000).

Secondly, define some fuzzy sets on the universe (13000, 20000). Let A_1=(not many), A_2=(not too many), A_3=(many), A_4=(many many), A_5=(very many), A_6=(too many), and A_7=(too many many). There is no restriction on the number of the fuzzy sets defined. The elements of each fuzzy set are u1, u2,......., and u7. The authors have determined the memberships for each element in the respective fuzzy sets. Thus, all the sets are presented as follows.

$$
\begin{aligned}
A_1 &= \{u1/1,\ u2/.5,\ u3/0,\ u4/0,\ u5/0,\ u6/0,\ u7/0\}; \\
A_2 &= \{u1/.5,\ u2/1,\ u3/.5,\ u4/0,\ u5/0,\ u6/0,\ u7/0\}; \\
A_3 &= \{u1/0,\ u2/.5,\ u3/1,\ u4/.5,\ u5/0,\ u6/0,\ u7/0\}; \\
A_4 &= \{u1/0,\ u2/0,\ u3/.5,\ u4/1,\ u5/.5,\ u6/0,\ u7/0\}; \\
A_5 &= \{u1/0,\ u2/0,\ u3/0,\ u4/.5,\ u5/1,\ u6/.5,\ u7/0\}; \\
A_6 &= \{u1/0,\ u2/0,\ u3/0,\ u4/0,\ u5/.5,\ u6/1,\ u7/.5\}; \\
A_7 &= \{u1/0\ u2/0,\ u3/0,\ u4/0,\ u5/0,\ u6/.5,\ u7/1\}.
\end{aligned}
\tag{2}
$$

For simplicity, we will also use A_1, A_2, ... , and A_7 as vectors whose elements are the corresponding memberships.

Thirdly, fuzzylise the data, i.e., find out the memberships of each year's enrollment in each fuzzy set A_i(i=1 to 7). The results are shown in Table 1.

Fourthly, from Table 1 obtain the historical experience knowledge about the evolution of the enrollment of this university. Assume that if the maximum membership of one year's enrollment is under A_i(i=1,2, ... ,or 7), then we treat this year's enrollment as A_j. For example, for 1982, the maximum membership is under A_3, then we say that the enrollment of 1982 is A_3, or many. Since we are going to find out the laws for any two successive years' enrollments in terms of fuzzy set and logical reasoning, we will develop such logical relationships as if the enrollment of year i is A_k then that of year i+1 is A_l, and so on. Using the symbols in [2], we can get all the logical relationships as follows.

$A_1 \text{-->} A_1$, $A_1 \text{-->} A_2$, $A_2 \text{-->} A_3$, $A_3 \text{-->} A_3$, $A_3 \text{-->} A_4$,

$A_4 \text{-->} A_4$, $A_4 \text{-->} A_3$, $A_4 \text{-->} A_6$, $A_6 \text{-->} A_6$ and $A_6 \text{-->} A_7$.

By definition, we know that we have defined a time-invariant fuzzy time series. Let $R_1 = A^T{}_1 x A_1$, $R_2 = A^T{}_1 x A_2$, $R_3 = A^T{}_2 x A_3$, $R_4 = A^T{}_3 x A_3$, $R_5 = A^T{}_3 x A_4$, $R_6 = A^T{}_4 x A_4$, $R_7 = A^T{}_4 x A_3$, $R_8 = A^T{}_4 x A_6$, $R_9 = A^T{}_6 x A_6$ and $R_{10} = A^T{}_6 x A_7$. Then, according to Theorem 2, we get

$$R(t,t-1) = R = \bigcup_{i=1}^{10} R_i \tag{3}$$

Some calculation yields:

$$R = \begin{bmatrix} 1 & 1 & .5 & .5 & 0 & 0 & 0 \\ .5 & .5 & 1 & .5 & .5 & 0 & 0 \\ 0 & .5 & 1 & 1 & .5 & .5 & .5 \\ 0 & .5 & 1 & 1 & .5 & 1 & .5 \\ 0 & .5 & .5 & .5 & .5 & .5 & .5 \\ 0 & 0 & 0 & 0 & .5 & 1 & 1 \\ 0 & 0 & 0 & 0 & .5 & .5 & .5 \end{bmatrix}$$

So, the forecasting model is:

$$A_i = A_{i-1} o R \tag{4}$$

where A_{i-1} is the enrollment of year i-1 and A_i the enrollment of year i in terms of fuzzy sets.

