

**Assistant Professor Miraç EREN, PhD**  
**E-mail: mirac.eren@omu.edu.tr**  
**Ondokuz Mayıs University, Samsun**

## **FORECASTING OF THE FUZZY UNIVARIATE TIME SERIES BY THE OPTIMAL LAGGED REGRESSION STRUCTURE DETERMINED BASED ON THE GENETIC ALGORITHM**

***Abstract.** Estimation obtained through classical regression model reveals the fitting (or prediction) and projection (or forecast) values with a certain error. This situation leads to loss of information and imprecision of data. However, if the imprecise information is converted to fuzzy data rather than single value, an estimation procedure can be obtained in which observation errors are hidden in fuzzy coefficients. Thus, it would be more realistic to make an interval estimate instead of a single value estimate with a certain margin of error. Therefore, in this study, a novel fuzzy least squares method developed for the variables expressed by LR-type fuzzy numbers, based on the optimal classical lagged regression model structure determined by the genetic algorithm, was addressed. a numerical example to explain how the proposed method is applicable was considered.*

***Keywords:** Genetic Algorithms, LR-Type Fuzzy Numbers, Fuzzy Least Squares Method, Time Series, Forecasting.*

**JEL Classification: C22, C53**

### **1. Introduction**

In the real world, the data cannot be considered and modeled precisely. For example, the water level of a river, the temperature in a room or the electricity consumption cannot be measured in an exact way because of their fluctuations[1]. Therefore, trying to find an applicable method with less error to estimate the variation of a dependent (output or response) variable according to the variation of explanatory (input or independent) variables is one of the most important goals for today's researchers. For this, regression analysis, which is made to determine the best fitting coefficients of the model afterward creating an appropriate mathematical model from the knowledge of the values of a given data set, is one of the most used statistical tools.

The classical statistical regression approach that is bounded by some strict assumptions about the given data provides a crisp relationship between explanatory and response variables, based on a crisp data set, from a statistical viewpoint. In addition, deviations between the observed and the estimated values are supposed to be due to measurement errors and/or random variations[2]. However, in many real applications, the relationships among variables where the available data may be very limited and imprecise, and variables may be interacting in an uncertain,

qualitative. At the same time, the deviations due to the imprecisely observed data or the indefiniteness of the system structure and parameters may be. In this case, the uncertainty is not due to randomness but fuzziness[3]. As a result, two types of uncertainty arise. First, there exists probabilistic uncertainty, which is solved overtime or through experimentation. Secondly, there is fuzzy uncertainty that has nothing to do with experimentation or time. In other words, some aspects of uncertainty measure the vagueness of the phenomena (due to their inconsistency to the existent criteria or vagueness in their definition) cannot be summarized in random terms. This kind of uncertainty is evaluated by a measure called possibility. Hence, error terms are deleted in fuzzy regression models and are, in fact, hidden in the fuzzy coefficients [4]. This fuzzy uncertainty described as ambiguity and vagueness has revealed the theory of fuzzy sets introduced by Zadeh [5] to build such a system as needed to deal with ambiguous and vague sentences or information[6].

For overcoming such limitations mentioned in classical statistical regression approach, fuzzy regression, which is an extension of the classical regression and is used in estimating the relationships among variables where the available data are very limited and imprecise, and variables are interacting in an uncertain, qualitative and fuzzy way, is introduced. Fundamental differences between fuzzy regression and classical regression are as follows:

- Fuzzy regression can be used to fit fuzzy data and crisp data into a regression model, whereas ordinary regression can only fit crisp data.

- Statistical regression analysis is based on some strict assumptions. The unobserved error term should mutually be independent and identically distributed. Lack of such assumptions would affect the effectiveness of the method. In this case, fuzzy regression can be replaced.

- In contrast to the ordinary regression that is based on probability theory, fuzzy regression is based on possibility theory and fuzzy set theory.

- Ordinary regression modeling data with randomness type of uncertainty but fuzzy regression modeling data with fuzziness type of uncertainty.

- In ordinary regression, the unfitted errors between a regression model and observed data are assumed as observation error that is a random variable. In fuzzy regression, the same unfitted errors are viewed as the fuzziness of the model structure[7].

From the methodological and conceptual aspects, it is possible to roughly divide the fuzzy linear regression (FLR) methods into two categories as linear programming (LP)-based methods and fuzzy least-squares methods.

Linear programming (LP)-based methods using the minimum fuzziness criterion aims to build fuzzy linear models by minimizing the system fuzziness subject to including the data points of each sample within a specified feasible data interval. This approach was first proposed by Tanaka, Hayashi and Watada [8], Asai [9], afterward was investigated and improved by many authors.

Fuzzy Least Squares and Least Absolutes Methods based on a distance on the space of fuzzy numbers extend the least absolutes/least squares errors criterions

to the fuzzy setting. This approach was first proposed by Diamond [10], Celmiņš [11], simultaneously, afterward, was investigated and improved by many researchers. In the literature, there are also studies that present the fitting (or prediction) and projection (or forecast) of time series in the framework of fuzzy regression. Again, it is possible these studies to divide based on linear programming (LP) and fuzzy least squares/ least absolutes. For studies on fuzzy linear programming (possibilistic) estimation of time series, look at: [12-15]. For studies on fuzzy least squares/ least absolutes estimation of time series, look at: [16]. In addition, there are also the fuzzy possibilistic methods [17-19] and fuzzy least squares/ least absolutes methods [20] combined with heuristic (evolutionary) techniques.

