Bulletin of the *Transilvania* University of Braşov • Vol 3(52) - 2010
Series III: Mathematics, Informatics, Physics, 125-132

AN APPLICATION OF FUZZY TIME SERIES TO THE ROMANIAN POPULATION

Adela SASU[1]

Abstract

In this paper we use a strategy for forecasting the population count of Romania using a method inspired by the Fuzzy set theory. Various forecasting methods have been developed on the basis of fuzzy time series data, but accuracy has been matter of concern in these methods. In fuzzy time series methods the forecasted values depend to some degree on our interpretation of the model's output, and hence different interpretations may lead to different results.

2000 *Mathematics Subject Classification:* 60G25, 62M10, 03E72, 94D05.
Key words: time series, fuzzy sets, forecasting.

1 Introduction

Many methods have been proposed for forecasting based on fuzzy time series. The initial work of Zadeh [9], [10] concerning fuzzy set theory has been applied to several diverse areas. Song and Chissom in [5], [6], [7] introduced a theory for fuzzy time series and applied fuzzy time series methods that modeled the enrollments of the University of Alabama. In recent years, a number of techniques have been proposed for forecasting based on fuzzy set theory methods ([1], [2], [3], [4], [8]).

The development stage of society, characterized by the implementation of market mechanism into all the spheres of human activity, sets greater requirements upon the perspective calculation of forecasted total population. The demographic forecasting, by being an integral part of social-economic development, allows us to assess the anticipated total population, the economically active population, the size of different age cohorts etc. These factors should be taken into consideration in formulating a scientifically feasible social-economic development policy and preparing the complex practical measures aimed at the implementation of this policy.

The development of information technologies and software resources has opened new opportunities for modeling demographic processes and handling forecasting problems. Researches carried out over the recent years prove that the application of traditional analyzing methods and modeling of the population growth process on the basis of processing

[1]Faculty of Mathematics and Informatics, *Transilvania* University of Braşov, Romania, e-mail: asasu@unitbv.ro

numeric/quantitative data does not produce the desired results and even involves considerable risks and errors. The principal causes for this undesirable circumstance stem from the fact that a great many forecasting models are not sufficiently efficient due to the incompatibility of highly accurate quantitative methods of classical mathematical apparatus with the great complexity of the population growth process. The other cause is that these methods aimed at the mathematical analysis of accurately determined systems are not capable to encompass certain characteristics of the research sphere.

2 A brief presentation of fuzzy time series

Time series represent consecutive series of observation that are conducted by equal time intervals and lies at the root of exploring real processes in economics, meteorology and natural sciences etc.

Formally, time series can be defined as a discrete function $x(t)$ whose argument and function values are dependent on discrete time moments as well as on argument values, function values at different time intervals. It is assumed that the time interval $t \in [0, T]$ of process $x(t)$ is observed, that is to say, parameter t varies along the time interval $[0, T]$ or assumes any integer belonging to this interval. For every fixed time moment $t = s$, the value of the function, beginning from this moment, is generally determined by the values of the function arguments at all the time moments ranging from $t = 0$ to $t = s - 1$, and the value of function at all the time moments ranging from $t = 0$ to $t = s - 2$.

Let us assume that $U = \{u_1, u_2, ..., u_n\}$ is a universal time set. The fuzzy set A of the universal set U is defined as follows

$$A = \{\mu_A(u_1)/u_1, \mu_A(u_2)/u_2, ..., \mu_A(u_n)/u_n\},$$

where $\mu_A(u_i)$ is the membership function, $\mu_A(u_i) : U \to [0, 1]$, $\mu_A(u_i)$ is a degree of belonging of u_i to the set A, "/" is a division sign.

Let us assume that $Y(t)$ $(t = 1, 2, 3, ...)$ which is a subset of set $\mathbb{R}$ of real numbers, is simultaneously a universal set on which a fuzzy set $\mu_i(t)$ is defined, that is, the membership function is time dependent. Let us define a set $F(t)$ arranged out of $\{\mu_i(t), t = 1, 2, ...\}$. More precisely, $F(t)$ is a set of fuzzy sets $F(t) = \{\mu_i(t), t = 1, 2, \}$. Then $F(t)$ is a fuzzy time series defined on a universal set $Y(t)$ $(t = 1, 2, 3.)$. $F(t)$ is time-dependent, which means, function $F(t)$ will assume different values at different time moments. The intensive changes in demographic processes that are caused by the influence of the social-demographic factors, have rendered the determination of perspective variation in total population one of the most important tasks to be tackled for demographic forecasting. To solve the task of forecasting the total population, we have introduced a model of fuzzy time series in this article. More precisely, the problem is described as following: for a given time interval, data pertaining to the total population of Romania or, to be more clear, the dynamics and respective variation of the total population are available, but the problem is to find the anticipated total population based on the variations of the previous years.

