

# Research on the Application of Fuzzy Variable Time Series

Yeping Peng

Software College of Jishou University, Zhangjiajie, Hunan Province, China

**Keywords:** fuzzy variable; time series; fuzziness; membership; data mining.

**Abstract:** In order to solve the problem of inaccurate setting of fuzzy data in the actual research process, the method of fuzzy mathematics is adopted. The importance of fuzzification of data variables is illustrated by using fuzziness. The definitions of fuzzy variables and time series of fuzzy variables are given. The necessity of fuzzification of variables is illustrated by examples of econometrics and data mining. The results show that many data in the study do have the characteristics of fuzziness, and the application of time series of fuzzy variables is helpful for more objective econometric models and time series mining. The introduction of time series of fuzzy variables has certain reference value for econometrics and data mining.

## 1. Introduction

In scientific research, whether variable data needed by the model can be obtained correctly often determines the reliability of the description of objective economic phenomena and the accuracy of the prediction of future data. However, in practical research, some variables with fuzzy attributes are often set incorrectly or even neglected, such as age, height, education, wealth etc., which are all variables with fuzzy boundaries. It is often one-sided to use the method of setting variables in general econometrics to set them. As a well prediction or evaluation cannot be achieved in this way, a time series based on fuzzy sets for predicting has been proposed. In the mode, firstly, the intervals of discourse domain and region are defined, and then the fuzzy sets and the fuzzy semantic variables are defined on the basis of discourse domain, by which the historical data are fuzzified and the categories of the fuzzy logical relations and the fuzzy logical relations are determined. However, in such a model, a lot of calculation is needed to determine the fuzzy logic relationship. When the fuzzy logic relationship is complex in the middle, the max-min operation will also take a lot of calculation time. However, the influence of region interval of fuzzy sets on prediction accuracy is not considered in this model. In addition, this model only gives fuzzy time series of discrete data, which are difficult to apply to econometric models and data mining. [1]

In this paper, another method of fuzzification of economic variables, the method of fuzzy variables, is proposed. This method is based on Zadeh's theory of fuzzy sets, and the time series of fuzzy variables obtained can be directly applied to econometric models. Thus, it overcome the shortcoming that the method of fuzzy time series cannot be applied to econometric models. The relationship between econometrics and fuzzy mathematics is linked by fuzzy variables, and the problem of variable fuzzification in econometric models can be solved. The time series of fuzzy variables obtained in this paper can also be applied to data mining and time series mining.

## 2. Definition of Time Series of Fuzzy Variables

Fuzzy variables are variables with fuzzy attributes, such as rapid economic development and high degree of aging. Since most economic data have fuzzy attributes, they are typically time series of fuzzy variables.

### 2.1 Definition of Fuzziness

Fuzzy variable data is variable data with fuzzy attributes in data. Since it has fuzzy attributes, it is necessary to find a quantitative index to describe the degree of fuzziness of fuzzy variables. A natural idea is that ordinary sets are not ambiguous and the number of indistinct degrees should be

zero; for  $U < U$ , if  $A(u) > 0.5$ ,  $A_c(u) < 0.5$ , then  $A$  is ambiguous and the number of indistinct degrees should be 1; and because  $|A(u) - 0.5| = |A_c(u) - 0.5|$ ,  $A$  and  $A_c$  should be equivalent. In addition, the magnitude of the fuzziness of the marked fuzzy set  $A$  should also have some properties: the farther away  $A(u)$  is from 0.5, the smaller its value is, and vice versa, the closer it is to 0.5, the larger its value is. This is ambiguity. The definition of ambiguity is an axiomatic definition, which cannot be directly applied to research. The following gives a computational ambiguity. [2]

If  $A \in \mathcal{F}(U)$ , then  $L(A) = \frac{2}{n} \sum_{i=1}^n |\mu_A(u_i) - A_{0.5}(u_i)|$  is defined as L-fuzziness of  $A$ , wherein  $A_{0.5}(u_i)$  is  $\lambda$  cut set of  $A$  and  $\lambda = 0.5$ . L-fuzziness can be used to measure the degree of fuzziness of fuzzy variables.