Combining different forecasting structures can be an effective way of improving forecasting performance. Therefore, in this study, to identify the functional relationship between variables with fuzzy input and output observations, a novel fuzzy least squares method developed for the variables expressed by LR typed fuzzy numbers, based on the optimal classical lagged regression model structure determined by the genetic algorithm, is applied. This paper is organized as follows. The following section (Section 2) introduces the preliminary concepts related fuzzy sets and fuzzy arithmetic. Section 3 presents the research methodology. In Section 4, the improved method for the case study is proposed. The results of the numerical example are given in Section 5. Finally, Section 6 concludes the study, and give recommendations for future research.

## 2. Preliminary Concepts about Fuzzy Sets and Fuzzy Arithmetic

A fuzzy set  $\tilde{A}$  on the universal set  $X$  is described by its membership function  $\tilde{A}(x): X \rightarrow [0,1]$  (where  $X = \mathbb{R}$  is assumed). The  $\alpha$ -cut of a fuzzy set  $\tilde{A}$  is defined as the crisp set  $A_\alpha = \{x \in \mathbb{R}: \tilde{A}(x) \geq \alpha\}$ ,  $\alpha \in (0,1]$ . If  $A_\alpha$  represented by  $[A_\alpha^l, A_\alpha^r]$  for any  $\alpha \in [0,1]$  is a non-empty compact interval, A fuzzy set  $\tilde{A}$  of  $\mathbb{R}$  is called a fuzzy number. The so-called LR-fuzzy numbers described by  $\tilde{N} = (n, l, r)_{LR}$  with central value  $n \in \mathbb{R}$ , left and right spreads  $l, r \in \mathbb{R}$ , decreasing left and right shape functions  $L, R: \mathbb{R}^+ \rightarrow [0,1]$ , with  $L(0) = R(0) = 1$  are also one of the specific type of fuzzy numbers, which are rich and flexible enough to cover most of the applications. Such LR-fuzzy numbers have the form of membership function [21]

$$\tilde{N}(x) = \begin{cases} L\left(\frac{n-x}{l}\right) & , \quad x \leq n, \\ R\left(\frac{x-n}{r}\right) & , \quad x > n. \end{cases}$$

According to this, the so-called triangular fuzzy numbers being a special type of LR-fuzzy number and denoted by  $\tilde{N} = (n, l, r)_T$  are defined by the following membership function:

$$\tilde{N}(x) = \begin{cases} \frac{x - (n - l)}{l} & , x \in [n - l, n], \\ \frac{(n + r) - x}{r} & , x \in (n, n + r]. \end{cases}$$

The  $\alpha$ -cut of the LR-triangular fuzzy number  $\tilde{N}$  is as follows

$$N_\alpha = [N_\alpha^l, N_\alpha^r] = [n - (1 - \alpha)l, n + (1 - \alpha)r], \quad \alpha \in [0, 1].$$

A number of results for the algebraic operations of LR-fuzzy numbers based on Zadeh's extension principle are present. Let  $\tilde{M} = (m, l_m, r_m)_{LR}$  and  $\tilde{N} = (n, l_n, r_n)_{LR}$  be two triangular fuzzy numbers and  $\lambda$  be a real number. Then

$$\lambda \otimes \tilde{M} = \begin{cases} (\lambda m, \lambda l_m, \lambda r_m)_{LR} & , \lambda > 0, \\ l_{\{0\}} & , \lambda = 0, \\ (\lambda m, |\lambda|l_m, |\lambda|r_m)_{LR} & , \lambda < 0, \end{cases}$$

Where  $l_{\{0\}}$  stands for the indicator function of the crisp zero. When  $I = [i_1, i_2]$  and  $J = [j_1, j_2]$  be two closed intervals, then

$$\lambda \otimes I = \begin{cases} [\lambda i_1, \lambda i_2] & , \lambda > 0, \\ 0 & , \lambda = 0, \\ [\lambda i_2, \lambda i_1] & , \lambda < 0. \end{cases}$$

$$I \otimes J = [i_1 + j_1, i_2 + j_2]$$

The distance between two these intervals

$$\mathcal{D}(I, J) = (i_1 - j_1)^2 + (i_2 - j_2)^2.$$

### 3. Research methodology

#### 3.1. Modeling of univariate time series

$Y_t$ , the sequence of data points indexed in successive equally spaced  $t = 1, 2, \dots, T$  time order, is called as the univariate time series. There are two main points of time series: the dynamic causal effect, seeking an answer to the question of "What is the effect of  $X$  on  $Y$  over time?" and the forecasting, seeking an answer to the question of "What are predicted future values of  $Y$ , given available information?". In this study, an autoregressive (AR) model is used to obtain the prediction and forecasting of a univariate time series. Because, the intuition behind

this model is, that the observed time series  $Y_t$  depends on a weighted linear sum of the past values,  $k$ , of  $Y_t$  and a random shock  $\varepsilon_t$ . From this perspective, an autoregressive model corresponds simply to a linear regression of the current value of the series against one or more prior values of the series, and can therefore be analyzed among other methods with the standard linear least squares technique, where the resulting estimation of the parameters, has a straight forward interpretation. Technical, the AR ( $k$ ) model can be formulated as follows:

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_k Y_{t-k} + \varepsilon_t,$$

where  $Y_t$  denotes the time-series and  $\varepsilon_t$  indicates a white-noise process. The value of  $p$  is called the order of the AR model.