Finally, calculate the output of the model. The results are shown in Table 2.

B. Interpretation of the output results.

The output results of model (4) are actually all fuzzy sets. If the results in the form of fuzzy sets can satisfy the requirement for the forecasting job, just stop here. But in many situations, quantitative results are desired. Therefore, translating the fuzzy output into a regular number is indeed a necessary step. Since human experience knowledge has been included in the model and all the memberships, which can be clearly seen from the above modeling procedure, the output surely contains this kind of experience knowledge. So, it is not difficult to jump to the conclusion that in the interpretation of the output results, experience knowledge is still needed. Using human subjective experience knowledge is one of the most important features of fuzzy forecasting methods. Some principles are presented to interpret the results and solve the problem. The principles are:

1. If the membership of an output has only one maximum, then select the midpoint of the interval corresponding to the maximum as the output value;

2. If the membership of an output has two or more consecutive maximums, then select the midpoint of the corresponding conjunct intervals as the output value;

3. Otherwise, use the standardized membership and the midpoint of each interval to calculate the "mean" using $\sum S_i M_i$ where S_i is the standardized membership and M_i the midpoint of interval ui. And then the "mean" will be taken as the output value.

Following the above principles, we can obtain predicted values for the enrollments from 1972 to 1991 as shown in Table 2.

IV. Evaluation of The Method

This method has at least the following two advantages:

1. Experience knowledge can be utilized from the very beginning until the end of the whole forecasting process. Therefore, the requirement for the historical data is not very strict. In effect, from the above modeling procedures we can see that without the historical data we can still set up the same forecasting model if we can manage to gather or if we possess the same knowledge about the development of this university's enrollment as obtained from the data.

2. This method is better than the Linear Regression Method (LRM) in forecasting this university's enrollments. LRM was used to forecast the enrollments for this university from 1972 to 1990 with time t as the predictor. The comparative results are shown in Table 3. From there we can tell that on the whole, the fuzzy time series model (FTM) produces more precise results than LRM.

V. Discussions and Concluding Remarks

The authors also compared the predicted values of FTM with those of non-linear regression models. It was found that when only a quadratic term was included in the model, the FTM had better results, but when a cubic term was included, the non-linear regression model produced better predicted values. The reason might be that the curve of the enrollment of the university is cubic in nature. We used only one time-invariant model of fuzzy time series in our study. In the case of quite severe non-linearity, we should look for different approaches using fuzzy time series. There are at least two ways. The first one is to use time-variant models, i.e., multi-models. The second one is to use feed-back prediction information, i.e., using the prediction errors to modify the model to improve the prediction. These two possible approaches are our near-future research objectives. More results will be reported soon.

References

[1] Bintley, H. (1987). Time series analysis with REVEAL, Fuzzy Sets and Systems, 23, (1).

[2] Song, Qiang & Chissom, B. S. (1991). Fuzzy time series and its models. Submitted to Fuzzy Sets and Systems.

[3] Wu, Wangming, et. al. (1985). Applied Fuzzy Sets Methodology, Beijing: Normal University Press.

