International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 11 Issue 2 ǁ February 2023 ǁ PP. 44-55

Forecasting of Demographic and Economic Variables in India using Fuzzy Time Series Model

M. N. Megeri[1] and Bheemanna[2]
[1] *Associate Professor, Department of Statistics, Karnatak University's Karnatak Arts College, Dharwad, 580001, Karnataka, India.*
[2]*Research Scholar, Department of Statistics, Karnatak University's Karnatak Arts College, Dharwad, 580001, Karnataka, India.*

Abstract: *Forecasting of demographic and economic variables is an essential component of research that will help society and the government plan for the best or worst in the future. The data on Demographic variables are collected from Census of India and economic variables are gathered from Economic survey of India from 1971 to 2020. The goal of this study is to predict the magnitude of all the selected variables for next ten years using the Abbasova and Mamedova (AM) model, in which the parameters are evaluated through the AM model algorithms to obtain appropriate results for each set of data. Findings of the study reveals that except for GDP all the selected variables were AM model fitted well and comparison shows rural population is best fitted as compared to the entire demographic and economic variable by using Mean Absolute Percentage Error (MAPE).*
Keywords: *Demographic variable, Economic variable, Fuzzy time series, Abbasov-Mamedova and MAPE.*

Date of Submission: 24-01-2023 Date of acceptance: 07-02-2023

I. INTRODUCTION

In 2019, the world's population had a life expectancy of 72.6 years, an increase of more than 8 years since 1990. According to World Population Prospect (2019), further advancements in survival are expected to result in a global average life expectancy of roughly 77.1 years in 2050.Current age structures will drive two-thirds of the anticipated global population growth until 2050. It would happen even if childbirth in today's high-fertility countries dropped to roughly two births per woman during her lifetime. Global fertility falls from just under 2.5 births per woman in 2019 to roughly 2.2 in 2050, and then to 1.9 in 2100, according to the medium-variant prediction. China, which has 1.43 billion people, and India, which has 1.37 billion, has long been the world's two most populated countries.

Forecasting is the process of creating future predictions based on a combination of past experiences, knowledge, and analysis of connected issues. It is regarded as the foundation process, the initial stage in the development of policies and objectives for both organizations and government agencies. Forecasting has gotten a lot of attention from scientists because of its importance in so many fields. Time series and regression models play essential roles in statistical forecasting, but they have a number of drawbacks in reality. The issues of predicting have not yet been completely solved, despite various discussions in the literature. The main technique of statistics is to search for principles and rules to establish an appropriate forecasting model based on previous data. When there are irregular changes or the time series is non-stationary, a regression model Galton (1888); Pearson (1896) requires several inadequate assumptions, whereas a time series model, such as ARIMA Box and Jenkins (1976), performs badly. Many studies have suggested various models to solve the disadvantages of these two models, including Zecchin et al. (2011); Wang and Fu (2006); Wang et al. (2001); Ren et al. (2016); Gupta and Wang (2010); Zhu and Wang (2010); Park (2010); Teo et al. (2001); Ghazali et al. (2009). These approaches are significant additions to the forecasting problem since they have produced positive results in the data sets studied. However, we were unable to attain optimal results in all situations.

The fuzzy time series (FTS) established by Song and Chissom (1993) can close the gap indicated above, based on Zadeh (1965) fuzzy theory. FTS has since been studied and proven to be more efficient than traditional statistical techniques Song and Chissom (1993); Tseng and Tzeng (2002)). Abbasov and Mamedova (AM) proposed a model to anticipate the population of Azerbaijan using data fluctuations indicated by language level (Abbasov and Mamedova, 2003). The AM model has been used in a variety of applications due to its superior performance for certain types of forecasting issues; for example, Sasu used the AM model to forecast Romanian population Sasu (2010). Other important FTS research includes the models in Chen (2004), Huarng (2001), and Singh (2008).Ha Che-Ngoc, Tai Vo-Van, Quoc-Chanh Huynh-Le, Vu Ho, Thao Nguyen-Trang., (2018)developed an enhanced Fuzzy Time Series forecasting model based on approaches for determining the appropriate parameters for each data set in the AM model. All of the approaches above, however, solely rely on previous fuzziness data and do not forecast. Furthermore, the model parameters are not thoroughly investigated in order to determine the best values for each data set. Only in a few situations is one model considered superior to the others. As a result, no one model can be regarded optimal in all situations.

To address the previous problem, this article presents techniques for identifying appropriate parameters in the AM model. Specifically, w, the number of elements in the data set utilized as previous knowledge to predict the data, is determined by the value of the partial autocorrelation function (PACF), and n, the number of fuzzy sets, is determined by an indicator that evaluates the compactness of the divided intervals. Following the determination of acceptable w and n, the best choice of C is found using an efficient method in order to minimize forecasting error. The demographic and economic variables illustrate the suggested theories in detail and show that this approach could improve forecasting accuracy.

The rest of this paper is structured as follows. Section 2 describes the AM model's technique and parameters, as well as some associated definitions and for the AM model, in which the appropriate parameters are obtained through methods and algorithms. Section 3 deals with interpretation, Section 4 with conclusions, and Section 5 with references.

II. METHODS AND MATERIALS

The Indian populations were gathered from Census, Registrar General India; Gross Domestic Product (GDP) data was collected from Economic Survey of India. The Birth Rate, Death Rate, Infant Mortality Rate, Total Fertility Rate, Under 5 Mortality Rate, Life Expectancy at Birth, and Age Dependency Ratio were obtained from Sample Registration System (SRS) publication. All the selected variables data from 1971 to 2020.

2.1 Related Definitions and Abbasov-MamadovaModel:

Definition 1: Let U be the discourse universe, with U = $u_1, u_2, ..., u_m$. The following is the definition of a fuzzy set A of U:

A = $\{\mu_A(u_1)/ u_1, \mu_A(u_2)/ u_2, \mu_A(u_2)/ u_2, ..., \mu_A(u_m)/ u_m\}$ (1)

Where$\mu_A(u_1)$ is membership function of A, $\mu_A(u_i) : U \epsilon [0, 1]$, $\mu_A(u_i)$ represent the degree of membership of u_i in A, $\mu_A(u_i)$ [0, 1] and $1 \leq i \leq m$.

The membership degree or grade of membership is the value of the membership function of a given u, $\mu_A(u)$. If the membership grade equals one, u belongs entirely to the fuzzy set. If the membership grade is zero, u does not belong to the set. If the grade of membership is between 0 and 1, u is a fuzzy set partial member.

$$\mu_A(u) = \begin{cases} = 1 & u \text{ is a full member of } A \\ \in (0, 1) & u \text{ is a partial member of } A \\ = 0 & u \text{ is not member of } A \end{cases}$$

In the above definition, a fuzzy set A in U is defined by a membership function $\mu_A(u)$. The membership function A can be defined in a variety of ways $\mu_A(u)$. Some common types of membership functions include trapezoidal membership function, triangular membership function, Gaussian membership function, and so on. Following that, we look at a specific situation of Definition 1 in which the universe is a time series and provide the definition proposed by (Song and Chissom (1993)) as follows.