### 3.2. Estimation procedure of fuzzy lagged time series model

Assume that  $\tilde{Y}_t$  and  $\tilde{Y}_{t-1}, \tilde{Y}_{t-2}, \dots, \tilde{Y}_{t-p}$  are respectively the fuzzy dependent and independent variables expressed as LR-fuzzy numbers of their observations. A lagged regression model for the aforementioned data is

$$\tilde{Y}_t = \beta_0 \oplus \beta_1 \tilde{Y}_{t-1} \oplus \beta_2 \tilde{Y}_{t-2} \oplus \dots \oplus \beta_k \tilde{Y}_{t-k} \oplus \varepsilon_t.$$

Here, in order to estimate the crisp coefficients vector  $\beta = [\beta_0 \ \beta_1 \ \dots \ \beta_k]^t$ , a novel method developed by Chachi, J., and Taheri, S. M. (2016) to identify the functional relationship between variables by means of a fuzzy regression model where both input and output observations are given as fuzzy data is considered. The prediction procedure adapted to the lagged regression model is constructed according to the following steps.

**Step 1:** Primarily, the regression model based on the  $\alpha$ -level sets of the fuzzy input-output data is constructed. By considering the  $\alpha$ -level sets of the observed data, the following form of the univariate lagged interval-valued (or multiple) regression models is considered

$$[y_{t,\alpha}^l, y_{t,\alpha}^r] = \beta_0(\alpha) \oplus \beta_1(\alpha) \otimes [y_{t-1,\alpha}^l, y_{t-1,\alpha}^r] \oplus \dots \oplus \beta_k(\alpha) \otimes [y_{t-k,\alpha}^l, y_{t-k,\alpha}^r].$$

Note that, the regression coefficients  $\beta_0(\alpha), \beta_1(\alpha), \dots, \beta_k(\alpha)$  are real-valued functions of  $\alpha \in [0,1]$ .

**Step 2:** The unknown vector parameter  $\beta = [\beta_0(\alpha) \ \beta_1(\alpha) \ \dots \ \beta_k(\alpha)]^t$  is evaluated by minimizing the following objective function

$$\mathcal{D}(\beta) = \sum_{t=k+1}^n \mathcal{D}(Y_{t,\alpha}, X_{t,\alpha}\beta)$$

This objective function is the sum of squared distances between the  $\alpha$ -level sets of observations of the dependent variable  $Y_{t,\alpha}$  and their estimated values  $X_{t,\alpha}\beta$ , where  $X_{t,\alpha} = [1 \ Y_{t-1,\alpha} \ \dots \ Y_{t-k,\alpha}]$  is the vector of the  $\alpha$ -level sets of independent variables.

Since the interval multiplication is applied, it is necessary to ensure that the minimum and maximum values of the  $\alpha$ -level sets are properly considered for both positive and negative cases of  $\beta_0(\alpha), \beta_1(\alpha), \dots, \beta_k(\alpha)$ . In each case, such a requirement can be overcome by considering the following substitution variables.

$$L_{tj,\alpha} = \begin{cases} X_{tj,\alpha}^l & \beta_j(\alpha) \geq 0, \\ X_{tj,\alpha}^r & \beta_j(\alpha) < 0, \end{cases} \quad R_{tj,\alpha} = \begin{cases} X_{tj,\alpha}^r & \beta_j(\alpha) \geq 0, \\ X_{tj,\alpha}^l & \beta_j(\alpha) < 0, \end{cases}$$

Where,  $j = 0, 1, \dots, k$ ,  $t = 1, \dots, T$  and  $X_{0j,\alpha}^l = X_{0j,\alpha}^r = 1$ . Then, if the objective function  $\mathcal{D}(\beta)$  is rewritten,

$$\mathcal{D}(\beta) = \|Y_\alpha^l - L_\alpha\beta\|^2 + \|Y_\alpha^r - R_\alpha\beta\|^2,$$

Where,  $\|\cdot\|$  is the Euclidean norm, and  $L_\alpha$  and  $R_\alpha$  are matrices with general elements  $L_{tj,\alpha}$  and  $R_{tj,\alpha}$ , respectively. By equating to zero the partial derivative of  $\mathcal{D}(\beta)$  according to the unknown parameter  $\beta$ , the best value of  $\beta$  which minimizes the objective function  $\mathcal{D}(\beta)$ , is found. As a result, the mentioned situation corresponds to the system of linear equations in form of

$$L_\alpha^t L_\alpha \beta - L_\alpha^t Y_\alpha^l + R_\alpha^t R_\alpha \beta - R_\alpha^t Y_\alpha^r = 0,$$