## 2.2 Definition of Fuzzy Variable Time Series

Let  $a_{u_{ij}}$  represent the time series data of a variable in the period  $j$ , then

$\mu_A: U \rightarrow [0,1]$ , wherein  $\mu$  is the membership function of fuzzy subset  $A$ , and  $\mu_A(u)$  is the membership degree of  $u$  to  $A$ .

$$Z_j = \frac{\sum_{i=1}^n a_{u_{ij}} \cdot \mu_A(u_i)}{\sum_{i=1}^n a_{u_{ij}}} \quad (1)$$

is called the fuzzy variable time series of the variable concerned in the period  $j$ , wherein  $\mu = 1, 2, \dots, n$ ,  $\sum_{i=1}^n a_{u_{ij}}$  is the sum of the variable data in the period  $j$ . [3]

The fuzzy variable time series  $Z_j$  can be used as the explanatory variable or the explanatory variable in the econometric model, and the fuzzy variable can be used to study the econometric model. At the same time, it can be applied to data mining as well.

## 3. Classification of Fuzzy Variable Time Series

### 3.1 Discrete Fuzzy Variable Time Series

Firstly, if the fuzzy set  $A$  is discrete, the membership degree of  $u_i$  to the fuzzy set  $A$ ,  $\mu_A(u_i)$  can be obtained by setting the fuzzy set, and then the time series of fuzzy variables can be obtained by formula (1). For example, in terms of the setting of the fuzzy variables of education level, people concerned can be divided into graduates, undergraduates, junior college, senior high school and vocational high school, junior high school and primary school. Set  $a_{u_{ij}}$  for the number of graduates of different educational levels in the year  $j$  and  $u_i$  for educational level. Among them,  $u_1$  stands for graduate student,  $u_2$  an undergraduate student,  $u_3$  is a junior college student,  $u_4$  a senior high school student or a vocational middle school student,  $u_5$  a junior high school student, and  $u_6$  a primary school student. They form a domain  $U$ ,  $U = \{u_1, u_2, u_3, u_4, u_5, u_6\}$ .  $J$  stands for year, for the fuzzy set high level of education, and  $\mu_A(u_i)$  is the membership degree of the  $u_i$  to  $A$ . The membership degree of  $u_1, u_2, u_3, u_4, u_5, u_6$  to the fuzzy set  $A$ , i.e., high level of education can be thus defined respectively. [4]

### 3.2 Continuous Fuzzy Variable Time Series

Although continuous time series of fuzzy variables are seldom used in practice and thus less studied, it still has certain theoretical significance. When time  $j$  is continuous, then

$$Z_x = \frac{\int_{x_0}^x \mu_A(u_i) \cdot a_{u_i y} dy}{\int_{x_0}^x a_{u_i y} dy}$$

is defined as time series of continuous fuzzy variables, wherein  $x_0 \leq x \leq 1$ .

## 4. Instantiation in Econometric Study

Traditional economic theory holds that there is a trend of co-change between consumption and income. The relationship between consumption and income can be expressed by  $C = aY + b$ , in which  $C$  is consumption,  $Y$  is income and  $a$  is marginal propensity to consume. Nowadays, with social

development, education level may also play a certain role in promoting the consumption level. People's consumption concept will change in terms of what education they have received. Therefore, it is necessary to include educational level into the consumption equation.[5]

However, education level is always expressed in vague words such as "high", "low", "relatively high". This indicated that it should be a vague variable. In some econometric studies, people's educational level is often represented by some simple proportions, such as the proportion of college graduates to the total population of the country each year, and so on. Such a simple setting ignores the human factors with other educational levels, so the obtained model is also impersonal. In this paper, the fuzzy variable is used to set the level of culture and education for all, and use the time series of the obtained fuzzy variable to get a more accurate consumption model.

If consumption level(CONSUME)is variable to be explained, income level (INCOME)and education level (EDUCATION) explanatory variables, a logarithmic entry model is obtained because of rapid development of both CONSUME and INCOME.