Definition 2: Let X(t), a subset of real numbers (t = 1, 2,...), be the universe of discourse through which the fuzzy sets f_i (t) are defined. If it is a collection of $f_1(t)$, $f_2(t)$,..,then,F(t) is considered a Fuzzy Time Series (FTS) defined on X (t).

Definition 3: Given the actual data X_i and the prediction value $\{\widehat{X_i}\}$, i =1, 2,...,m, respectively, we have the following popular criteria to examine the established model:

2.2 Mean Absolute Percentage Error:The mean absolute percentage error (MAPE), also known as the mean absolute percentage deviation, is used to assess the accuracy of a forecast system. It is calculated as the average absolute percent error minus actual values divided by real values for each time period and expressed as a percentage.

$$\mathbf{MAPE} = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{\left|\widehat{X_i} - X_i\right|}{X_i} * 100\right)$$

Where $\widehat{X_i}$ denote estimated value, X_i denote actual values and m denote the number of observation.

2.3 Abbasov-Mamedova (AM) Model:

Let$X_t$value of the variable corresponds to year t = 1, 2,..., m based on historical data. The six steps of the AM model are as follows.

Step 1: Calculate the variation V_t between the current and previous value using formula (4). Then, using formula(5), define the universal set U.

$V_t = X_t - X_{t-1}$(4)

$$U = [V_{min} - D_1,\ V_{max} + D_2] \quad (5)$$

Where V_{min} denotes the smallest variation, V_{max} is the largest variation, and D_1 and D_2 arepositive numbers.

Step 2: Divide the universal set U into n equal-length intervals u_i, i= 1, 2,...,n, with at least one variation value in each interval u_i.Then, for each interval, determine the u_m^i middle points.
Step 3: On the universal set U, define the fuzzy set A_i, i = 1, 2,...,n, by using formula:

$$\mu_{A_i}(u) = \frac{1}{1+[C \times (u-u_m^i)]^2} \text{ (6)},$$

Where u is a generic element of the universal set U, $u_m^i$is the midpoint of the corresponding interval u_i, i = 1, 2,..., n, and C is a constant.
Step 4: Using formula (6), convert the input data time-point variations into fuzzy values.
Step 5: Choose an integer w, l>w>1, where l is the number of years prior to the current year that will be used in the experiment. Based on the options selected, we create an operation matrix using the w and Mamdani inference systems. $O_w(t)$ of size i × j (where i is the number of rows, which corresponds to the year sequence t-2, t-3,..., t-w, and j is the number of columns, a number of variation intervals which corresponds to the year sequence t-2, t-3,..., t-w) a criteria matrix K(t) of dimension 1 × j (row matrix matching to the year's fuzzy variation in total population t-1). The relationship matrix R(t) is then computed as follows.

$$R(t)\,[i, j] = O_w[i, j] \cap K(t)\,[\,j]$$

Or

$$R(t) = O_w\,(t) \otimes K\,(t) = \begin{bmatrix} R_{11} & R_{12} \cdots & R_{1j} \\ R_{21} & R_{22} \cdots & R_{2j} \\ R_{i1} & R_{i2} \cdots & R_{ij} \end{bmatrix}$$

Where$O_w(t)$ denotes the operation matrix, K(t) is the criteria matrix, and ⊗ denotes the minimum operator(∩). In a fuzzy form, define F(t), the fuzzy forecasting of variations for the year t, as shown below.

$$F\,(t) = [\max\,(R_{11}, \dots, R_{i1})\,, \dots, \max\,(R_{1j}\,, \dots, R_{ij}\,)] \\ = [\mu_{A_1}(V_t), \mu_{A_2}(V_2) \dots, \mu_{A_m}(V_t)].$$

Step 6: Defuzzify the data acquired in the fifth phase using formula (7).

$$V(t) = \left[\frac{\sum_{i=1}^{m} \mu_{A_i}(V_t) \times u_m^i}{\sum_{i=1}^{m} \mu_{A_i}(V_t)}\right] \quad (7)$$

Where $\mu_{A_i}(V_t)$ is the value of the membership function of the predicted variation in interval i, V(t) is the defuzzified prediction variation.
The following formula can be used to calculate the forecast value X(t) for year t.

$$X_t = X\,(t-1) + V\,(t), \quad (8)$$

Where X(t-1) is the actual value for the year t-1 and V (t) is the variation for the year t.
The number of equal-length intervals n, the positive integer w, and the constant C all have an effect on the projected result in the AM model. However, in the studies of Abbasov and Mamedova (2003) and Sasu (2010), these parameters were only identified based on personal experiences. As a result, this approach is unsuitable for dealing with many kinds of time series. Song and Chissom conducted a poll on the specific data for w and determined that w = 2 are the best. They also came to the conclusion that the predictions were the result of an improved fuzzy time series forecasting model and were better when we used a less sophisticated model with a smaller w (Song and Chissom , 1993).
However, because this conclusion is based on only a few unique polls, it lacks universality. In fact, w is the number of prior times that have a significant impact on the current value of the time series (it is similar to the partial autocorrelation p in the autoregressive integrated moving average, ARIMA). As a result, using a model with a set value of w for all types of time series is not feasible. When working with monthly or quarterly data, for example, w = 7 is considered an inappropriate parameter. According to the above statements, it is certain that the AM model's forecasting ability can be greatly enhanced if its parameters are determined in an appropriate way. To overcome the restrictions indicated above, we used a method called MAM (Modified Abbasov-Mamedova model) in this section, which can identify the parameters n, w, and C in a reasonable manner. The following are the specifics of the proposed method:

2.4 Determines the class interval:

The number of class interval is given by sturges rule as

$$n = 1 + 3.332 * log_{10}(N), \quad (9)$$

Where n denotes the number of the class interval, N denotes the number of observation.

2.5 Determine w:

In the literature, the parameters are chosen based on prior experience. Although Song and Chissom conducted a survey on specific datasets and concluded that w = 2 is the best, their finding is based on only a few unique surveys and lacks generality. Based on the Partial Autocorrelation Function (PACF) of the time series, an effective approach for determining w is proposed here.