Where  $L_\alpha^t$  and  $R_\alpha^t$  are the transposes of matrices  $L_\alpha$  and  $R_\alpha$ , respectively. Subject to the existence of  $(L_\alpha^t L_\alpha + R_\alpha^t R_\alpha)^{-1}$ , the least squares estimate of  $\beta$  is obtained as

$$\beta = (L_\alpha^t L_\alpha + R_\alpha^t R_\alpha)^{-1} (L_\alpha^t Y_\alpha^l + R_\alpha^t Y_\alpha^r).$$

**Step 3:** Finally, the optimal solutions  $\beta_j(\alpha), j = 0, 1, \dots, k$ , to estimate the regression coefficients  $\beta_j$ , are aggregated as

$$\hat{\beta}_j = \int_0^1 \beta_j(\alpha) d\alpha, \quad j = 0, 1, \dots, p.$$

Thus, the optimal model is obtained as

$$\hat{Y}_t = \hat{\beta}_0 \oplus \hat{\beta}_1 \otimes \tilde{Y}_{t-1} \oplus \dots \oplus \hat{\beta}_k \otimes \tilde{Y}_{t-k}.$$

#### 4. Method proposed for the case study

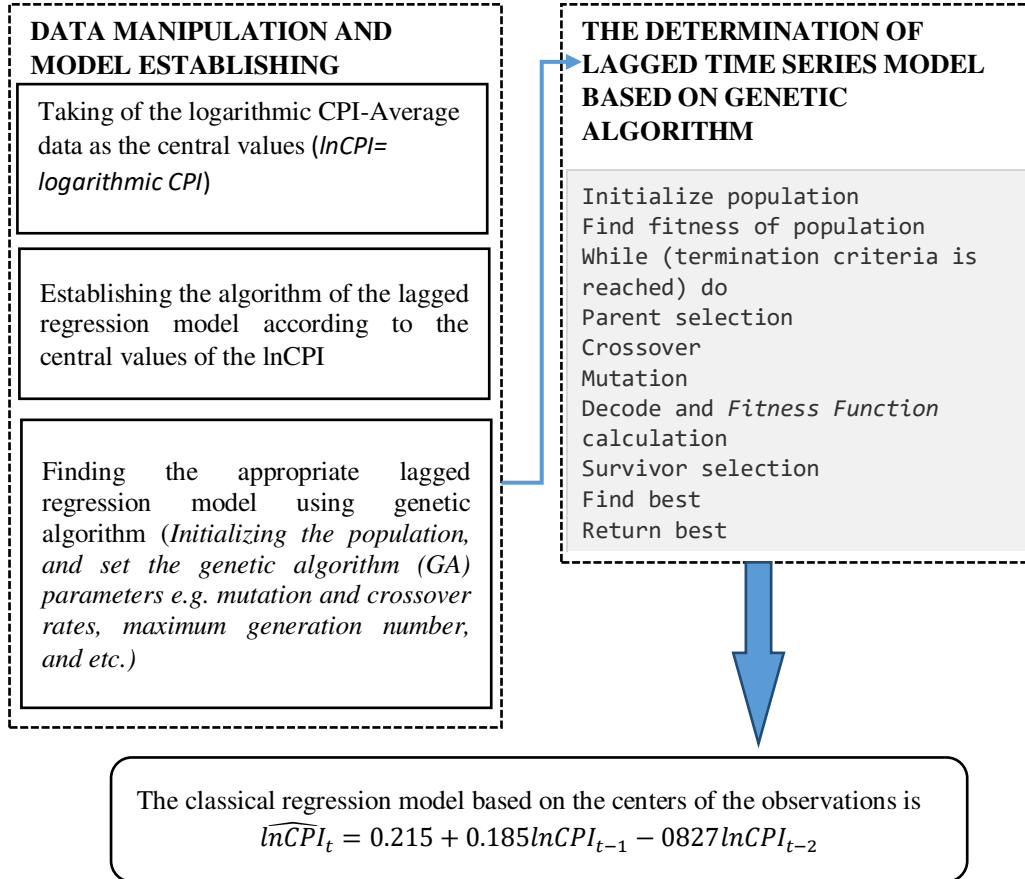
Forecasting of the Fuzzy Univariate Time Series by the Optimal Lagged Regression Structure Determined Based on the Genetic Algorithm

In this research, the consumer price index (CPI) forecasting is addressed as a numerical example to explain how the proposed method is applicable to obtain a suitable regression model for fuzzy observations. Here, the objective is an annual forecast based on the monthly CPI (base year 1978-1979) data between 1982 and 2016. Therefore, it is necessary to create a regression model suitable for the fuzzy data obtained by converting the monthly consumer price index data into annual data by taking their lower limit, average and upper limit values. Hence, error terms are deleted in the fuzzy regression model and are, in fact, hidden in the fuzzy coefficients. According to this, conversion of monthly CPI data to annual data is