If the explanatory variable of educational level is set by the fuzzy variable time series obtained above, the multiple regression model can be obtained by Eviews statistical software:

$$\ln \text{ CONSUME}=0.66 \cdot \ln \text{ INCOME}+ 2.50 \cdot \text{ EDUCATION}+ 13.22$$

The marginal propensity to consume is 0.66, which shows that when the logarithm of people's income increases by 1%, the logarithm of consumption increases by 0.66%. The contribution of education development level to consumption is 2.50. If education level increases by 1%, the logarithm of consumption will increase by 2.50%. This shows that although the level of cultural and educational development of the whole people is slow, its impact on people's consumption is still very important. Improving the cultural quality of the whole society will inevitably lead to an increase in consumption. [6]

If a simple proportion (such as the ratio of the number of university graduates to the total population of the whole country) is used as the explanatory variable of the level of social cultural and educational development, the following consumption model can be obtained:

$$\ln \text{ CONSUME}=0.67 \cdot \ln \text{ INCOME}+ 0.011 \cdot \text{ EDUCATION} + 13.88$$

Here, the contribution of education level to consumption is only 0.011, and the impact of education development level on consumption is obviously underestimated.

#### 4.1 Mining of Fuzzy Variable Time Series

A temporal database is a database consisting of sequence values or events that change over time. Time series database is very common in practical applications, such as daily fluctuations in the stock market, dynamic product processing, scientific experiments, medical treatment and so on. However, time series with fuzzy attributes are relatively rare.

#### 4.2 N-order moving average

By giving a set of values for the time series variable Y, i.e., (y<sub>1</sub>, y<sub>2</sub>...), the n-order moving average is obtained:

$$\frac{y_1+y_2+y_3 + \dots + y_n}{n}, \frac{y_2+y_3+y_4 + \dots + y_{n+1}}{n}, \frac{y_3+y_4+y_4 + \dots + y_{n+2}}{n} \dots$$

Moving average can reduce the total amount of change in the data set. Therefore, replacing time series with moving average can reduce undesirable fluctuations, so it is also called smoothing of time series.[7]

Via calculation on the fuzzy variable time series of education level above, the 3<sup>rd</sup>-order moving average can be obtained.

#### Summary

This paper has studied time series of data with fuzzy attributes and proposed a new method of fuzzification of economic variables. Firstly, from the analysis of fuzziness, some data do have fuzzy attributes and should be solved by using the method of fuzzy mathematics. Secondly, fuzzy variable time series can be applied to econometrics to solve the problem of fuzzification of variable data,

which can make the model more objective and reliable. Thirdly, fuzzy variable time series can also be applied to data mining and get directly perceived results.

### **Acknowledgement**

This research was financially supported by the General Project of Hunan Provincial Department of Education (No. 16C1312).

### **References**

- [1] Q. Song, B. S. Chissom, New models for forecasting enrollments: fuzzy time series and neural network approaches, Annual Meeting of the American Educational Research Association, 21 (1993) 1-27.
- [2] S. Askari, N. Montazerin, A high-order multi-variable Fuzzy Time Series forecasting algorithm based on fuzzy clustering, Expert Systems with Applications, 42, 4 (2015) 2121-2135.
- [3] L. I. Guannan, M. Shan, Combined forecasting model on time series of complex systems based on fuzzy variable weight optimized by adaptive genetic algorithm, Journal of Central South University, 44, 11 (2013)4542-4547.
- [4] S. M. Chen, Forecasting enrollments based on fuzzy time series, Fuzzy Sets & Systems, 81, 3(1996)311-319.
- [5] K. Huarng, H. K. Yu, A Type 2 fuzzy time series model for stock index forecasting, Physica A Statistical Mechanics & Its Applications, 353(2005)445-462.
- [6] H. S. Lee, M. T. Chou, Fuzzy forecasting based on fuzzy time series, International Journal of Computer Mathematics,81, 7 (2004)781-789.
- [7] X. Zhou, K. Zou, Y.Wang, Fuzzy Variable Time Series Based on Fuzzy Membership Function and Econometrics,2010 Third International Symposium on Knowledge Acquisition and Modeling, (2010) 225-228.