If $\emptyset_{kk}$ is the PACF at lag k (k = 1, 2,...), then $\emptyset_{kk}$ can be found using the recursive formula (Durbin (1960)) as follows:

$$\emptyset_{p+1,j} = \emptyset_{p,j} - \emptyset_{p+1,p+1}\emptyset_{p,p-j+1,}$$

$$\emptyset_{p+1,p+1} = \frac{r_{p+1} - \sum_{j=1}^{p} \emptyset_{p,j} r_{p+1-j}}{1 - \sum_{j=1}^{p} \emptyset_{p,j}\, r_j}$$

Where r_k is the autocorrelation function (ACF) at lag k, $\emptyset_{1,1} = r_1$.

At lag k, the PACF evaluates only the direct correlation between v_t and v_{t-k}, while the linear dependences between the intermediate variables are ignored. As a result, it can exactly indicate the number of past years on which the current year is based. We may calculate the appropriate w based on this result.

It is important to note that when PACF has the highest value at lag 1, w is assumed to be 2, allowing the AM model conditions to be fitted. In reality, the PACF and appropriate w for the fuzzy time series model can be calculated using statistical tools such as R programming with store in the AnalyzeTS package.

2.6 Determine the Constant C:

Given an AM model with certain parameters w and n, the constant C has a significant influence on the value of $\mu_{Ai}(u_i)$ as well as the predicted result. Previous research (Abbasov and Mamedova (2003); Sasu (2010)) did not provide recommendations on how to calculate an appropriate C for each unique dataset. This subsection presents an algorithm for Determining Optimal C (DOC) using the Modified Abbasov and Mamedova (MAM) approach and the five steps mentioned below:

Step 1: Produce an integer k (k > 499), a very tiny positive value, where k is the number of points split every iteration and ε is the error.

Step 2: When t = 0, assign the following values: $a^{(0)} = 0$, $b^{(0)} = 1$, $C^{(0)} = \frac{1}{2}$, and n(0) = 1.

Step 3: Calculate the values for t = i, i ≥ 1

$$a^{(t)} = a^{(t-1)} + \left[n^{(t-1)} - 1\right] \Delta C^{(t-1)}$$

$$b^{(t)} = a^{(t-1)} + \left[n^{(t-1)} + 1\right] \Delta C^{(t-1)}$$

$$\Delta C^{(t)} = \frac{b^{(t)} - a^{(t)}}{k}$$

$$\text{If } a = 0 \text{ and } b = 1, \text{then } C_i^{(t)} = a^{(t)} + i\Delta C^{(t)}, t = 1,2,\ldots,k-1.$$

$$\text{If } a = 0 \text{ and } b \neq 1, \text{then } C_i^{(t)} = a^{(t)} + i\Delta C^{(t)}, t = 1,2,\ldots,k.$$

$$\text{If } a \neq 0 \text{ and } b = 1, \text{then } C_i^{(t)} = a^{(t)} + i\Delta C^{(t)}, t = 1,2,\ldots,k-1.$$

$$\text{If } a \neq 0 \text{ and } b \neq 1, \text{then } C_i^{(t)} = a^{(t)} + i\Delta C^{(t)}, t = 1,2,\ldots,k.$$

Step 4: Run the Abbasov-Mamedova model with all of the ${C_i}^{(t)}$ values from Step 3. Find ${C_n}^{(t)}$for which the criteria is currently the best.

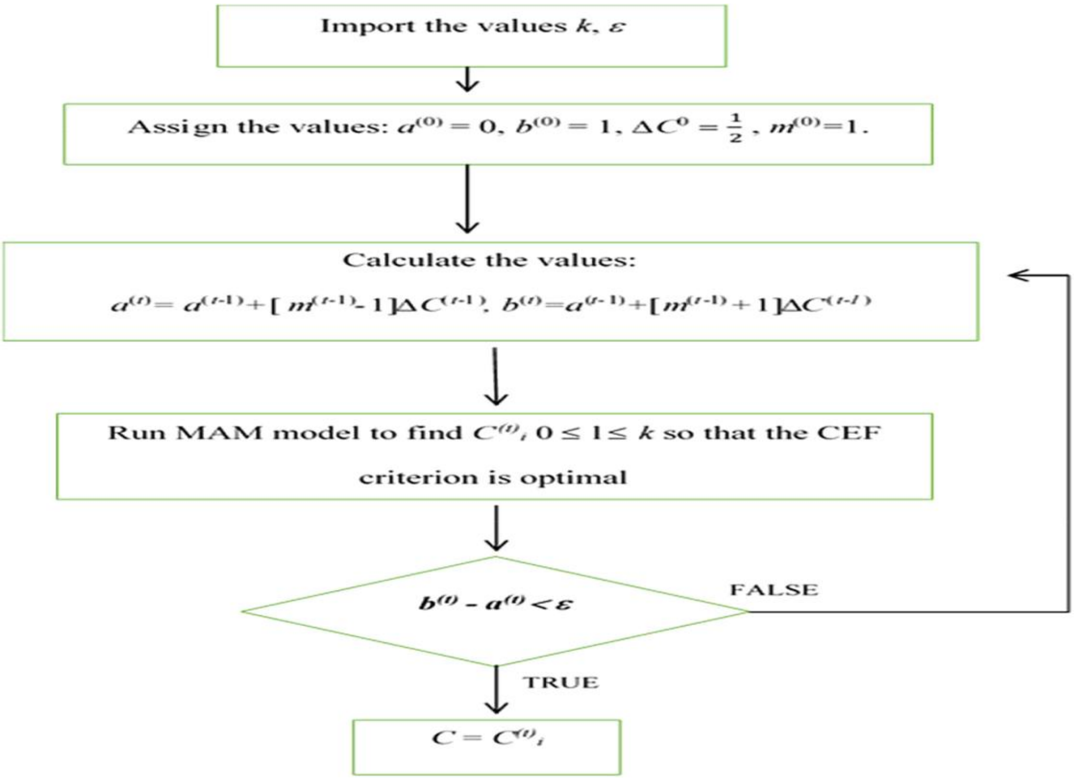


Figure 2. Algorithm to determine C

Step 5: With the new n, execute Steps 3 and 4 to determine

$$C = C_n^{(m)} = a^{(t)} + n\, \Delta C^{(m)} \text{until} b^{(t)} - a^{(t)} < \varepsilon$$

Keep in mind,

(i) (k+1) C values are examined in each iteration. k = 1000 is used in the numerical examples in this study.

(ii) $\varepsilon$is an extremely small number that was chosen at arbitrarily. The smaller ε is, the more iterations and computing time are needed. In fact, depending on the value of, the best value of C can be found with an acceptable error ε. In numerical examples, the value ε = 0.000001 is used.

(iii) Criteria are used to evaluate the forecasting model (Criteria Evaluate Forecasting). In this study, we use proven models using MAPE.

The DOC algorithm is illustrated in figure 2.