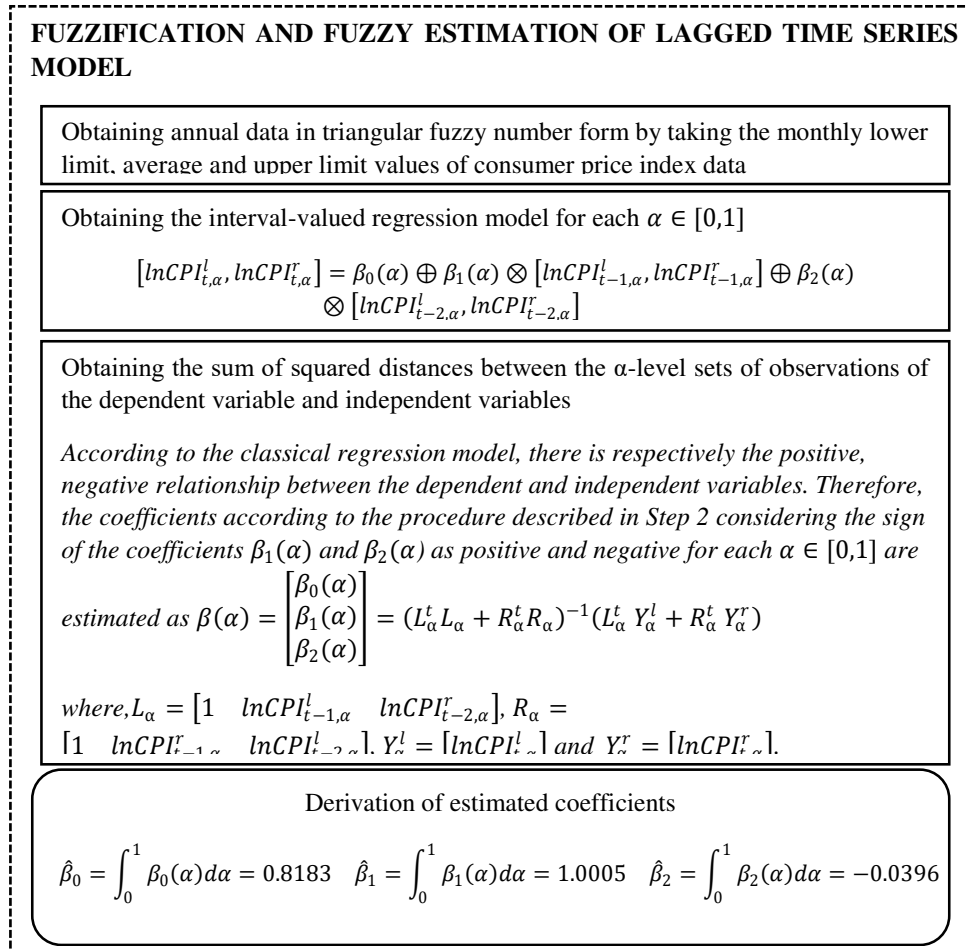
**Table 1. Annual CPI data expressed in triangular fuzzy numbers**

Years	CPI- Lower	CPI- Average	CPI- Upper	Years	CPI- Lower	CPI- Average	CPI- Upper
1982	(373,	410,	462)	2000	(2561547,	2960721,	3401704)
1983	(481,	539,	633)	2001	(3473079,	4545060,	5820672)
1984	(655,	800,	947)	2002	(6174953,	6733431,	7598668)
1985	(1000,	1160,	1366)	2003	(7761927,	8506320,	9007595)
1986	(1410,	1561,	1785)	2004	(9049038,	9208410,	9599091)
1987	(1837,	2168,	2767)	2005	(9682418,	10136773,	10700413)
1988	(2933,	3801,	4848)	2006	(10120727,	10952013,	11634954)
1989	(5113,	6447,	8183)	2007	(11125675,	11910289,	12610805)
1990	(8493,	10547,	13141)	2008	(12034795,	13155479,	13908329)
1991	(13697,	17503,	22484)	2009	(13178158,	13975633,	14785709)
1992	(24485,	30053,	37748)	2010	(14256818,	15172005,	15747233)
1993	(39859,	50392,	64695)	2011	(14955449,	16157062,	17375868)
1994	(66961,	106102,	150181)	2012	(16542725,	17591030,	18446882)
1995	(160158,	206323,	263667)	2013	(17751610,	18908996,	19812036)
1996	(285265,	366475,	462066)	2014	(19127577,	20584646,	21481504)
1997	(486939,	672724,	916498)	2015	(20512553,	22164543,	23318354)
1998	(979423,	1225733,	1526376)	2016	(22477401,	23885562,	25308122)
1999	(1588988,	1943553,	2449441)				

To get rid of the heteroscedasticity problem in the regression model to be established regarding the CPI forecasting, it would be better to use their natural logarithms instead of the CPI data in Table 1. According to this, flowchart for prediction in the fuzzy logic framework of the logarithmic CPI data is







**Figure 1. Flowchart of the proposed methodology**

If the phase of “Establishing the algorithm of the lagged regression model according to the central values of the CPI” in the Flowchart is further elaborated, the algorithm is

```

Fitness Function (gen) {
max.lag=Length(gen)
X=NULL
AIC.Criterion=NULL
DO FOR i=1, max.lag
    IF gen[i]==1 THEN
        X= [, append (lnCPIt-i)]
    ENDDO
Result=Regress (lnCPIt,X)
AIC.Criterion=AIC (Result)
Return (AIC.Criterion)
}
    
```

