Etemadi Fuzzy Linear Regression (EFLR)

Sepideh Etemadi
Isfahan University of Technology

Mehdi Khashei (khashei@in.iut.ac.ir)
Isfahan University of Technology https://orcid.org/0000-0002-2607-2665

Research Article

Keywords: Fuzzy Sets and Systems, Generalizability, Predicting, Fuzzy linear regression, Managerial Decision Making, Reliability-based Modeling

Posted Date: January 3rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-981004/v1

Abstract

Modeling and forecasting are among the most powerful and widely-used tools in decision support systems. The Fuzzy Linear Regression (FLR) is the most fundamental method in the fuzzy modeling area in which the uncertain relationship between the target and explanatory variables is estimated and has been frequently used in a broad range of real-world applications efficaciously. The operation logic in this method is to minimize the vagueness of the model, defined as the sum of individual spreads of the fuzzy coefficients. Although this process is coherent and can obtain the narrowest α-cut interval and exceptionally the most accurate results in the training data sets, it can not guarantee to achieve the desired level of generalization. While the quality of made managerial decisions in the modeling-based field is dependent on the generalization ability of the used method. On the other hand, the generalizability of a method is generally dependent on the precision as well as reliability of results, simultaneously. In this paper, a novel methodology is presented for the fuzzy linear regression modeling; in which in contrast to conventional methods, the constructed models' reliability is maximized instead of minimizing the vagueness. In the proposed model, fuzzy parameters are estimated in such a way that the variety of the ambiguity of the model is minimized in different data conditions. In other words, the weighted variance of different ambiguities in each validation data situation is minimized in order to estimate the unknown fuzzy parameters. To comprehensively assess the proposed method's performance, 74 benchmark datasets are regarded from the UCI. Empirical outcomes show that, in 64.86% of case studies, the proposed method has better generalizability, i.e., narrower α-cut interval as well as more accurate results in the interval and point estimation, than classic versions. It is obviously demonstrated the importance of the outcomes' reliability in addition to the precision that is not considered in the traditional FLR modeling processes. Hence, the presented EFLR method can be considered as a suitable alternative in fuzzy modeling fields, especially when more generalization is favorable.

1. Introduction

The Fuzzy Linear Regression (FLR) is one of the most popular and widely used fuzzy methods to explain, evaluate, and forecast the target variable with aid of explanatory variables in incomplete, imprecise, noisy, linguistic, qualitative, or vague situations. Regularly, the FLR is one of the best methods of modeling when the number of observations for modeling is low, distributional assumptions are violated, the relationship between input and output variables is vague or the phenomenon is ambiguous. This method is based on possibility and fuzzy set theories that instead of error terms, it is included the fuzzy coefficients [1]. possibilistic regression analysis is among the most discussed topic in the field of an FLR model that was presented by Tanaka *et al.* [2]. This approach minimizes the total vagueness of the method, specified as the sum of individual spreads of the fuzzy parameters, subject to the threshold of the estimations to cover the support of the observations for a specific h-level. Also, to estimate the FLR coefficients, a mathematical linear programming problem with symmetric triangular fuzzy coefficients was formulated [3]. The FLR model is used in various applications such as medicine, engineering, finance, environment, management, energy, etc., successfully in the literature.

Černý, M. and Hladík [4] determined the fuzzy parameters of a possibilistic linear regression based on reduction of the fuzzy-valued method to an interval-valued method on a given α-cut in the inflation-consumption application. Hesamian and Akbari [5] introduced the fuzzy penalized method to estimate unknown FLR parameters with exact explanatory variables. The performance of the presented model was assessed in two applied simulation cases. The model was assessed in the prediction of house pricing, also in the prediction of

gasoline mileage in miles per gallon. The outcomes indicated that the developed method has higher precision in comparison with several common fuzzy multiple regression. Zhou *et al.* [6] introduced two types of FLR methods with symmetric and asymmetric triangular fuzzy numbers for parameters in the field of affordable levels of house prices in china. Jiang and Liao [7] proposed a fuzzy least absolute regression that comprises mixed qualitative and quantitative data. The coefficients of the method are achieved by solving a linear programming model. The model was implemented in the house lease price evaluation example. Pandit *et al.* [8] evaluated the efficiency of FLR based on the average width of the forecasting intervals in the agriculture field. The outcomes illustrated that the FLR methodology outperformed the simple and multiple linear regression (MLR) techniques.

Amiri *et al.* [9] predicted the reference evapotranspiration rates for grass crops in greenhouse conditions by the FLR method efficiently. The proposed model can be applied as a guide in determining the irrigation scheduling program. Chen and Nien [10] presented an intuitionistic FLR by considering that the input, output variables, and the coefficients of the method were intuitionistic fuzzy numbers (IFNs). The mathematical programming problem by minimizing the absolute distances between the actual and forecasted was utilized to obtain the optimal IFN coefficients. The efficiency and applicability of the proposed model through some case study in water and soil resources were confirmed. Also, the prediction of slump flow of concrete was accomplished based on the proposed model desirably. Karim and Kamsani [11] predicted the value of the water quality index at selected rivers in Malaysia to evaluate water pollution using FLR efficiently. Chachi [12] formulated the weighted least-squares fuzzy regression for the assessment of suspended load in the field of hydrology engineering. The coefficients of the model were estimated through an iterative re-weighted algorithm.

Khayum *et al.* [13] predicted producing biogas by co-digesting spent tea waste through an FLR approach efficiently. Bajestani *et al.* [14] predicted retinopathy in diabetic patients using a type-2 fuzzy regression model to deal with high levels of uncertainty in medical information. This study can be applied to conduct patients and doctors to decrease the cost of diabetes control and remedy by optimizing the number of check-ups. İçen and Günay [15] utilized a fuzzy expert system in the Monte Carlo approach to estimate the parameter intervals in the FLR. The method was assessed through two case studies that the first related to the prediction of cognitive response times of the control room crew in a nuclear power plant to an uncommon incident and the second one is relevant to acceptance patterns to a hospital with indicators of health service availability and indigency. Liu *et al.* [16] applied FLR with asymmetric triangular fuzzy numbers for parameters in establishing quality function deployment. This model can be used in converting vague and qualitative customer demand into quantitative technical specifications of productions. Gholizadeh *et al.* [17] employed an FLR model considering cutting speed, cutting depth, speed, and tooltip radius as inputs to predict the surface roughness efficiently. This approach can be used to obtain the favorable quality of the machining surface finish in the manufacturing industry.