III. RESULT AND DISCUSSION

The study investigates selected demographic and economic indicators from 1971 through 2020. Total population, Urban population, and Rural population, Crude birth rate, Crude death rate, Infant mortality rate, Under 5 mortality rate, Total fertility rate, and Life expectancy at birth are demographic variables that have been chosen for analysis, as have GDP and age dependency ratio are economic variables. According to the Abbasov and Mamedova model (2003), the detailed technique is broken down into six steps.

Step 1: Table 2 shows the annual total populations from 1971 to 2020, as well as variation in each year. The difference between the population values in the current year and the prior year is the variation for the current year. For Total population, variation for 2010 is equal to1190152847-1170720342 = 19583620. To define the universal set U, first find the smallest and biggest variation total population over period 1971- 2020, then choose appropriate non-negative numbers D_1, D_2 to ensure the smoothness of the interval's bounds. The universal set U is thus defined as:U = [V_{min}-D_1, V_{max} + D_2], where V_{min} = 13690146 is the smallest variation value, V_{max} = 20943652is the largest variation value, D_1 = 0, and D_2 = 0. As a result, the universal set U is specified as [13690146, 20943652] for Total population. As similarly , all the variables of the universal set U is specified as [3358042, 11226082], [9717570, 10332103], [-2, 0.70], [-1.40, 2], [-11, 14], [-0.30, 0.10], [-5.60, -2.20], [0.1, 1.2], [3361,1872933], and [-1, 0] for total population, Urban population, Rural population, Birth rate, Death rate, IMR, TFR, Under 5 mortality, GDP and Age Dependency ratio are respectively.

Step 2: Fuzzy sets are defined using the universal set U. In this example, "the variation in total population" is a language variable with the following linguistic values: A1 denotes (very low level population growth (VLLPG); A2 denotes (low level population growth (LLPG)); A3 denotes (changeless population growth (CPG)); A4 denotes (moderate population growth (MPG)); A5 denotes (normal-level population growth (NLPG)); A6 denotes (high-level population growth (HLPG)); A7 denotes (very high-level population growth (VHLPG)); Every linguistic value corresponds to a fuzzy variable, which is assigned to a corresponding fuzzy set that determines the meaning of this variable based on a set of rules.The fuzzy variable < VLLPG, [13690146, 14726361], A_1 >, for example, gives the linguistic value "very-low-level population growth," where A_1 is a fuzzy set defined on the domain [13690146, 14726361] of the universal set U. The universal set U must be subdivided into many intervals of equal length. Then n value is computed from equation (9). The appropriate number of equal intervals from equation (9) is 7. For demographic and economic variables, this set U is partitioned into seven equal length and middle values.

The intervals' middle points are determined as follows:

$u_m^1 = 14208254, u_m^2 = 15244469, u_m^3 = 16280684, u_m^4 = 17316899, u_m^5 = 18353114, u_m^6 = 19389329, u_m^7 = 20425544$. Similar linguistic variable with grading is used for all the selected variables.

The PACF computed and presented in Figure 3 is used as the criterion to determine the appropriate w. According to Figure 3, for w, PACF reaches its maximum value at lag 1, and then the w is chosen as 2. In summary, the Abbasov-Mamedova model employs n = 7 for all the variables from equation (9) and w = 2, 2, 2, 8, 3, 16, 6, 2, 10, 2, 8 based on PACF for Total population, Urban population, Rural population, Birth rate, Death rate, IMR, TFR, Under 5 mortality, Life expectancy at birth, GDP, and Age Dependency ratio respectively.

Table 1: The universal set U must be divided into 7 equal intervals and middle value for demographic and economic variables.

Variables		u_1	u_2	u_3	u_4	u_5	u_6	u_7
Total Population	L	13690146	14726361	15762576	16798791	17835007	18871222	19907437
	U	14726361	15762576	16798791	17835007	18871222	19907437	20943652
	M	14208254	15244469	16280684	17316899	18353114	19389329	20425544
Urban population	L	3358042	4482048	5606053	6730059	7854065	8978071	10102076
	U	4482048	5606053	6730059	7854065	8978071	10102076	11226082
	M	3920045	5044051	6168056	7292062	8416068	9540073	10664079
Rural population	L	9717570	9805360	9893151	9980941	10068732	10156522	10244313
	U	9805360	9893151	9980941	10068732	10156522	10244313	10332103
	M	9761465	9849256	9937046	10024837	10112627	10200417	10288208
Birth rate	L	-20.00	-16.14	-12.29	-8.43	-4.57	-0.71	3.14
	U	-16.14	-12.29	-8.43	-4.57	-0.71	3.14	7.00
	M	-18.07	-14.21	-10.36	-6.50	-2.64	1.21	5.07
Death rate	L	-14.00	-9.14	-4.29	0.57	5.43	10.29	15.14
	U	-9.14	-4.29	0.57	5.43	10.29	15.14	20.00
	M	-11.57	-6.71	-1.86	3.00	7.86	12.71	17.57
IMR	L	-110.00	-74.29	-38.57	-2.86	32.86	68.57	104.29
	U	-74.29	-38.57	-2.86	32.86	68.57	104.29	140.00
	M	-92.14	-56.43	-20.71	15.00	50.71	86.43	122.14
TFR	L	-3.00	-2.43	-1.86	-1.29	-0.71	-0.14	0.43
	U	-2.43	-1.86	-1.29	-0.71	-0.14	0.43	1.00
	M	-2.71	-2.14	-1.57	-1.00	-0.43	0.14	0.71
Under 5 mortality	L	-56.00	-51.14	-46.29	-41.43	-36.57	-31.71	-26.86
	U	-51.14	-46.29	-41.43	-36.57	-31.71	-26.86	-22.00
	M	-53.57	-48.71	-43.86	-39.00	-34.14	-29.29	-24.43
Life expectancy at birth	L	0.1	0.26	0.41	0.57	0.72	0.88	1.04
	U	0.26	0.41	0.57	0.72	0.88	1.04	1.20
	M	0.17	0.33	0.49	0.65	0.80	0.96	1.12

GDP	L	3361.00	270442.70	537524.40	804606.10	1071687.90	1338769.60	1605851.30
	U	270442.70	537524.40	804606.10	1071687.90	1338769.60	1605851.30	1872933.00
	M	136901.90	403983.60	671065.30	938147.00	1205228.70	1472310.40	1739392.10
Age Dependency Ratio	L	-1.00	-0.86	-0.71	-0.57	-0.43	-0.29	-0.14
	U	-0.86	-0.71	-0.57	-0.43	-0.29	-0.14	0.00
	M	-0.93	-0.79	-0.64	-0.50	-0.36	-0.21	-0.07