**Figure 2. The proposed fitness function algorithm**

**5. Results and discussions**

According to the regression coefficients estimated at the end of the presented methodology regarding the prediction of the logarithmic CPI data using to explain how the proposed method is applicable to obtain a suitable regression model for fuzzy observations, the fuzzy linear regression model is

$$\widehat{\ln CPI}_t = 0.8183 \oplus 1.0005 \otimes \widehat{\ln CPI}_{t-1} \ominus 0.0396 \otimes \widehat{\ln CPI}_{t-k}.$$

The arranged form of the established regression model for the forecasting of the logarithmic CPI data, of which the syntax of the code is written in MATLAB, and its goodness of fit is

**Table 2. The performance of established fuzzy regression model for logarithmic CPI data**

	$\widehat{\ln CPI}_t = (\widehat{\ln CPI}_t, \widehat{\ln CPI}_t^l, \widehat{\ln CPI}_t^r)$	Performance (Goodness of fit)
Classical regression based model	$\widehat{\ln CPI}_t = 0.215 + 0.185 \widehat{\ln CPI}_{t-1} - 0.0827 \widehat{\ln CPI}_{t-2}$ $\widehat{\ln CPI}_t^l = \widehat{\ln CPI}_t - t_{\frac{0.05}{2}, 35-3} \sqrt{\widehat{Var}(\widehat{\ln CPI}_t)}$ $\widehat{\ln CPI}_t^r = \widehat{\ln CPI}_t + t_{\frac{0.05}{2}, 35-3} \sqrt{\widehat{Var}(\widehat{\ln CPI}_t)}$	$\gamma^2 = 332.572$

Forecasting of the Fuzzy Univariate Time Series by the Optimal Lagged Regression Structure Determined Based on the Genetic Algorithm

Proposed model	$\widehat{\ln CPI}_t = 0.8183 + 1.0005 \ln CPI_{t-1} - 0.0396 \ln CPI_{t-2}$ $\widehat{\ln CPI}_t^l = 1.0005 \ln CPI_{t-1}^l - 0.0396 \ln CPI_{t-2}^l$ $\widehat{\ln CPI}_t^r = 1.0005 \ln CPI_{t-1}^r - 0.0396 \ln CPI_{t-2}^r$	$\gamma^2 = 48.793$
----------------	--	---------------------

Here, the sum of squared distances between fuzzy numbers is  $\gamma^2 = \sum_{t=1}^T [\ln CPI_t - \widehat{\ln CPI}_t]^2 + [(\ln CPI_t - \ln CPI_t^l) - (\widehat{\ln CPI}_t - \widehat{\ln CPI}_t^l)]^2 + [(\ln CPI_t + \ln CPI_t^r) - (\widehat{\ln CPI}_t + \widehat{\ln CPI}_t^r)]^2$ . In addition,  $\widehat{\ln CPI}_t^l$  and  $\widehat{\ln CPI}_t^r$  demonstrated in the classical regression based model corresponds to lower and upper values of prediction interval for new observations, respectively (namely,  $\widehat{\ln CPI}_t - t_{\frac{0.05}{2}, 35-3} \sqrt{\widehat{Var}(\widehat{\ln CPI}_t)} \leq \widehat{\ln CPI}_t \leq \widehat{\ln CPI}_t + t_{\frac{0.05}{2}, 35-3} \sqrt{\widehat{Var}(\widehat{\ln CPI}_t)}$ ). The smaller the  $\gamma^2$  value, which is the sum of squared distances between prediction and real observation value, is, the higher the goodness of fit would be. According to this, as observed from  $\gamma^2$  values in Table 3, the performance of the proposed model will be higher than the conventional regression equation with a predictive interval.

So, Logarithmic CPI and its fitting (prediction) values of the proposed model are as in Figure 3.