Rahmawati and Sarno [18] developed FLR for determining the rate of an anomaly in a port container. The proposed model outperformed the support vector regression, radial basis function kernel, and MLR models. Wang *et al.* [19] developed an FLR based on approximate Bayesian computation. An engineering application is applied to show the advantages of the presented model. The results were assessed based on measuring criteria such as distance metric and the degree of the fitting index. Yan *et al.* [20] established FLR to predict the payments per claim and the reserve values for each accident year. The linear regression and fuzzy least square models are applied to calculate the coefficients of the FLR. The outcomes indicated that the linear programming approach is more effective to estimate the FLR coefficients. The proposed payments per claim approach assist

insurance companies to extract more accurate reserves. Ferraro [21] utilized the FLR to assess the progress of reforestation in the environmental domain, as well as to assess students' satisfaction with a course.

Soltani *et al.* [22] predicted the discharge capacity of a Nickel–Metal Hydride battery to assess the remaining useful life by a modified fuzzy c-regression. Hesamian and Akbari [23] established an FLR model with fuzzy targets, fuzzy varying parameters, and exact predictors. To estimate fuzzy parameters, a locally weighted approximation concept, and a popular M-estimator were integrated. The outcomes indicated the efficiency of the presented method in comparison with other approaches, especially, in case studies where outliers exist in the dataset. Also, the effectiveness of the model through two simulation examples was confirmed. Chen and Nien [24] presented an FLR that all coefficients in the model are fuzzy. The model was evaluated by some simulation examples. accordingly, the proposed model based on distance, mean similarity, and credibility measures superior to existing models. Hosseinzadeh and Hassanpour [25] offered an FLR model by considering a set of crisp inputs and regression parameters as Gaussian fuzzy numbers. Also, a nonlinear programming method was constructed based on a new distance between Gaussian fuzzy numbers in order to calculate the regression parameters.

Khammar *et al.* [26] introduced objective function in the form of different loss functions and under the averages of differences between the α-cuts of errors for estimating the coefficients of the FLR model. The model was implemented on some simulated and real-world applications. Jiang *et al.* [27] formulated a multi-objective optimization problem to FLR with crisp inputs that include three objectives as minimizing the mean absolute percentage error of modeling, minimizing the effect of outliers, and minimizing the fuzziness of fuzzy outputs. The superiority of the proposed model over Tanaka's fuzzy regression, Peters' fuzzy regression, fuzzy least-squares regression, and probabilistic fuzzy regression methods in the tea maker design application is obviously confirmed. Chen and Nien [28] built mathematical programming problems based on the least absolute deviations to present intuitionistic fuzzy regression methods with intuitionistic fuzzy coefficients. The performance of the presented methods was assessed based on similarity and distance measures. Outcomes demonstrated that the obtained model superior to the existing approach. Hose and Hanss [29] utilized the least-squares model to estimate the parameters in the FLR.

Although the FLR models are powerful in uncertain modeling and applied successfully in different domains, they are similar to all fuzzy models in the literature that have the same operation logic on the methodology of modeling and estimating unknown parameters. The logic of constructing such fuzzy models is to minimize the vagueness of all fuzzy parameters simultaneously in the training data in order to achieve maximum generalization and better performance in the test data. So, many attempts have been made to minimize the ambiguity of models to increase the generality and effectiveness of obtained results. Although this is a common and logical procedure for measuring the performance of models, it does not take into account the stability and reliability of the results. Accordingly, the generalizability in these types of models is only regarded to be dependent on the precision of training results. Although precision is among the effective factors influencing the model's generalization, it is not the unique factor explaining how to change the model's generalizability.

In fact, due to the varying situations and changing data patterns, the reliability of results should be considered as a key factor in the development of forecasting models. In other words, making stable and reliable results are necessary to adopt appropriate decisions. Hence, reliability is another effective factor in the model's generalizability that should be taken into account. However, Etemadi and Khashei [30] recently have maximized

reliability in the multiple linear regression modeling process to achieve maximum generalizability. In other words, they recently have proposed a new reliability-based casual forecasting statistical modeling methodology. The efficiency of the proposed multiple linear regression based on achieving a higher quality of made decisions in the diverse fields of real-world decision problems was confirmed. Accordingly, precision, as well as reliability of obtained results simultaneously affect the generalization. In spite of the vital role of reliability in making generalizable results, the impact of the reliability is not incorporated in the process of estimating the unknown fuzzy coefficients of the traditional FLR. This issue is especially more prominent in the field of fuzzy causal models. The current study addresses the gap and proposes a reliability-based approach in FLR models to obtain reliable and consistent forecasts. The fundamental idea of the proposed model is minimizing the range of performance deviation to ensure more reliable outcomes.

In fact, significant effort has been made in the literature to increase the generalizability of forecasting models. All of them are based on precision-based logic to increase the generalizability of the results for the unknown test set. However, due to fluctuations and changes in the variables affecting the systems under study, achieving reliable results is another important issue that should be considered but it has been ignored in the traditional precision-based FLR model in the literature. For this purpose, in this paper, based on maximizing reliability to fill the gap and improve decision-making processes, a novel reliability-based methodology is developed for the FLR modeling entitled Etemadi Fuzzy Linear Regression (EFLR) models; which considers reliability in the modeling procedure, and the presented FLR method is constructed according to the concept of maximizing reliability rather than precision in order to obtain more generalization and consequently make more quality decisions in real-world applications.