Note: U is Upper value, L is Lower value, M is middle value

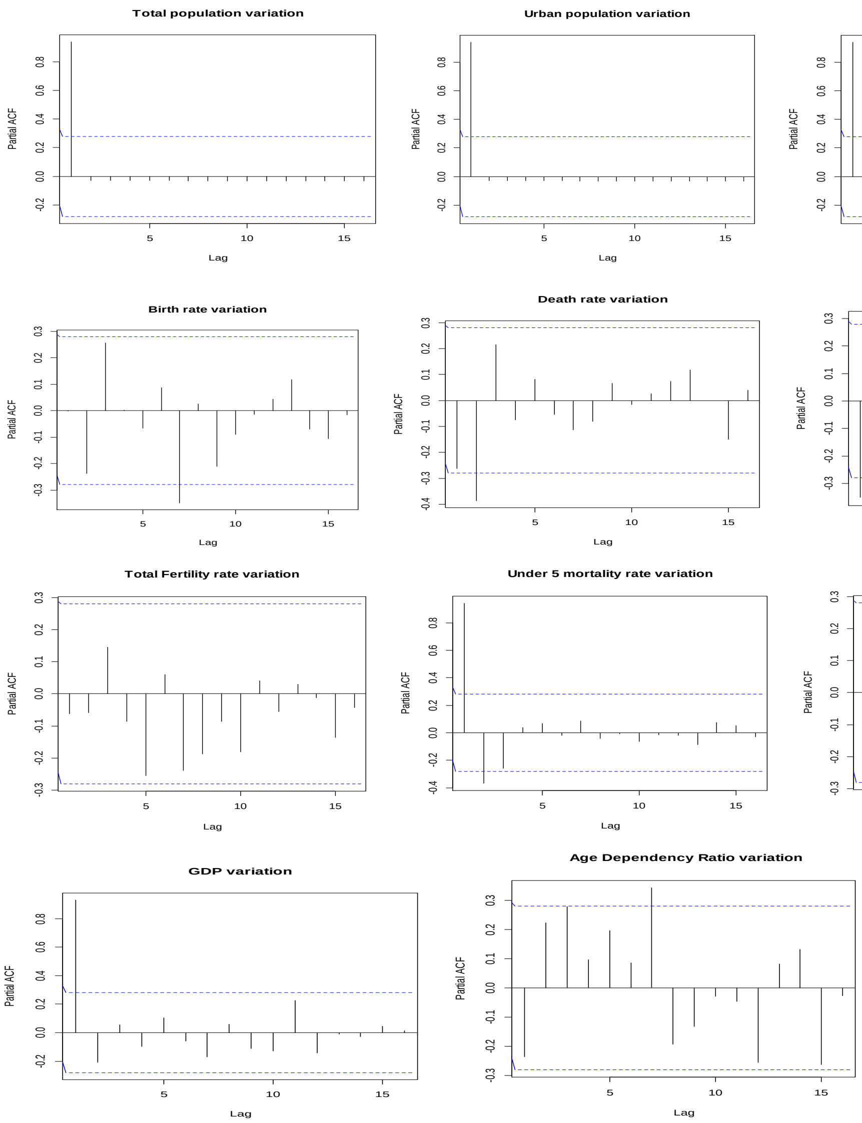


Figure3: PACF according to w

With n = 7 and w = 2, performing the DOC algorithm, the optimal value of C is 0.00002563264 for total population. As similarly w = 2, 2, 8, 3, 16, 6, 2, 10, 2, 8 performing the DOC algorithm, the optimal value of C is 0.000005924608, 0.000302948864, 0.007756070912, 0.507482469248, 0.065993775232, 0.000000027392, 0.000000027392, 0.999999999872, 0.999996037632, and 0.000000208256 for Urban population, Rural population, Birth rate, Death rate, IMR, TFR, Under 5 mortality, Life expectancy at birth, GDP, and Age Dependency ratio respectively.

Steps 3 and 4: On the universal set U, define the fuzzy sets A_1, A_2,..., A_7, and convert the input data into fuzzy values using formula (6). Figure 4 depicts an exemplary evolution of the continuous membershipfunctions of fuzzy sets A_i. Table 1 shows the fuzzification results for all years with the last five digits for the sake of brevity.

Specific Fuzzy Sets of seven levels :
$A1$=[0.169515 0.001684 0.000386 0.000167 0.000092 0.000059 0.000040]
$A2$=[0.001161 0.115540 0.001763 0.000394 0.000169 0.000093 0.000059]
$A3$=[0.000314 0.001119 0.083173 0.001848 0.000403 0.000171 0.000094]
$A4$=[0.000143 0.000308 0.001078 0.062485 0.001939 0.000412 0.000174]
$A5$=[0.000094 0.000171 0.000403 0.001848 0.083173 0.001119 0.000314]
$A6$=[0.000059 0.000093 0.000169 0.000394 0.001763 0.115540 0.001161]
$A7$=[0.000040 0.000059 0.000092 0.000167 0.000386 0.001684 0.169515]

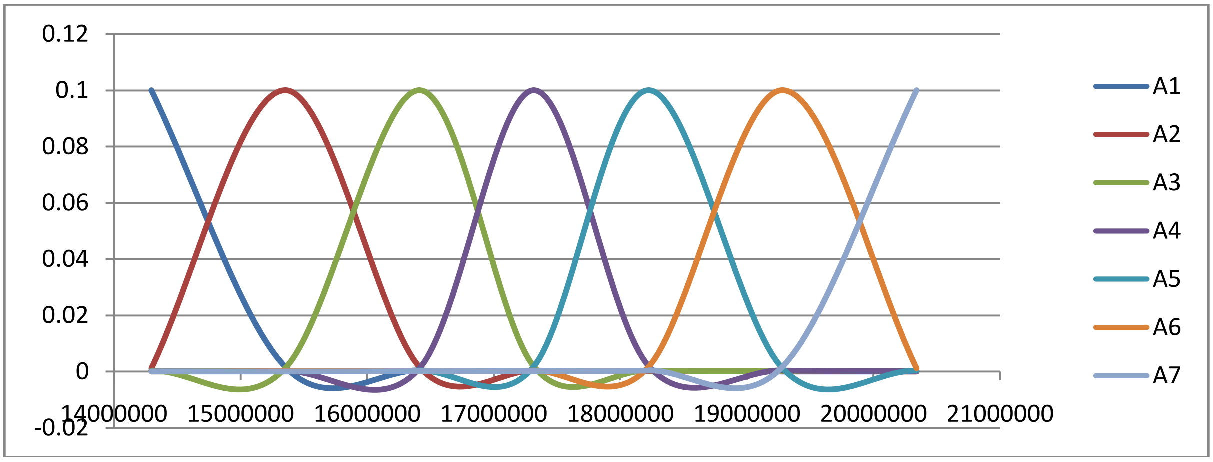


Figure 4: Gaussian Membership functions of values of fuzzy sets of linguistic variable "variation in total population".