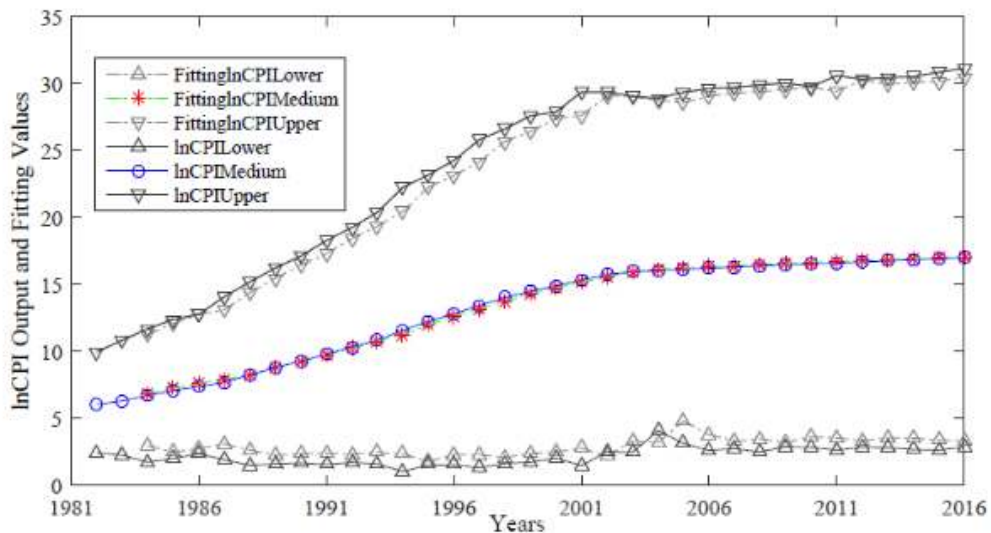
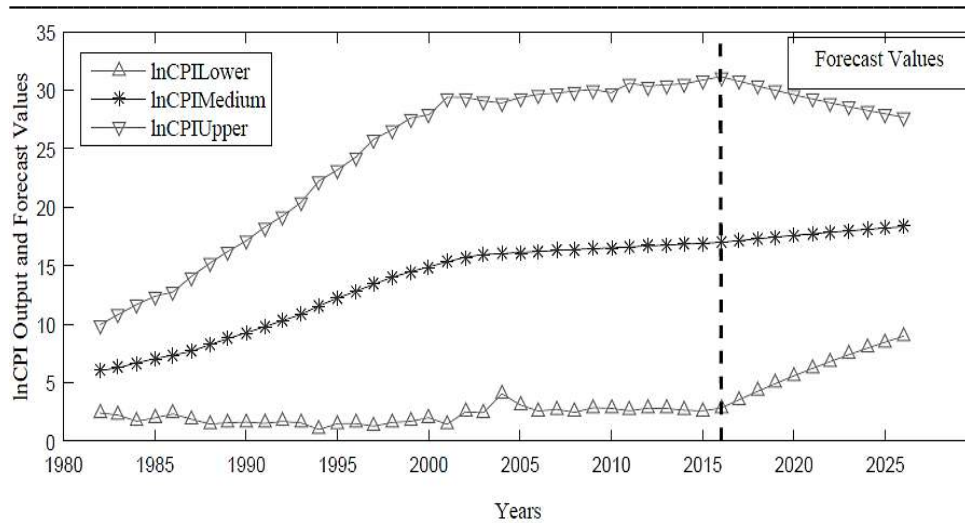


Figure 3. Logarithmic CPI and fitting (prediction) values

If a projection according to the resulting fuzzy regression model is made, the findings obtained will be as Figure 3.



**Figure 4. The projection of logarithmic CPI**

As a result, Real and projection CPI values are obtained by taking the anti-log of real and projection lnCPI values.

**Table 3. Forecasting CPI in 2020 and 2026**

	(CPI-Lower, CPI-Medium, CPI-Upper)
2020	(5.588, 17.590, 29.614)
2026	(8.979, 18.328, 27.694)

**6. Conclusions and Future Researches**

One of the main issues in linear regression models is imprecision or vagueness in the definition and/or observation of output and/or of inputs. Data imprecision may be due to several causes: (i) imprecision in measuring the empirical phenomena observed; (ii) vagueness of the variables of interest (inputs and/or outputs) when they are expressed in linguistic terms; (iii) partial or total ignorance about the variables’ values on specific instances; (iv) categorization (granularity) of the variables of interest. Therefore, In order to make the available information suitable, Converting imprecise data into fuzzy data could be more effective than replacing them with a single value. This situation is the same for time series, too. For example, when a monthly time series is converted to a yearly series, the monthly average is usually taken over the years. Then, the fitting (or prediction) and projection (forecast) values are revealed with a certain error through the regression model. This situation leads to loss of information and imprecision of data. However, if the imprecise information is converted to fuzzy data rather than single value, an estimation procedure can be obtained in which

## Forecasting of the Fuzzy Univariate Time Series by the Optimal Lagged Regression Structure Determined Based on the Genetic Algorithm

observation errors are hidden in fuzzy coefficients. Thus, it would be more realistic to make an interval estimate instead of a single value estimate with a certain margin of error. For this, there are many fuzzy regression models based on linear programming and fuzzy least squares methods in the literature. In this study, a fuzzy least squares based approach, a novel method developed by Chachi, J., and Taheri, S. M. (2016) to identify the functional relationship between variables with fuzzy input and output observations has been adapted to the lagged regression created to forecast univariate time series. As a case study, the consumer price index (CPI) data between 1982 and 2016 has been addressed. Accordingly, firstly, after taking the logarithmic CPI-average data as the central values, the optimal lagged regression model according to the central values of the logarithmic CPI was determined by the genetic algorithm. Then, annual data obtained in triangular fuzzy number form by being taken the monthly lower limit, average and upper limit values of logarithmic CPI data was predicted by the fuzzy least squares method according to the optimal classical lagged regression model structure determined by the genetic algorithm. As a result, prediction and projection CPI values were obtained by taking the anti-log of prediction and projection values of the logarithmic CPI data. Here, the presented method is discussed for a univariate model. In addition, as a future research, the method discussed can be adapted to multivariate models. Because, to establish a relationship with the regression between monthly and annual data, Monthly data are usually converted into annual data by being taken their average. However, this condition causes loss of information. It would be more realistic to use the improved fuzzy regression model presented in the research instead of estimating it with a certain error-sharing regression model to overcome this problem.