The core logic of the presented method is based on the point that a model with the most precision may further change in the face by different data conditions in comparison with a reliability-based model. Thus, a reliability-based model may have a more generalization capability, especially in high volatility complex uncertain real-world situations. This goal is achieved by minimizing fluctuations in model performance. In other words, the proposed approach is based on the assumption that reducing changes in model performance leads to enhancing the reliability of the model's results. To implement the proposed method, in the first phase the training data is split into training and validation sets. Then, the fuzzy parameters are determined in a way that the changes in these models' performances are minimized. On the other hand, the most important challenge in using the presented approach is that the complexity of the reliability-based methodology is higher than the classic precision-based methodology. Accordingly, the choice of the Etemadi reliability-based approach seems to be a tradeoff between the cost of modeling and achieving better and more efficient decisions in real-world problems.

To illustrate the appropriateness and effectiveness of the presented Etemadi FLR, the reliability-based modeling performance is evaluated by 74 benchmark data sets with various domains and different characteristics from UCI, and the results are analyzed and compared with the classic FLR model. So, these data sets are modeled with both the proposed reliability-based and traditional fuzzy precision-based models, and then their obtained confidence intervals are compared together. Empirical outcomes illustrate that the presented method has outperformed the classic FLR model in terms of interval width criterion. Accordingly, the role of reliability as an effective factor in the quality of forecasts is definitely confirmed. So, the presented model can be utilized as a suitable alternative for the accuracy-based FLR model in making effective decisions. The remaining paper is systematized in the following fashion. In Section 2, the proposed Etemadi Fuzzy Linear Regression (EFLR) model is mathematically formulated. The description of considered datasets to assess the performance of the

presented method is presented in Section 3. The performance of the presented method and comparing it with the conventional model is discussed in Section 4. Finally, the discussion, future research suggestions, and conclusion are stated in Sections 5 and 6, respectively.

2. The Etemadi Fuzzy Linear Regression (Eflr) Method

In the statistical linear regression models, the concept of the error term is used in order to estimate the unknown parameters by assuming the linear relationship between target and explanatory variables. This parameters estimation process has come from the fact that in such situations the underlying systems and consequently their associated data are considered to be crisp. Thus, the error term can be defined as a function of the difference between actual and fitted values. While in the fuzzy environments, the basic concept is that the residuals between fitted and actual values are not created by measurement errors, rather by the coefficient uncertainty in the model. However, in both crisp and fuzzy versions of linear regression, the performance of the model is maximized in the training data in order to obtain the maximum level of generalization. Although this process is logical and common for achieving the unknown coefficients of linear regression with maximum generalization capability, it is not the only possible way. On the other hand, the reliability of obtained results is another effective factor affecting the generalization that is considered in none crisp as well as FLR.

Accordingly, in this paper, a novel branch of fuzzy linear regression (FLR) models, entitled Etemadi fuzzy linear regression (EFLR) is presented in which the reliability is maximized instead of minimizing the ambiguity. On the other hand, in conventional FLR models, the fuzzy parameters are estimated in such a way that the ambiguity of the model is minimized. While, in the presented method, these fuzzy parameters are estimated in such a way that the variety of the ambiguity of the model is minimized in different data conditions. Therefore, in the proposed model, after dividing the data into training and test data, the training data itself is divided into two training and validation parts. Then, in each iteration of the estimating procedure of the presented model, one data from the validation dataset is added to the training data. In this way, the best values of parameters in the proposed model are values that in adding new data, have the lowest changes.

In general, a k-variable FLR including the target variable and $k-1$ explanatory variables $X_2, X_3, \ldots, X_k$ can be written as follows:

$$\tilde{Y}_t = \tilde{\beta}_1 + \tilde{\beta}_2 X_{2t} + \tilde{\beta}_3 X_{3t} + \ldots + \tilde{\beta}_k X_{kt} = \sum_{k=1}^{k} \tilde{\beta}_i x_i = X' \tilde{\beta} \qquad t = 1, 2, \ldots, N^T \quad (1)$$

where, $\tilde{\beta}_1$ is the intercept, $\tilde{\beta}_2$ to $\tilde{\beta}_k$ are partial slope coefficients, and $N^T = N^{Train} + N^{Validation} + N^{Test}$ is the total number of sample size. Then, the fuzzy triangular numbers are considered for parameters ($\tilde{\beta}_k \; k = 1, 2, \ldots, k$) as follows:

$$\mu_{\tilde{\beta}_k}\left(\beta_k\right) = \begin{cases} 1 - \dfrac{\left|\alpha_k - \beta_k\right|}{c_{1k}} & \alpha_k - c_{1k} \leqslant \beta_k, \\ 1 - \dfrac{\left|\alpha_k - \beta_k\right|}{c_{2k}} & \beta_k \leqslant \alpha_k + c_{2k}, \\ 0 & otherwise, \end{cases} \quad k = 1, 2, \ldots, k \tag{2}$$

where, $\mu_{\tilde{\beta}_k}\left(\beta_k\right)$ is the membership function of the fuzzy set that is shown by the parameter $\tilde{\beta}_k$ $k = 1, 2, \ldots, k$, α_k is the center of the k^{th}fuzzy number, c_{1k} and c_{2k} are the left and right spread around the center of the kth fuzzy number, respectively. After that, applying the extension principle, the membership function of the fuzzy number $\tilde{Y} = X'\tilde{\beta}$ can be described as follows:

$$\mu_{\tilde{y}}\left(y_t\right) = \begin{cases} 1 - \dfrac{\left|y_t - x_t\alpha\right|}{c'_1\left|x_t\right| + c'_2\left|x_t\right|} & for x_t \neq 0, \\ 1 & for x_t = 0, \quad y_t = 0, \\ 0 & for x_t = 0, \quad y_t \neq 0, \end{cases} \quad t = 1, 2, \ldots, N^T \tag{3}$$

where, α and represent vectors of the centers and spreads, respectively. In this way, the total vagueness for the i^{th}data of the validation sample, S_i $i = 1, 2, \ldots, N^{Validation}$, described as the sum of individual spreads of the fuzzy parameters, can be represented as follows:

$$Minimize \quad S_i = \sum_{t=1}^{N^{Train}+i} c'_{1i}\left|x_t\right| + c'_{2i}\left|x_t\right| \qquad i = 1, 2, \ldots, N^{Validation} \tag{4}$$