The following principles are recommended for the solution of this problem:

1. Since the proposed method is applied to demographic forecasting for the identification of the model or for finding the extent to which it conforms to (reflects) the real process, we should, first of all, give a "retrospective forecast" that comprises the following:

   - one of the previous years ($t = s$) is selected as a forecast year and the total population is calculated for this year based on the variations in total population of prior years ($t = s - 1, s - 2, , s - k$);
   - the obtained results are compared with the retrospective data (real data of the s-th year) and the subsequent error is estimated;
   - the experiment is carried out over a fixed time interval;
   - the effectiveness of the method is evaluated based on the value of the subsequent error.

2. If positive results are obtained, the model should be applied to computing the estimated total population.

Abbasov and Mamedova applied these principles in [1] to study the total population in Azerbaijan.

3 Forecasting Methodology

We applied the methodology proposed by Abbasov and Mamedova in [1] for studying the evolution of population forecasting of Romania.

In accordance with the description of the problem, the following forecasting methodology is:

1. Definition of the universal set U containing the interval between the smallest and the greatest variations in total population.

2. Division of the universal set U into equal length intervals containing variation values corresponding to different population growth rates.

3. The qualitative description of variation values of total population as a linguistic variable, that is to say, determining the respective values of linguistic variable or the set of fuzzy sets $F(t)$.

4. Fuzzifying the input data or the conversion of numerical values into fuzzy values. This operation enables us to reflect the corresponding numerical/qualitative values of qualitative representations of population growth rates in the value of membership function.

5. Selection of parameter $w > 1$, corresponding to the time period prior to the concerned year, calculation of fuzzy relationships matrix and forecasting of population growth in the next year.

The application of the proposed methodology to population forecasting is described briefly in the experimental section.

4 Experimental Section

In the experimental section we used annual data to represent the total population(thousand persons) of Romania between 1988-2009 taken from the site www.census.gov, more precisely from http://www.census.gov/ipc/www/idb/country.php.

Table 1 gives the dynamics of the total population over 1988-2009 (input data for "retrospective forecast") and the variation in total population between every next and previous year. The variation for the current year is understood to be the difference between the sizes of population in current and previous years. For example, the variation for 1990 is equal to $22866 - 22852 = 14$. To define a universal set U, first of all, the smallest and greatest variation values must be found over the period $[1988, 2009]$; these extremal values are denoted as D_1 and D_2, respectively. After that, the universal set U can be defined as $U = [V_{min} - D_1, V_{max} + D_2]$, where $V_{min} = -66000$ is the smallest variation (year 1977), $V_{max} = 83000$ is the greatest variation (year 1989), $D_1 = 3000$, $D_2 = 2000$. Thus, the universal set U will be as $U = [-69000, 85000]$.

The universal set U must be divided into several equal intervals. In our case, this set U is divided into seven equal length intervals: $u_1 = [-69000, -47000]$, $u_2 = [-47000, -25000]$, $u_3 = [-25000, -3000]$, $u_4 = [-3000, 19000]$, $u_5 = [19000, 41000]$, $u_6 = [41000, 63000]$, $u_7 = [63000, 85000]$. If we take into account the fact that forecasting with fuzzy time series exhibits the least average error, it is necessary to find the middle points of the intervals: $u_m^1 = -58000$, $u_m^2 = -36000$, $u_m^3 = -14000$, $u_m^4 = 8000$, $u_m^5 = 30000$, $u_m^6 = 52000$, $u_m^7 = 74000$.