Step 5: Using the min - max operator to forecast the population in 2021, we get the following results:
O^2 (2021) = [0.000035 0.000049 0.000075 0.000126 0.000256 0.000772 0.011174]
K (2021) = [0.000034 0.000047 0.000070 0.000116 0.000227 0.000630 0.005638]
R (2021) = [0.000034 0.0000470.000070 0.000116 0.0002270.000630 0.005638]
As a result, the fuzzy forecasting of variation for the year 2021 is
F (2021) = [0.000034 0.000047 0.000070 0.000116 0.000227 0.000630 0.005638]
Step 6: Finally, use formula (7) to compute the variations in 2021

$$\mathrm{V}\ (2021) = \frac{0.000034 * 14208254 + 0.000047 * 15244469 + \cdots + 0.005638 * 20425544}{0.000034 + 0.000047 + \cdots + 0.005638}$$
$$= 20096652$$

Hence, the forecasting population in 2021 is
X(2021) = X(2020) + V(2021)
= 1392789204+ 20096652
=1412885856

Repeat for the remaining items; the forecasted results are shown in Table 3, 4 and Figure 5.The above procedure is same for all the variables.

Table 3: Prediction of the demographic variables from 2021 to 2030.

Year	Total population	Urban Population	Rural Population	Birth Rate	Death Rate	IMR	TFR	Under 5Mortality	Life Expectancy
2021	1412885856	483191692	929670598	19.1	5.9	29	1.9	29.2	70.5
2022	1432982508	493475451	939459928	18.5	5.8	28	1.8	26.6	71.2
2023	1453080694	503692225	949249126	17.9	5.7	27	1.7	24.0	71.8
2024	1473179558	513814205	959038267	17.3	5.6	26	1.6	21.3	72.5
2025	1493281060	523769168	968827182	16.7	5.5	25	1.5	18.5	73.1
2026	1513384746	533485756	978615911	16.0	5.4	24	1.4	15.6	73.8
2027	1533493380	543004192	988404217	15.4	5.3	23	1.3	12.7	74.4
2028	1553607416	552498877	998192062	14.7	5.2	22	1.2	9.7	75.1
2029	1573731127	562002776	1007979082	14.1	5.1	21	1.1	6.8	75.7

2030	1593866739	571505701	1017765085	13.4	5.0	20	1.0	3.8	76.4

Table 4: Forecast of the Economic variables from 2021 to 2030.

Year	GDP	Age Dependency Ratio
2021	21770780	48.5
2022	23174760	48.0
2023	24614016	47.5
2024	26055055	47.0
2025	27519924	46.5
2026	28986677	46.0
2027	30458698	45.5
2028	31930855	45.0
2029	33403166	44.5
2030	34875476	44.0

According to Table 3, the population of India is predicted to grow by 1.59 billion people during the next ten years, from 1.39 billion by 2020 to 1.59 billion in 2030, with a peak of nearly 0.2 billion around 2030, representing a 14.4 per cent increase over ten years at a rate of 1.4 per cent per annually. The urban population was 0.47 billion in 2020 and is expected to increase to 0.52 billion in the next five years and 0.57 billion by 2030, a 20.8 per cent increase from the initial period of forecasting at a rate of approximately 2 per cent per year. The rural population was 0.91 billion in 2020 and is expected to increase to 0.96 billion by 2025, indicating a 5.3% increase in the last half decade, with a further increase to 1.02 billion forecasted by the last decade, showing a 10.6% increase over 2020 at a rate of one per cent per annual.

The forecasting of the urban population shows nearly double the percentage growth of the rural population and approximately one-half of the total population. Therefore, urban population growth is highest as compared to rural and total population. This is an alarm to the policy makers framing policies regarding the increase in urban infrastructure. In the future, over population will have a number of consequences, including Socio-demographic and economic problems that are increased social polarization - theft, crime and social insecurity, affect to the demographic factors and also economic problems like unemployment, food shortages.

Growing population pressure affects social insecurity and presents an environmental risk to existing land resources. This issue may be solved by improving land production. Importantly, agricultural productivity has grown dramatically as a result of improved farming methods and technology. As a result, innovative technology and sustainable land management are critical to the agricultural production potential in India.

The rural population will increase to 0.91 billion in ten years. Therefore, it is already implemented and more technology in agriculture is needed to improve the economy's boost. And also, women are more likely to participate in work; fertility may be decreased. Work in these areas is often more focused on agriculture than in urban areas.

In 2020, the birth rate in India was 19.6 per 1,000 people, with its a predicted drop to 16.7 per 1,000 people in 2025, its 15.3% decline last roughly five years, and 13.4 per 1,000 people by 2030, a 31.6 percent decrease from 2020. Early marriage delay, family planning, female literacy, young desire, urban growth, contraception, and health care are among the most significant factors connected to fertility decrease.

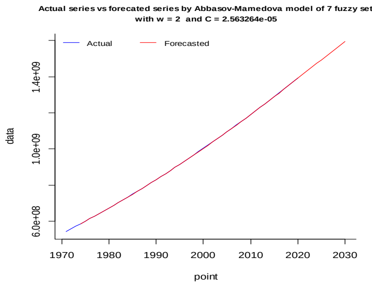
In 2020, the death rate in India was 6.0 per 1,000 people, with its predicted fall to 5.5 per 1,000 people in the last half decade, an 8.3% decrease, and 5.0 per 1,000 people by the last decade, a 16.6% decrease from 2020. In India in 2020, the infant mortality rate was 30 deaths per 1,000 children, with its projected drop to 25 deaths per 1,000 children in the last five years, a 16.6 percent decrease from 2020, and 20 deaths per 1,000 children by 2030, indicating 33.3% decrease in the previous decade. The total fertility rate in India was 2.0 children per woman in 2020, and it is expected to fall to 1.5 children per woman in 2025, a 25% drop from 2020, and 1.0 child per woman by 2030, a 50% decrease from the last decade. The assumption is that the Total fertility rate will continue to fall at its current rate.

India's under-5 mortality rate was 31.8 children per 1000 live births in 2020. It is predicted to fall to 18.5 children per 1000 live births in 2025, a 41.8 percent decrease in the last half decade, and 3.8 children per 1000 live births by 2030, an 88 percent decrease over 2020. Forecasted under-5 mortality is about five times the percentage drop of the death rate, roughly triple the total fertility rate, and roughly double the infant mortality rate. As a result, the under-5 mortality rate has declined the least in comparison to the death rate, total fertility rate, and infant mortality rate because of innovations in health-care technologies and female education.