At the same time, considering the imposed threshold level, $h_i \in \{0, 1\}$ $i = 1, 2, \ldots, N^{Validation}$, the S_i can be achieved by solving the linear mathematical programming as follows:

$$Minimize \quad S_i = \sum_{t=1}^{N^{Train}+i} c'_{1i}\left|x_t\right| + c'_{2i}\left|x_t\right|$$

$$subject.\ to \begin{cases} x'_t\alpha_i + \left(1 - h_i\right)c'_{1i}\left|x_t\right| \geqslant y_t & t = 1, 2, \ldots, N^{Train} + i \quad i = 1, 2, \ldots, N^{Validation} \\ x'_t\alpha_i - \left(1 - h_i\right)c'_{2i}\left|x_t\right| \leqslant y_t & t = 1, 2, \ldots, N^{Train} + i \quad i = 1, 2, \ldots, N^{Validation} \\ c_{1i} \geqslant 0;\ c_{2i} \geqslant 0 & i = 1, 2, \ldots, N^{Validation} \end{cases}$$

(5)

Now, in the proposed method, the weighted variance of different ambiguities in each validation data situation is minimized in order to estimate the unknown fuzzy parameters. In this way, we have that

$$\underset{i}{Minimize} \quad VAR\left(\sum_{t=1}^{N^{Train}+i} c'_{1i}\left|x_t\right| + c'_{2i}\left|x_t\right|\right) = \sum_{i=1}^{N^{Validation}} w_i\left(\sum_{t=1}^{N^{Train}+i} c'_{1i}\left|x_t\right| + c'_{2i}\left|x_t\right| - S\right)^2$$

$$subject.\ to \begin{cases} x'_t\alpha_i + \left(1-h_i\right)c'_{1i}\left|x_t\right| \geqslant y_t & t = 1, 2, \ldots, N^{Train}+i \quad i = 1, 2, \ldots, N^{Validation} \\ x'_t\alpha_i - \left(1-h_i\right)c'_{2i}\left|x_t\right| \leqslant y_t & t = 1, 2, \ldots, N^{Train}+i \quad i = 1, 2, \ldots, N^{Validation} \\ c_{1i} \geqslant 0;\ c_{2i} \geqslant 0;\ \sum_{i=1}^{N^{Validation}} w_i = 1 & i = 1, 2, \ldots, N^{Validation} \end{cases}$$

(6)

where, VAR is the weighted variance function and w_i is the weight of the ith data of the validation sample. Finally, since there is usually no preference between validation data situations; thus, we have that $w_i = 1 \Big/ N^{Validation} \quad i = 1, 2, \ldots, N^{Validation}$. In this way, Eq. (6) will be transformed to Eq. (7) as follows, in which, S_E is the total vagueness, α_E is the vector of the centers, c_{1E} and c_{2E} are vectors of the left and right spreads of the Etemadi method, respectively.

$$Minimize \quad S_E = \sum_{i=1}^{N^{Validation}} \sum_{t=1}^{N^{Train}+i} c'_{1E}\left|x_t\right| + c'_{2E}\left|x_t\right|$$

$$subject.\ to \begin{cases} x'_t\alpha_E + \left(1-h_i\right)c'_{1E}\left|x_t\right| \geqslant y_t & t = 1, 2, \ldots, N^{Train}+i \quad i = 1, 2, \ldots, N^{Validation} \\ x'_t\alpha_E - \left(1-h_i\right)c'_{2E}\left|x_t\right| \leqslant y_t & t = 1, 2, \ldots, N^{Train}+i \quad i = 1, 2, \ldots, N^{Validation} \\ c_{1E} \geqslant 0;\ c_{2E} \geqslant 0 \end{cases}$$

(7)

3. Description Of Datasets

In order to comprehensively assess the presented Etemadi FLR modeling procedure's performance compared with conventional FLR, 74 benchmark datasets in different domains such as medicine, engineering, energy, transportation, social sciences, environment, etc., as well as diverse applications in each domain in the UCI regression section are regarded. These datasets comprised of real or simulated instances from 1987 to 2020. Moreover, the sample size's variation of these datasets is 15 to 434874 data points and the number of explanatory variables in the models is varied from 2 to 85. Also, these cases have several attribute types, covering two main classes of 1) Single and 2) Mixed attributes. The information of 30 datasets, as an example, is reported in Table (1).

4. Empirical Result

Data set #25 entitled "Servo" is chosen in this paper as an example, and implementing the proposed and classic modeling process and comparing the results about it, has been described. This dataset is related to the simulation of a servo system contain a servo amplifier, a motor, a lead screw/nut, and a sliding carriage of some sort [31]. This case involves 167 instances of 4 explanatory variables as two gain settings namely pgain and vgain, as well as two choices of mechanical linkages namely motor and screw (respectively X_1 to X_4) to predict

rise time, or the time needed for the robot to respond to a step-change in a position setpoint (*Y*). Descriptive statistics of the variables are presented in Table (2). The chart of the rise time (*Y*) is displayed in Fig. (1).

Firstly, this dataset is dispart into training and testing subsets. Then, based on the presented modeling process, a portion of the training dataset is dedicated to the validation dataset. In this paper, 85% of data is randomly selected as the training dataset, and 15% of the remaining data is considered as the testing dataset. Also, 10% of the training dataset is randomly chosen as the validation dataset. Afterward, Due to the special process of selecting validation data and the possibility of influencing validation data on the performance of the presented model, to eliminate all the possible effects, each model has been repeated 100 times by using different data for validation. In this way, the estimated functions of the conventional and Etemadi fuzzy linear regression models are obtained as follows, respectively:

$$\begin{aligned} \mathrm{Y} = \ & \langle 1.2895; 7.363225; 1.9790 \rangle - \langle 0.000; 0.15747; 0.000 \rangle X_1 - \langle 0.000; 0.28061; 0.000 \rangle X_2 \\ & - \langle 0.000; 1.46833; 0.000 \rangle X_3 + \langle 0.000; 0.523872; 0.000 \rangle X_4 \end{aligned} \tag{8}$$

$$\begin{aligned} \mathrm{Y} = \ & \langle 1.2775; 7.388273; 1.9567 \rangle - \langle 0.000; 0.17129; 0.000 \rangle X_1 - \langle 0.000; 0.28619; 0.000 \rangle X_2 \\ & - \langle 0.000; 1.4576; 0.000 \rangle X_3 + \langle 0.000; 0.51969; 0.000 \rangle X_4 \end{aligned} \tag{9}$$