### REFERENCES

- [1] H.-C. Wu (2003), *Fuzzy Estimates of Regression Parameters in Linear Regression Models for Imprecise Input and Output Data*; Computational Statistics & Data Analysis 42(1), 203-217;
- [2] A. Azadeh, O. Seraj, M. Saberi (2010), *An Integrated Fuzzy Regression–Analysis of Variance Algorithm for Improvement of Electricity Consumption Estimation in Uncertain Environments*; The International Journal of Advanced Manufacturing Technology 53(5-8), 645-660;
- [3] M. Namdari, J.H. Yoon, A. Abadi, S.M. Taheri, S.H. Choi (2014), *Fuzzy Logistic Regression with Least Absolute Deviations Estimators*; Soft Computing 19(4) (2014) 909-917;
- [4] S. Pourahmad, T. Ayatollahi, S. Mohammad, S.M. Taheri (2011), *Fuzzy Logistic Regression: A New Possibilistic Model and Its Application in Clinical Vague Status*. Iranian Journal of Fuzzy Systems 8(1), 1-17;
- [5] L.A. Zadeh (1965), *Fuzzy Sets*; Information and Control 8(3) (1965) 338-353;

- 
- [6]W. Zeng, Q. Feng, J. Li(2017), *Fuzzy Least Absolute Linear Regression*; Applied Soft Computing 52; 1009-1019;
- [7]A. Azadeh, O. Seraj, M. Saberi (2008), *A Total Fuzzy Regression Algorithm for Energy Consumption Estimation*; Industrial Informatics, INDIN 2008. 6th IEEE International Conference on IEEE, pp. 1562-1568;
- [8]H. Tanaka, I. Hayashi, J. Watada (1989), *Possibilistic Linear Regression Analysis for Fuzzy Data*; European Journal of Operational Research 40(3), pp. 389-396;
- [9]H.T.-S.U.-K. Asai (1982), *Linear Regression Analysis with Fuzzy Model*; IEEE Trans. Systems Man Cybern 12; 903-907;
- [10]P. Diamond(1987), *Least Squares Fitting of Several Fuzzy Variables*; 2nd Int. Fuzzy Syst. Assoc. IFSA World Congress, pp. 329-331;
- [11]A. Celmiņš (1987), *Least Squares Model Fitting to Fuzzy Vector Data*; Fuzzy sets and systems 22(3); 245-269;
- [12]S. Muzzioli, B. De Baets (2013), *A Comparative Assessment of Different Fuzzy Regression Methods for Volatility Forecasting*; Fuzzy Optimization and Decision Making 12(4), 433-450;
- [13]S. Pourahmad, S.M.T. Ayatollahi, S.M. Taheri, Z.H. Agahi (2011), *Fuzzy Logistic Regression Based on the Least Squares Approach with Application in Clinical Studies*; Computers & Mathematics with Applications 62(9), 3353-3365;
- [14]M.N. Mehr, F.F. Samavati, M. Jeihoonian(2011), *Annual Energy Demand Estimation of Iran Industrial Sector by Fuzzy Regression and ARIMA*; Fuzzy Systems and Knowledge Discovery (FSKD), Eighth International Conference on IEEE, pp. 593-597;
- [15]S.-P. Chen, J.-F. Dang (2008), *A Variable Spread Fuzzy Linear Regression Model With Higher Explanatory Power and Forecasting Accuracy*; Information Sciences 178(20); 3973-3988;
- [16]L. Abdullah, N. Zakaria (2012), *Matrix Driven Multivariate Fuzzy Linear Regression Model in Car Sales*; *Journal of Applied Sciences*(Faisalabad) 12(1), pp. 56-63;
- [17]K.-P. Lin, P.-F. Pai, Y.-M. Lu, P.-T. Chang (2013), *Revenue Forecasting Using a Least-Squares Support Vector Regression Model in a Fuzzy Environment*; Information Sciences 220, 196-209;
- [18]A. Azadeh, A. Ziaiefar, K. Pichka, S.M. Asadzadeh (2013), *An Intelligent Algorithm for Optimum Forecasting of Manufacturing Lead Times in Fuzzy and Crisp Environments*; International Journal of Logistics Systems and Management, 16(2), pp. 186-210;
- [19]A. Azadeh, M. Saberi, O. Seraj (2010), *An Integrated Fuzzy Regression Algorithm for Energy Consumption Estimation with Non-Stationary Data: A Case Study of Iran*; Energy 35(6); 2351-2366;

Forecasting of the Fuzzy Univariate Time Series by the Optimal Lagged Regression Structure Determined Based on the Genetic Algorithm

- 
- [20] **M.H. Mashinchi, M.A. Orgun, M. Mashinchi, W. Pedrycz**(2011), *A Tabu–Harmony Search-Based Approach to Fuzzy Linear Regression*; IEEE Transactions on Fuzzy Systems 19(3); 432-448;
- [21] **H.-J. Zimmermann** (2011), *Fuzzy Set Theory—and its Applications*; Springer Science & Business Media.