Fuzzification of variations for each year is made using the formula:

$$\mu_{A^{mn}}(u_i) = \frac{1}{1 + [C * (U - u_m^i)]^2} \tag{1}$$

where A^{mn} is the fuzzy set of the corresponding variation for the years $[1988, 2009]$ (the last two digits of the year), C is a constant (in our case $C = 0.0001$), U is the variation taken from table 1 and u_m^i is the middle point of the corresponding defined interval u_i. Fuzzy sets are defined on the universal set U. In this case "the variation in total population" is a linguistic variable that assumes the following linguistic values:

A_1 =(very low level population growth (VLLPG));
A_2 =(low level population growth(LLPG));
A_3 =(unchanged population growth (CPG));
A_4 = (moderate population growth (MPG));
A_5 =(normal-level population growth (NLPG));
A_6 = (high-level population growth (HLPG));
A_7 =(very high-level population growth (VHLPG)).

To every linguistic value here corresponds a fuzzy variable which, according to a certain rule, is assigned against a corresponding fuzzy set determining the meaning of this variable.

Years	Total population (thousand person)	Variation (thousand person)
1988	22769	
1989	22852	83
1990	22866	14
1991	22826	-40
1992	22797	-29
1993	22769	-28
1994	22739	-30
1995	22693	-46
1996	22628	-65
1997	22562	-66
1998	22516	-46
1999	22481	-35
2000	22452	-29
2001	22428	-24
2002	22404	-24
2003	22380	-24
2004	22356	-24
2005	22330	-26
2006	22304	-26
2007	22276	-28
2008	22247	-29
2009	22215	-32

Table 1: The dynamics and variation of population growth for the period [1988-2009].

If the value of variable U in formula (1) is accepted as the middle point of the corresponding interval, the fuzzy set A_i $(i = 1, ..., 7)$ will be represented as in Figure 1.

For evaluating the effectiveness of the application to handling demographic forecasting problems, the total population has been calculated over a certain time period. Results obtained from the retrospective analysis of population forecasting are given in table 2.

The essence of the conducted experiment consists of the following:

1. the dynamics of the total population for the examined period is considered to be unknown;

2. with the aid of the proposed methodology, the total population size was forecasted for every year selected from the time interval [1988,2009] based on the changes in population growth rates of the previous years;

3. in order to test the model's degree of adequacy, the observed (real) dynamics of the total population over the time interval [1988,2009] was compared with the corresponding results of the model application and the consequent model error computed.

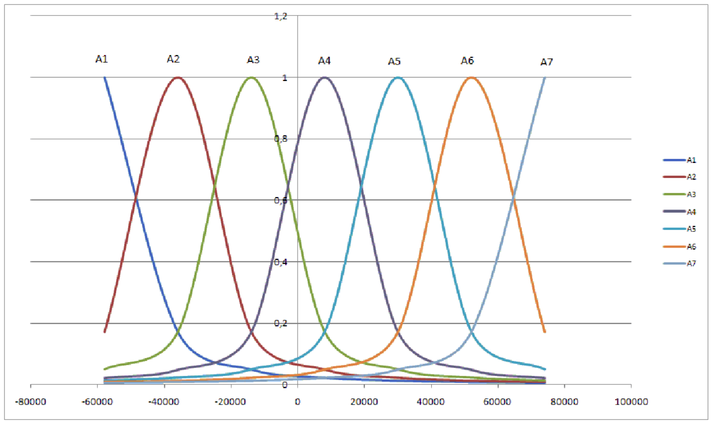

Figure 1: Membership function of values of fuzzy sets of linguistic variable "variation in total population".

The error of the proposed method is computed by the following formula:

$$\delta(t) = \frac{V^t_{obsev} - V^t_{forec}}{N^t_{obsev}} \cdot 100\% \tag{2}$$

where V^t_{obsev} is the variation in total population for the t-th year; V^t_{forec} is the variation in total population for the t-th year; N^t_{obsev} is the observed total population for the t-th year, $1988 \leq t \leq 2009$. The average error coresponding to the results shown in table 2 is 0.000522.

The comparative analysis of the observed and forecasted data and the consequent error of the approximated method have confirmed the high efficacy of the model and made us believe that its application to demographic forecasting would serve our purpose.

Figure 2, depicting graphically the dynamics of the observed (actual) and forecasted total population sizes, displays the two data's remarkable closeness that, in its turn, necessitates continuing the researches conducted in this direction.