Table (1): The general information of the considered case studies from UCI datasets

No.	Title	year	Number of Instances/Attributes	Attribute Type	Field of Application and Domain
1-	Skill Craft	2013	3338/ 19	Mixed: Real, Integer	Skills and Art (Forecasting capabilities)
2-	Metro Interstate Traffic	2019	48204/ 8	Mixed: Real, Integer	Transportation (Metro traffic forecast)
3-	Parkinson Speech with Types of Sound Recordings	2014	520/ 26	Mixed: Real, Integer	Medicine (Parkinson)
4-	Parkinsons Telemonitoring	2009	5875/ 20	Mixed: Real, Integer	Medicine (Parkinson)
5-	SolarFlare	1989	323/ 9	Single: Categorical	Energy (Solar Flare)
6-	Wiki 4HE	2015	904/ 47	Mixed: Categorical, Integer	Education (Applying the Wikipedia in teaching)
7-	Road Network	2013	434874/ 3	Single: Real	Civil Engineering (Road height forecast)
8-	Appliances energy prediction	2017	19735/ 27	Single: Real	Energy (Appliances)
9-	Beijing $PM_{2.5}$	2017	41757/ 11	Mixed: Integer, Real	Environment (Air quality assessment)
10-	Residential Building Data Set	2018	372/ 73	Real	Financial (Forecast construction costs and sales prices)
11-	Carbon Nanotube	2018	10721/ 5	Single: Real	Materials Engineering (Nano)
12-	KEGG Metabolic Network (Undirected)	2011	64608/ 25	Mixed: Integer, Real	Network (Medicine)
13-	KEGG Metabolic Network (Directed)	2011	53413/ 18	Mixed: Integer, Real	Network (Medicine)
14-	Facebook Comment Volume	2016	10044/ 44	Mixed: Real, Integer	Social media (Forecast the number of comments)
15-	Computer Hardware	1987	209/ 7	Integer	Computer (Predicting the performance of CPU)
16-	Gas sensor temperature modulation	2019	295534/ 19	Single: Real	Chemical engineering (Experiments)

No.	Title	year	Number of Instances/Attributes	Attribute Type	Field of Application and Domain
17-	WLAN for indoor localization from wristband	2017	17786/ 7	Mixed: Real, Integer	Telecommunications (Smartphone)
18-	GNFUV Unmanned Surface Vehicles	2018	1649/ 2	Single: Real	Smart vehicles (Marine)
19-	Insurance Company Benchmark	2000	9822/ 85	Mixed: Integer, Categorical	Insurance
20-	Online News Popularity	2015	39644/ 55	Mixed: Integer, Real	Social media (Forecasting the number of subscriptions)
21-	Online Video and Transcoding Time	2015	68784/ 16	Mixed: Integer, Real	Social media (Video)
22-	Parking Birmingham	2019	35717/ 3	Single: Real	City Services (Parking)
23-	$PM_{2.5}$ of Five Chinese Cities	2017	22687/ 13	Mixed: Categorical, Real	Environment (Air quality assessment)
24-	Query Analytics Workload	2019	200000/ 7	Single: Real	Crime (Predicting average beat)
25-	**Servo**	**1993**	**167/ 4**	**Mixed: Integer, Categorical**	**Electrical engineering (System simulation)**
26-	Student Performance	2014	395/ 31	Integer	Educational (Evaluating)
27-	Gas Turbine CO and NOx	2019	7158/ 10	Single: Real	Energy (Forecasting hourly energy of turbine)
28-	Pedestrian in Traffic	2019	2061/ 12	Single: Real	Transportation (Traffic)
29-	clickstream for online shopping	2019	165474/ 11	Mixed: Integer, Real	Online store (Predicting the order price)
30-	CNN-based stock market prediction	2019	1130/ 44	Single: Real	Finance (prediction of S&P indicator)

Table (2): The descriptive statistics of the elected dataset (Servo)

General information	Sample	X_1	X_2	X_3	X_4	Y
Minimum	Training Data	1.000	1.000	3.000	1.000	0.1312
Maximum		5.000	5.000	6.000	5.000	7.1001
Mean		1.1257	1.2430	0.8274	1.1996	1.0178
S. D.		1.3743	1.4269	1.0082	1.3824	1.4958
Skewness		0.1498	0.1748	0.4586	0.4510	2.1736
Kurtosis		-1.1421	-1.2954	-0.8326	-1.0605	3.7999
Minimum	Testing Data	1.000	1.000	3.000	1.000	0.2438
Maximum		5.000	5.000	6.000	5.000	5.1000
Mean		2.5769	2.9231	3.6538	2.1538	2.1675
S. D.		1.6043	1.5211	0.9356	1.2551	1.6960
Skewness		0.4475	0.1399	1.4182	1.0021	0.5821
Kurtosis		-1.4677	-1.4638	1.2506	0.1749	-1.4853

The interval width (IW) of the FLR and EFLR methods, and the improvement percentage of the presented model compared to its conventional version for the Servo case are presented in Table (3). This case's outcomes illustrate that the presented Etemadi FLR can improve the performance of the conventional FLR model. To ensure the consistency of the results and remove the possible effects of data on models' performance, the aforementioned process for the Etemadi and conventional FLR models is implemented for all 74 different case studies in the regression section. The interval width values of the Etemadi FLR and traditional FLR, as well as the improvement percentage of the proposed Etemadi model against the conventional one, for 30 previously selected data sets are reported in Table (4).