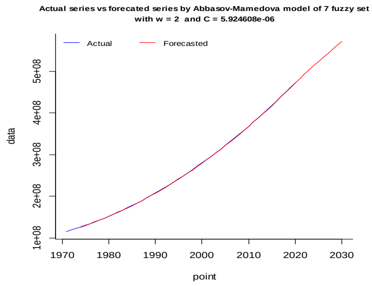
Table 5: Prediction of the accuracy test

Variables	MAPE
Total Population	0.025
Urban Population	0.104
Rural Population	0.003
Birth rate	1.095
Death rate	3.256
Infant Mortality Rate	2.358
Total Fertility Rate	1.577
Under 5 Mortality	0.274
Life Expectancy	0.386
GDP	49.83
Age Dependency Ratio	0.783

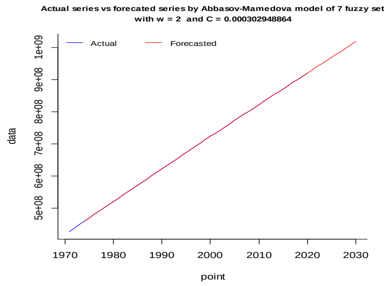
Total population



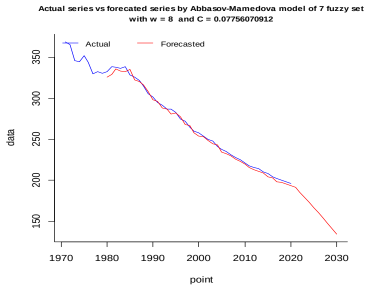
Urban population



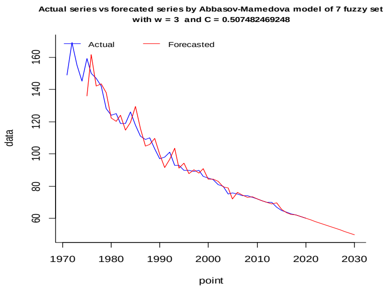
Rural population



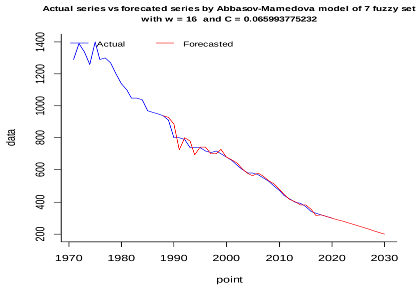
Birth rate



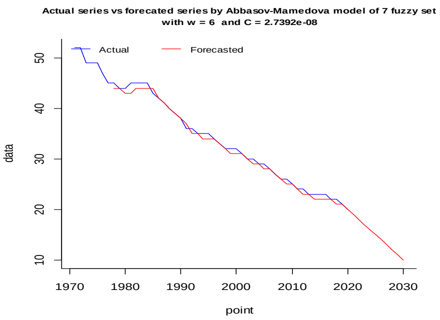
Death rate



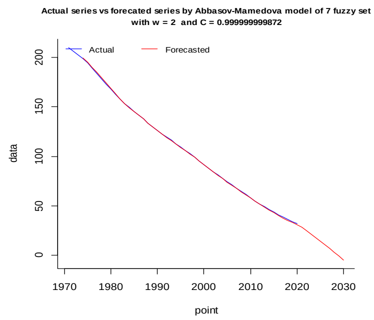
Infant Mortality rate



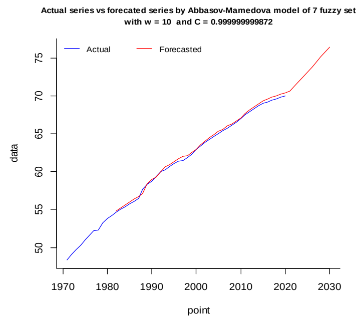
Total fertility rate



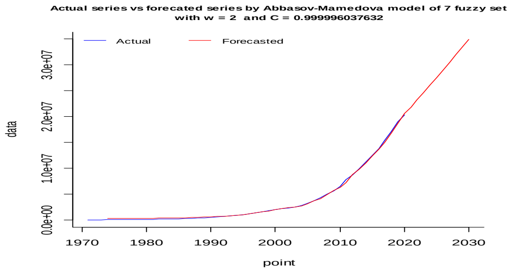
Under-5 mortality rate



Life Expectancy at Birth



Gross Domestic Product (GDP)



Age Dependency ratio

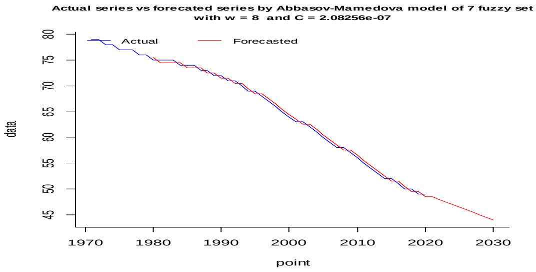


Figure 6: Forecasting of demographic and economic variables for fuzzy time series from 2020 to 2030.

According to table 3, the birth rate is falling, and the death rate is continuing to decline. Therefore, the forecasted population has been increasing due to the large number of people in reproductive age group and immigration also increases because of the increase in job opportunities and due to the impact globalization, more companies and jobs are created. Because of the implementation of family planning India, there may be lower infant mortality rate and also death rate is declining as there is improvement in the health infrastructure in the country, this indicates that Indian demographic transition is expected to be the third stage from the late expanding stage.

India's life expectancy at birth was 70 years in 2020, it is an increase to 73.1 years in 2025 with 4.4 per cent increase from the last five years, and predicted to 76.4 years by 2030. 9.1 per cent increase from the last decade. Longer life expectancy generates a higher return on human capital, encouraging more investment in education and health, increasing economic growth and decreasing mortality.

According to table 4, India's GDP was 203 trillion in 2020, with a projected growth to 275 trillion in the next half decade, representing 35.3 per cent increase from 2020, and 348 trillion in the last decade, indicating 71.4 per cent increase over the previous ten years. The Age Dependency Ratio in India was 49 in 2020, with a predicted fall to 46.5 in 2025, a 5.1% decrease in the last half decade, and 44 by 2030, a 10.2% decrease in the last decade. The age dependence ratio is decreasing, and as the working-age population grows, the GDP will increase. Then, as our projected population grows, GDP will also increase, ensuring that both variables have equal future value, or a balanced both variables.

Forecasts for the ten-year period 2021–2030 are provided in table 5, and figure.6 which demonstrates that the Abbasov and Mamedova technique produces good forecasting results except for GDP since MAPE value is minimum value except GDP. The established model fits almost all of the actual values well and best fit is observed in case of rural population followed by total population and urban population, next position of better fit is observed in case of vital rates and poor fit is observed for economic variable (GDP). The Figure 6 demonstrates that the predicted and actual values for all selected variables are nearly identical. The Fuzzy Time Series model is best suited for forecasting the entire selected variable for India except GDP.