Table 1

	A1	A2	A3	A4	A5	A6	A7
1990	0	0	0	.3	.5	.8	1
1989	0	0	0	.25	.55	1	.8
1988	0	0	.1	.5	.8	1	.7
1987	0	.1	.5	1	.8	.1	0
1986	0	.2	1	.7	.2	0	0
1985	.2	.8	1	.2	0	0	0
1984	.2	.8	1	.2	0	0	0
1983	.2	.8	1	.2	0	0	0
1982	.2	.8	1	.2	0	0	0
1981	0	.2	.8	1	.5	0	0
1980	0	.1	.5	1	.9	.2	0
1979	0	.1	.5	1	.9	.2	0
1978	0	.5	1	.7	.2	0	0
1977	0	.6	1	.6	.1	0	0
1976	.2	.8	1	.2	0	0	0
1975	.2	.8	1	.2	0	0	0
1974	.8	1	.8	.1	0	0	0
1973	1	.9	.2	0	0	0	0
1972	1	.8	.1	0	0	0	0
1971	1	.5	0	0	0	0	0

Table 2

year	output membership	standardized membership	predicted value
1972	1,1,.5,.5,0,0,0	.286, .286, .143, .143, 0, 0, 0	14000
1973	1,1,.8,.5,.5,.1,.1	.25, .25, .2, .125, .125, .025. .025	14000
1974	1,1,.9,.5,.5,.2,.2	.2325, .2325, .209, .116, .116, .047, .047	14500
1975	.8,.8,1,.8,.5,.5,.5	.163,.163, .204, .163, .102, .102, .102	15500
1976	.5, .5, 1, 1, .5, .5, .5	.111, .111, .222, .222, .111, .111. .111	16000
1977	.5, .5, 1, 1, .5,.5,.5	.111, .111, .222, .222, .111, .111, .111	16000
1978	.5, .5, 1, 1, .5, .6, .5	.109, .109, .217, .217, .109, .13, .109	16000
1979	.5, .5, 1, 1, .5, .7, .5	.106, .106, .213, .213, .106, .149, .106	16000
1980	.1, .5, 1, 1, .5, 1, .5	.0217, .108, .217, .217, .108, .217, .108	16813
1981	.1, .5, 1, 1, .5, 1, .5	.0217, .108, .217, .217, .108, .217, .108	16813
1982	.2, .5, 1, 1, .5, 1, .5	.0425, .106, .213, .213, .106, .213, .106	16789
1983	.5, .5, 1, 1, .5, .5, .5	.111, .111, .222, .222, .111, .111, .111	16000
1984	.5, .5, 1, 1, .5, .5, .5	.111, .111, .222. .222, .111, .111, .111	16000
1985	.5, .5, 1, 1, .5, .5, .5	.111, .111, .222, .222, .111, .111, .111	16000
1986	.5, .5, 1, 1, .5, .5, .5	.111, .111, .222, .222, .111, .111, .111	16000
1987	.2, .5, 1, 1, .5, .7, .5	.045, .114, .227, .227, .114, .159, .114	16000

1988	.1, .5, 1, 1, .5, 1, .5	.027, .108, .217, .217, .108, .217, .108	16813
1989	0, .5, .5, .5, .5, 1, 1	0, .125, .125, .125, .125, .25, .25	19000
1990	0, .5, .5, .5, .5, 1, 1	0, .125, .125, .125, .125, .125, .25, .25	19000
1991	0, .5, .5, .5,. 5, .8, .8	0, .138, .138, .138, .138, .222, .222	19000

Table 3

year	predicted value of FTM*	residual of FTM*	predicted value of LRM	residual of LRM	actual values
1990	19000	328	18033	1295	19328
1989	19000	-30	17809	1161	18970
1988	16813	1337	17585	565	18150
1987	16000	859	17360	-501	16859
1986	16000	-16	17136	-1152	15984
1985	16000	-837	16912	-1749	15163
1984	16000	-855	16688	-1543	15145
1983	16000	-503	16464	-967	15497
1982	16789	-1356	16239	-806	15433
1981	16813	-425	16015	373	16388
1980	16813	106	15791	1128	16919
1979	16000	807	15567	1240	16807
1978	16000	-139	15342	519	15861
1977	16000	-397	15118	485	15603
1976	16000	-689	14894	417	15311
1975	15500	-40	14670	790	15460
1974	14500	196	14446	251	14696
1973	14000	-133	14221	-354	13867
1972	14000	-437	13997	-434	13563

* FTM=Fuzzy Time Series