Table (3): The performance metrics of the Etemadi and classic FLR methods

Data #25	Interval Width		Improvement
	Etemadi Fuzzy Linear regression	Conventional Fuzzy Linear regression	
Servo	**3.2342**	3.2685	**1.05%**

In this section, the Etemadi and conventional FLR models' results are analyzed and compared together in terms of number and rate of superiority, as well as the interval width improvement percentage in terms of interval estimations. The outcomes of the implementation of 74 datasets illustrate that in 48 cases (64.86% of cases) the Etemadi FLR model compared to the conventional model in terms of interval width had a closer confidence interval. Based on these consequences, it can be inferred that both precision and reliability have an influence on the fuzzy linear regression model's generalization capability, while the importance of reliability is greater than precision. Moreover, it is shown the efficiency of the presented Etemadi FLR in comparison with its conventional version.

In another analysis, the Etemadi fuzzy linear regression model has improved the conventional one in terms of the interval width. The outcomes illustrate that the generalizability of the Etemadi FLR method is improved at 1.3682%, in comparison with classic fuzzy linear regression. These outcomes illustrate that the Etemadi FLR method is slightly superior to the traditional version. These results highlight the importance and influence of the reliability-based fuzzy linear methods on the decisions' quality. Hence, it may be more sensible that in wholly unknown situations or in blindly electing an FLR model for modeling purposes, the Etemadi FLR has been preferred over the classic one.

Table (4): The performance metrics in the precision and reliability-based FLR methods

No.	Title	Interval Width		Improvement Percentage
		Etemadi FLR method	Conventional FLR method	
1-	Skill Craft Master Table	**0.000939**	0.000943	**0.379**
2-	Metro Interstate Traffic Volume	**3827.8464**	3841.1803	**0.347**
3-	Parkinson Speech with Types of Sound Recordings	44.0550	**42.2169**	-4.354
4-	Parkinsons Telemonitoring	**34.7866**	35.2481	**1.309**
5-	Solar Flare	**0.3629**	0.3822	**5.060**
6-	Wiki 4HE	**4.2579**	4.3273	**1.604**
7-	Road Network (Denmark)	61.7252	**61.7031**	-0.036
8-	Appliances energy prediction	313.4679	**310.4242**	-0.980
9-	Beijing $PM_{2.5}$	**236.2951**	246.2549	**4.044**
10-	Residential Building	**188.0655**	197.3205	**4.690**
11-	Carbon Nanotubes	0.0990	**0.0989**	-0.095
12-	KEGG Metabolic Network (Undirected)	0.2684	**0.2683**	-0.055
13-	KEGG Metabolic Network (Directed)	**0.3082**	0.3082	**0.017**
14-	Facebook Comment Volume Dataset	**449.0989**	449.6862	**0.131**
15-	Computer Hardware	**128.6973**	134.0650	**4.004**
16-	Gas sensor temperature modulation	**18.8421**	18.9342	**0.486**
17-	WLAN for indoor localization from wristband	**37.7864**	37.8450	**0.155**
18-	GNFUV Unmanned Surface Vehicles	**2.3273**	2.3436	**0.693**
19-	Insurance Company Benchmark	0.6992	**0.6950**	-0.602
20-	Online News Popularity	69936.5420	**69864.0571**	-0.104
21-	Online Video and Transcoding Time	61.8273	**61.6588**	-0.273
22-	Parking Birmingham	**1381.7846**	1446.3447	**4.464**
23-	$PM_{2.5}$ of Five Chinese Cities	**283.3752**	285.8687	**0.872**
24-	Query Analytics Workloads	0.37193	**0.37192**	-0.003
25-	**Servo**	**3.2342**	3.2685	**1.049**
26-	Student Performance	**6.4958**	6.7651	**3.981**
27-	Gas Turbine CO and NOx	**3.2494**	3.4265	**5.168**

No.	Title	Interval Width		Improvement Percentage
		Etemadi FLR method	Conventional FLR method	
28-	Pedestrian in Traffic	33.4583	**33.2512**	-0.623
29-	Clickstream for online shopping	0.8786	**0.8784**	-0.025
30-	CNN-based stock market prediction	0.4533	**0.4527**	-0.136
Average				**1.039**

5. Discussions And Future Works

The outcomes of the Etemadi reliability and classic precision-based FLR methods are discussed, and some potential suggestions for future researches are expressed.

- Based on the obtained outcomes and analyzes, reliability in addition to precision is an effective factor on the decisions' quality made based on fuzzy causal modeling processes in various domains such as medicine, finance, environment, engineering, energy, transportation, and social media.
- The superiority of the reliability-based models' generalizability over precision-based in all areas of under-study shows that the reliability is more effective than accuracy in improving the quality of a broad range of decision issues. In this regard, it seems that reliability-based modeling methodology with a focus on minimizing performance changes in the face of different data conditions, deal with the uncertainty of decision issues and achieve better results.
- It can be numerically concluded that the Etemadi reliability-based models can further improve the result's generalizability in terms of interval width in comparison with similar precision-based models. Accordingly, by focusing on Etemadi's reliability-based methodology in a broad range of real-world decision-making problems, better quality, and more efficient results can be obtained.
- Based on obtained outcomes, it seems data structure is an influential factor in determining the superiority of proposed FLR modeling methods. Although, to answer the challenge of selecting the reliability-based or precision-based modeling approaches in a given decision-making problem, further research by the large scale of datasets with diverse characteristics is needed. In other words, it seems that the choice between reliability-based modeling approaches or precision-based approaches in various decision-making problems depends on the characteristics of the data such as complexity, uncertainty, unknown patterns, linear and nonlinear structures, etc., as well as the modeling process, that further investigation and research are required.
- Based on the Etemadi reliability-based modeling process, it is determined that the complexity of the reliability-based methodology is higher than the classic precision-based methodology, in other words, the cost of modeling the Etemadi reliability-based approach is higher than the classic one. Accordingly, the choice of the Etemadi reliability-based approach seems to be a tradeoff between the cost of modeling and achieving better and more efficient decisions in real-world problems.