IV. CONCLUSION

This work offers a Fuzzy Time series forecasting model capable of interpolating historical data and projecting the future. In describing actual uses of our approach, we illustrate how we used algorithms to find the appropriate parameters for each data set, such as the constant C and division intervals for the universal set. We are using package R software for this work. The demographic and economic variables are well fitted for forecasting results except GDP from MAPE. They conclude that the prediction accuracies for total population, urban population, rural population, birth rate, death rate, IMR, TFR, Under-5 mortality rate, life expectancy at birth, GDP, and age dependency ratio are 99.98%, 99.90%, 99.99%, 99.91%, 97.85%, 98.75%, 99.53%, 99.61%, 51.27%, and 99.32%, respectively. Among the other variables, the rural population is the best fit. The rural population outperforms the other populations, and the birth rate is well within the rates. When compared to GDP, the age dependence ratio is the best outfit.

In the next 10 years of forecasting, these things may happen: The population has grown due to a consistent drop in the mortality rate and a modest decline in the birth rate due to government improvements in policymaking, infrastructure, saturation, and healthcare facilities. Population growth is influenced by fertility, death, and migration rates. Population growth is the sum of natural growth and migratory impacts. Thus, a high rate of natural growth can be offset by a high level of net out-migration, and a low rate of natural growth can be offset by a high level of net in-migration. However, migration has a far smaller impact on population growth rates than changes in birth and death rates.Apart from the fall in birth and death rates, migration also plays a significant role in population growth. So, because the population is increasing, in-migration will be greater than out-migration. There is a precaution that must be considered during migration. As a result, immigration policies must be implemented. The life expectancy at birth for the next ten years is forecast. There is a two-year increase from 2021 to 2030 because of the decline in the death rate and birth rate due to improvements in healthcare facilities. The birth rate, death rate, IMR, TFR, and under-5 mortality rate are all diminishing as a result of the melioration of healthcare facilities, female education and enhancing population health. The GDP is very high, which may be due to the fact that the age dependency ratio has declined. That indicates the number of working-age people is increasing. After a ten-year forecast, the demographic transition is still in the late-expending stage due to a reduction in the birth rate and a sustained fall in the mortality rate.

Acknowledgement: Authors are thankful for the financial support from University Research Scholarship (URS), Karnataka University, Dharwad.

REFERENCES

[1]. Abbasov, A.M., Mamedova, M.H. (2003). Application of fuzzy time series to population forecasting. Vienna Univ. Technol. **12**, 545–552.

[2]. Box, G.E.P., Jenkins, G.M. (1976). Time series analysis: forecasting and control. Holden-Day Series in Time Series Analysis, Revisededn. Holden-Day, San Francisco.

[3]. Chen, S.M. (1996). Forecasting enrollments based on fuzzy time series. Fuzzy Sets Syst.**81**(3),311–319.

[4]. Chen, S.M., Hsu, C.C. (2004). A new method to forecast enrollments using fuzzy time series.Int. J. Appl. Sci. Eng.**2**(3), 234–244.

[5]. Galton, F.: Co-relations and their measurement, chiefly from anthropometric data. Proc. Roy. Soc. Lond. **45**(273–279), 135–145 (1888).

[6]. Ghazali, R., Hussain, A.J., Al-Jumeily, D., Lisboa, P. (2009). Time series prediction using dynamic ridge polynomial neural networks. In: 2009 Second International Conference on Developments in eSystems Engineering (DESE), pp. 354–363. IEEE.

[7]. Gupta, S., Wang, L.P. (2010). Stock forecasting with feed forward neural networks and gradual data sub-sampling. Aust. J. Intell. Inf. Process. Syst. **11**(4), 14–17.

[8]. Ha Che-Ngoc, Tai Vo-Van, Quoc-Chanh Huynh-Le, Vu Ho, Thao Nguyen-Trang., (2018). An improved fuzzy times series forecasting model, studies in computational intelligence, 474-490.

[9]. Huarng, K. (2001). Heuristic models of fuzzy time series for forecasting. Fuzzy Sets Syst.**123**(3), 369–386.

[10]. Park, D.C. (2010). A time series data prediction scheme using bilinear recurrent neural network. In: 2010 International Conference on Information Science and Applications (ICISA), pp 1–7. IEEE.

[11]. Pearson, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. Philos. Trans. Roy. Soc. Lond. Ser. A Contain. Pap. Math. Phys. Character **187**, 253–318.

[12]. Ren, Y., Suganthan, P.N., Srikanth, N., Amaratunga, G. (2016). Random vector functional link network for short-term electricity load demand forecasting. Inf. Sci. **367**, 1078–1093.

[13]. Sasu, A. (2010). An application of fuzzy time series to the Romanian population. BulletinTransilv. Univ. Brasov,**3**, 52.

[14]. Singh, S.R. (2008). A computational method of forecasting based on fuzzy time series. Math.Comput. Simul. **79**(3), 539–554.

[15]. Song, Q., Chissom, B.S. (1993). Forecasting enrollments with fuzzy time series part I. Fuzzy Sets Syst. **54**(1), 1–9.

[16]. Teo, K.,Wang, L., Lin, Z. (2001). Wavelet packet multi-layer perceptron for chaotic time series prediction: effects of weight initialization. In: Computational Science-ICCS 2001, pp 310–317.

[17]. Tseng, F.M., Tzeng, G.H. (2002). A fuzzy seasonal ARIMA model for forecasting. Fuzzy Sets Syst. **126**(3), 367–376.

[18]. Wang, L., Teo, K.K., Lin, Z. (2001). Predicting time series with wavelet packet neural networks. In: Proceedings of the International Joint Conference on Neural Networks, IJCNN 2001. vol 3, pp 1593–1597. IEEE.

[19]. Wang, L., Fu, X. (2006). Data Mining with Computational Intelligence. Springer Science & Business Media, New York.

[20]. Zadeh, L.A. (1965). Fuzzy sets. Inf. Control **8**(3), 338–353.

[21]. Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning, parts 1-3, Inform. Sci.8 (1975) 199-249;8(1975) 301-357; 943-80.

[22]. Zecchin, C., Facchinetti, A., Sparacino, G., De Nicolao, G., Cobelli, C. (2011). A new neural network approach for short-term glucose prediction using continuous glucose monitoring time-series and meal information. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, pp 5653–5656. IEEE.

[23]. Zhang, G.P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing**50**, 159–175.

[24]. Zhu, M., Wang, L. (2010). Intelligent trading using support vector regression and multilayer perceptrons optimized with genetic algorithms. In: The 2010 International Joint Conference on Neural Networks (IJCNN), pp. 1–5. IEEE.