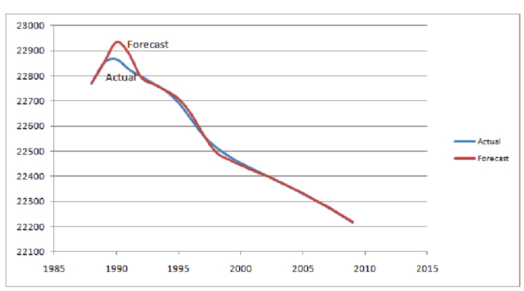

Figure 2: Results of the retrospective analysis of the total population forecasting.

Years	Actual		Forecasted		Error (%)
	Total population (thousand person)	Variation (thousand person)	Total population (thousand person)	Variation (thousand person)	
1988	22769		22771		
1989	22852	83	22852	81	0.000088
1990	22866	14	22935	83	0.003017
1991	22826	-40	22889	-46	0.000262
1992	22797	-29	22794	-95	0.002895
1993	22769	-28	22767	-27	0.000043
1994	22739	-30	22741	-26	0.000175
1995	22693	-46	22709	-32	0.000616
1996	22628	-65	22649	-60	0.000220
1997	22562	-66	22566	-83	0.000753
1998	22516	-46	22496	-70	0.001065
1999	22481	-35	22467	-29	0.000266
2000	22452	-29	22444	-23	0.000267
2001	22428	-24	22422	-22	0.000089
2002	22404	-24	22403	-19	0.000223
2003	22380	-24	22380	-23	0.000044
2004	22356	-24	22356	-24	0
2005	22330	-26	22332	-24	0.000089
2006	22304	-26	22304	-28	0.000089
2007	22276	-28	22278	-26	0.000089
2008	22247	-29	22248	-30	0.000044
2009	22215	-32	22218	-30	0.00009

Table 2: Results obtained from the retrospective analysis of population forecasting

5 Conclusion

The methodology proposed in this article enables us to forecast demographic processes on the basis of fuzzy time series. A peculiar trait of the methodology consists of its capability to forecast the required indicator by utilizing incomplete, fuzzy input data. The described approach, by entering the dynamics of the total population until the previous year into an experimental base, helps make forecast calculations for any distant perspective. This, in its turn, allows us to take into account the trend of previous population growth rates and as a result achieve more accurate forecasts.

As it is evident, although exploring the dynamics of total population provides us with its primary, aggregate characteristics, it does not mirror its reproduction process or the structure of population. Therefore, in the future, the range of forecasted population characteristics is intended to be extended by including other population indicators such as age structure, new-borns, the dead, migration etc.

References

[1] Abbasov, A. M. and Mamedova, M. H., *Application of fuzzy time series to population forecasting*, Proceedings of 8th Symposion on Information Technology in Urban and Spatial Planning, Vienna University of Technology, February 25 - March 1, 545-552, 2003.

[2] Chen, S. M., *Forecasting enrollments based on high-order fuzzy time series*, Cybernetics and Systems: An International Journal, **33**, (2002), 1-16.

[3] Chen, S. M. and Hsu, C. C., *A New Method to Forecast Enrollments Using Fuzzy Time Series*, International Journal of Applied Science and Engineering, **2** (3), (2004), 234-244.

[4] Pandey, A. K., Sinha, A. K. and Srivastava, V. K., *A Comparative Study of Neural-Network and Fuzzy Time Series Forecasting Techniques Case Study: Wheat Production Forecasting*, International Journal of Computer Science and Network Security, **8** (9), (2008), 382-387.

[5] Song, Q. and Chissom, B. S., *Fuzzy time series and its models*, Fuzzy Sets and Systems, **54** (1993),269-277.

[6] Song, Q. and Chissom, B. S., *Forecasting enrollments with fuzzy time series Part I*, Fuzzy Sets and Systems, **54** (1993),1-9.

[7] Song, Q. and Chissom, B. S., *Forecasting enrollments with fuzzy time series Part II*, Fuzzy Sets and Systems, **62** (1994),1-8.

[8] Stevenson, M. and Porter, J. E., *Fuzzy Time Series Forecasting Using Percentage Change as the Universe of Discourse*, World Academy of Science, Engineering and Technology, **55** (2009), 154-157.

[9] Zadeh, L. A., *Fuzzy sets*, Information and Control **8** (3)(1965), 338-353.

[10] Zadeh, L. A., *Fuzzy Sets, Fuzzy Logic, Fuzzy Systems*, World Scientific Press, 1996.