Future research basis can be as follows:

- Proving the effect of the reliability and/or the precision factors on the performance and generalizability of fuzzy causal modeling mathematically.
- Illustrating the significant effect of the reliability and/or the precision factors on the performance and generalizability of fuzzy causal modeling approaches, statistically.
- Investigating the significant effect of the used fuzzy causal model features on the performance and generalizability of the precision and the reliability-based methodologies, statistically.
- Studying the influential variables on the performance and generalizability of reliability and the precision-based methodologies, statistically.
- Implementing the Etemadi approach on various kinds of existing fuzzy models and comparing obtained outcomes.
- Modeling and recognizing features influencing the superiority of precision and reliability-based fuzzy models.
- Proposing a weighting algorithm to identify the relative importance of the reliability and classic precision-based FLR models.
- Developing a specialized allocation approach to diagnosing the proper modeling methodology for predetermined data.
- Considering precision and reliability simultaneously in the modeling procedure and presenting optimal combinations of precision and reliability-based methods.

6. Conclusion

In this paper, a novel Etemadi reliability-based FLR model with a focus on maximizing confidence instead of precision to improve the generalizability of the outcomes is presented. While in the previous researches, merely precision is considered as a prominent factor affecting the generalizability of the FLR models, the outcomes achieved from this paper show the superiority of the generalizability of the reliability-based FLR method over the precision-based one. Accordingly, the proposed reliability-based FLR method and classic precision-based method have been implemented on 74 UCI benchmark datasets with various application domains, then results are compared and analyzed in terms of the interval width. The results demonstrated that in 64.86% of case studies the proposed Etemadi FLR model has a closer confidence interval than the classic model. The proposed method has improved the generalizability of the classic method to 1.3682%. Therefore, it seems that reliability in addition to precision affects the generalizability of the FLR. However, based on the outcomes, the effect of reliability on increasing quality of real-world decisions or generalizability of the FLR model has higher efficiency. This study illustrates the importance and efficiency of the proposed Etemadi FLR method in achieving higher generalizability.

Declarations

Funding: No specific financial support was received to carry out this study.

Conflicts of interest/Competing interests: The authors declare that they have no competing interests.

Consent to participate: The authors declare that they consent to participate in this study.

Consent for publication: The authors declare that they consent for the publication of this study.

Availability of data and material: Data is available from the corresponding author on reasonable request.

Code availability: Code used in the study are available from the corresponding author on reasonable request.

Authors' contributions: All authors have the same contribution to preparing this manuscript.

References

1. Chukhrova N, Johannssen A (2019) Fuzzy regression analysis: systematic review and bibliography. Appl Soft Comput 84:105708. https://doi.org/10.1016/j.asoc.2019.105708
2. Tanaka H, Uejima S, Asai K (1982) Linear Regression Analysis with Fuzzy Model. *IEEE Transactions on Systems, Man, and Cybernetics*, *12*(6), pp. 903-907. *10.1109/TSMC.1982.4308925.*
3. Khashei M, Hejazi SR, Bijari M (2008) A new hybrid artificial neural networks and fuzzy regression model for time series forecasting. Fuzzy sets systems 159(7):769–786. https://doi.org/10.1016/j.fss.2007.10.011
4. Černý M, Hladík M (2018) Possibilistic linear regression with fuzzy data: tolerance approach with prior information. Fuzzy Sets Syst 340:127–144. https://doi.org/10.1016/j.fss.2017.10.007
5. Hesamian G, Akbari MG (2019) Fuzzy lasso regression model with exact explanatory variables and fuzzy responses. Int J Approximate Reasoning 115:290–300. https://doi.org/10.1016/j.ijar.2019.10.007
6. Zhou J, Zhang H, Gu Y, Pantelous AA (2018) Affordable levels of house prices using fuzzy linear regression analysis: the case of Shanghai. Soft Comput 22(16):5407–5418. https://doi.org/10.1007/s00500-018-3090-4
7. Jiang L, Liao H (2020) Mixed fuzzy least absolute regression analysis with quantitative and probabilistic linguistic information. Fuzzy Sets Syst 387:35–48. https://doi.org/10.1016/j.fss.2019.03.004
8. Pandit P, Dey P, Krishnamurthy KN (2021) Comparative assessment of multiple linear regression and fuzzy linear regression models. SN Computer Science 2(2):1–8. https://doi.org/10.1007/s42979-021-00473-3
9. Amiri MJ, Zarei AR, Abedi-Koupai J, Eslamian S (2019) The performance of fuzzy regression method for estimating of reference evapotranspiration under controlled environment. *International Journal of Hydrology Science and Technology*, *9*(1), pp.28-38. *10.1504/IJHST.2019.10017431.*
10. Chen LH, Nien SH (2020) Approach for Establishing Intuitionistic Fuzzy Linear Regression Models Based on Weakest T-norm Arithmetic. *IEEE Transactions on Fuzzy Systems*. *10.1109/TFUZZ.2020.2978457.*
11. Karim SAA, Kamsani NF (2020) Water Quality Index Prediction Using Multiple Linear Fuzzy Regression Model: Case Study in Perak River, Malaysia. Springer Nature
12. Chachi J (2018) A weighted least squares fuzzy regression for crisp input-fuzzy output data. IEEE Trans Fuzzy Syst 27(4):739–748. 10.1109/TFUZZ.2018.2868554
13. Khayum N, Rout A, Deepak BBVL, Anbarasu S, Murugan S (2019) Application of fuzzy regression analysis in predicting the performance of the anaerobic reactor co-digesting spent tea waste with cow manure. *Waste and Biomass Valorization*, pp.1–14. https://doi.org/10.1007/s12649-019-00874-9
14. Bajestani, N.S., Kamyad, A.V., Esfahani, E.N. and Zare, A., 2018. Prediction of retinopathy in diabetic patients using type-2 fuzzy regression model. *European Journal of Operational Research*, *264*(3), pp.859–869. https://doi.org/10.1016/j.ejor.2017.07.046.

15. İçen D, Günay S (2019) Design and implementation of the fuzzy expert system in Monte Carlo methods for fuzzy linear regression. Appl Soft Comput 77:399–411. https://doi.org/10.1016/j.asoc.2019.01.029
16. Liu Y, Han Y, Zhou J, Chen Y, Zhong S (2018) Establishing the relationship matrix in QFD based on fuzzy regression models with optimized h values. Soft Comput 22(17):5603–5615. https://doi.org/10.1007/s00500-017-2533-7
17. Gholizadeh H, Javadian N, Fazlollahtabar H (2018) Fuzzy regression integrated with genetic-tabu algorithm for prediction and optimization of a turning process. The International Journal of Advanced Manufacturing Technology 96(5):2781–2790. https://doi.org/10.1007/s00170-018-1655-0
18. Rahmawati D, Sarno R (2019) Anomaly detection using control flow pattern and fuzzy regression in port container handling. *Journal of King Saud University-Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2018.12.004
19. Wang N, Reformat M, Yao W, Zhao Y, Chen X (2020) Fuzzy Linear regression based on approximate Bayesian computation. Appl Soft Comput 97:106763. https://doi.org/10.1016/j.asoc.2020.106763
20. Yan C, Liu Q, Liu J, Liu W, Li M, Qi M (2019) Payments per claim model of outstanding claims reserve based on fuzzy linear regression. Int J Fuzzy Syst 21(6):1950–1960. https://doi.org/10.1007/s40815-019-00617-x
21. Ferraro MB (2017) On the generalization performance of a regression model with imprecise elements. International Journal of Uncertainty Fuzziness Knowledge-Based Systems 25(05):723–740. https://doi.org/10.1142/S0218488517500313
22. Soltani M, Telmoudi AJ, Belgacem YB, Chaari A (2020) Parameters identification and discharge capacity prediction of Nickel–Metal Hydride battery based on modified fuzzy c-regression models. Neural Comput Appl 32(15):11361–11371. https://doi.org/10.1007/s00521-019-04631-w
23. Hesamian, G. and Akbari, M.G., 2020. A robust varying coefficient approach to fuzzy multiple regression model. *Journal of Computational and Applied Mathematics*, *371*, p.112704. https://doi.org/10.1016/j.cam.2019.112704.
24. Chen LH, Nien SH (2020) A new approach to formulate fuzzy regression models. Appl Soft Comput 86:105915. https://doi.org/10.1016/j.asoc.2019.105915
25. Hosseinzadeh E, Hassanpour H (2021) Estimating the parameters of fuzzy linear regression model with crisp inputs and Gaussian fuzzy outputs: A goal programming approach. Soft Comput 25(4):2719–2728. https://doi.org/10.1007/s00500-020-05331-7
26. Khammar AH, Arefi M, Akbari MG (2021) A general approach to fuzzy regression models based on different loss functions. Soft Comput 25(2):835–849. https://doi.org/10.1007/s00500-020-05441-2
27. Jiang H, Kwong CK, Chan CY, Yung KL (2019) A multi-objective evolutionary approach for fuzzy regression analysis. Expert Syst Appl 130:225–235. https://doi.org/10.1016/j.eswa.2019.04.033
28. Chen LH, Nien SH (2020) Mathematical programming approach to formulate intuitionistic fuzzy regression model based on least absolute deviations. *Fuzzy Optimization and Decision Making*, pp.1-20. https://doi.org/10.1007/s10700-020-09315-y
29. Hose D, Hanss M (2019) Fuzzy linear least squares for the identification of possibilistic regression models. Fuzzy Sets Syst 367:82–95. https://doi.org/10.1016/j.fss.2018.10.003
30. Etemadi S, Khashei M (2021) Etemadi Multiple Linear Regression. Measurement 186:110080. https://doi.org/10.1016/j.measurement.2021.110080

31. Dua D, Graff C (2019) UCI Machine Learning Repository http://archive.ics.uci.edu/ml. Irvine, CA: University of California, School of Information and Computer Science

Figures

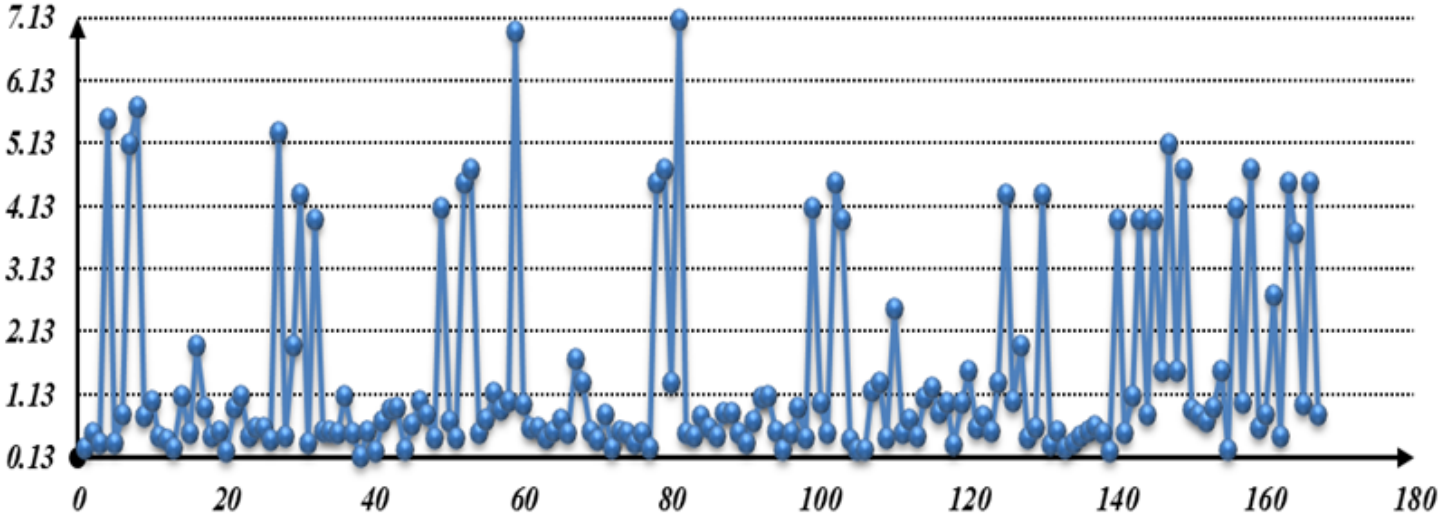

Figure 1

The chart of the target variable (rise time)