Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Computation of Fuzzy Linear Regression Model using Simulation Data

Aliya Syaffa Zakaria[1,*], Muhammad Ammar Shafi[1], Mohd Arif Mohd Zim[2], Nurnadiah Nordin[2], Siti Mukhaiyarah Mahtar[2]

[1] Faculty of Technology Management and Business, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Johor, Malaysia
[2] Department of Medicine, Hospital Al-Sultan Abdullah, UiTM Bandar Puncak Alam, 42300 Puncak Alam, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 21 August 2023
Received in revised form 28 February 2024
Accepted 11 April 2024
Available online 12 May 2024

Keywords:

Fuzzy linear regression; Parameter; Mean square error; Root mean square error; Computation

ABSTRACT

Regression analysis is a powerful technique for determining the causal influence on population outcome, although it is susceptible to outliers. It also oversimplifies the real-world data and problem, as data is rarely linearly separable. This paper analyses the simulation data of a fuzzy linear regression model (FLRM) model, which can aid in minimizing the interference of unwanted information, hence improving the precision of results. The aim of the fuzzy linear regression model (FLRM) is used to determine the best prediction model with the lowest measurement error. The model provides a fundamental mathematical and statistical framework for acknowledging the data imprecision. This study has applied the fuzzy linear regression model (FLRM) to evaluate the data in 15 rows of respondent models. The study implemented measurement error of cross-validation technique which is mean square error (MSE) and root mean square error (RMSE), to enhance data accuracy. Microsoft Excel and Matlab were applied to obtain the result. The simulation result indicates that comparing models with two measurement errors should be used to determine the optimal results. MSE and RMSE value with six parameters of a degree of fitting *H*-value have been calculated in the study. It concludes that a degree of fitting *H*-value of 0.0 is proven as a good model with the lowest MSE and RMSE measurement errors.

1. Introduction

Regression analysis is a statistical method for estimating the relationship between variables that have a cause and an effect. Regression analysis is a potent method for comprehending (including forecasting and explaining) the causal influence on a population's result (Jeon [4]). Nevertheless, regression models are highly sensitive to outliers. A data point that dramatically deviates from most other observations is referred to as an outlier. Variability in measurement may be indicative of an experimental error, and an outlier in regression analysis might cause significant difficulty. Actual world data and problems are also oversimplified by regression analysis models, as data are rarely linearly separable. Fuzzy linear regression provides tools for studying the relationship between variables when certain assumptions of multiple linear regression fail. It also provides a fundamental

* *Corresponding author.*
E-mail address: syaffayila98@gmail.com

https://doi.org/10.37934/araset.45.1.7178

mathematical and statistical framework for acknowledging the imprecision of data. In all statistical studies, researchers seek the most recent techniques for minimizing the statistical measurement error value (Tanaka *et al.,* [10]: Zadeh [11]: Zolfaghari *et al.,* [12]). A wide variety of fuzzy linear models can be used for approximating a linear dependence according to a set of observations in fuzzy regression analysis. (Skrabanek & Marek [8]). It can also aid in reducing the interference of unnecessary information, thereby improving the precision of the results (Kang *et al.,* [3]).

A fuzzy regression model is applied to evaluate the functional relationship between dependent and independent variables in a fuzzy environment. Fuzzy linear regression analysis is also an effective alternative to generally utilized statistics-based regression techniques. Numerous forms of fuzzy regression models are presented in the literature, along with a diversity of estimation techniques for fuzzy model parameters (Pérez *et al.,* [1]). Previous research has explored and developed a variety of applications of fuzzy linear regression and its advantages. A recent study has demonstrated that a fuzzy approach, including fuzzy linear regression, is an appropriate framework for lentil yield management compared to multiple linear regression, which is ineffective as some of its preconditions are not met, as in the data, where some variables were fuzzy numbers (Sorkheh *et al.,* [9]). Furthermore, it clearly shows its ability to deal with the perceptual uncertainties involved in strength prediction issues by providing a standard equation to estimate output values. This leads to accurate predictions of the predicted cement strength values, thereby enhancing the design and providing a valuable modelling tool for the engineering field (Gkountakou & Papadopoulos [2]).

2. Materials and Methods

2.1 Materials

Statistical analysis is adaptable and applicable to various fields, particularly the linear regression technique. Fuzzy linear regression is a form of regression analysis in which fuzzy numbers represent certain model elements. Fuzzy linear functions were proven to be a helpful technique for ambiguous occurrences in linear regression models. The statistical software Matlab and Microsoft Excel were used to analyse the data.

2.2 Methods

2.2.1 Fuzzy linear regression (Tanaka et al., [10])

To formulate a fuzzy linear regression model, the following were assumed to hold (Tanaka *et al.,* [10]):

i. The data can be represented by a fuzzy linear model:

$$\mathrm{Y_e}^* = \mathrm{A_1}^*\mathrm{x_{e\,1}} + \ldots + \mathrm{A_g}^*\mathrm{x_{e\,g}} \triangleq \mathrm{A}^*\mathrm{x_e}, \tag{1}$$

Where,
Fuzzy parameter A_g
Variable of fuzzy parameter x_e
Equation of the fuzzy parameter Y_e^*

$$\mu_{Y_{e^*}}(y) = 1 - \frac{|y_e - x_e^T\alpha|}{\varsigma^T|x_e|} \tag{2}$$

ii. The degree of the fitting of the estimated fuzzy linear model $Y_e{}^{*}$ =$A{*}x_e$ to the given data Y_e = (y_e , ε_e) was measured by the following index h_e , which maximizes h subject to $Y_e^h \subset Y_e^{*h}$ where:

$$Y_e^h = \{y|\mu_{Ye}(y) \geq h\}$$
$$Y_e^* = \{y \,|\mu_{Y_e^*}(y) \geq h\} \tag{3}$$

Which is h -level sets. This index h_e is illustrated in Figure 1. The degree of the fitting of the fuzzy linear model for all data Y_1 , ..., Y_N is defined by *min* $_f$ [h_f].

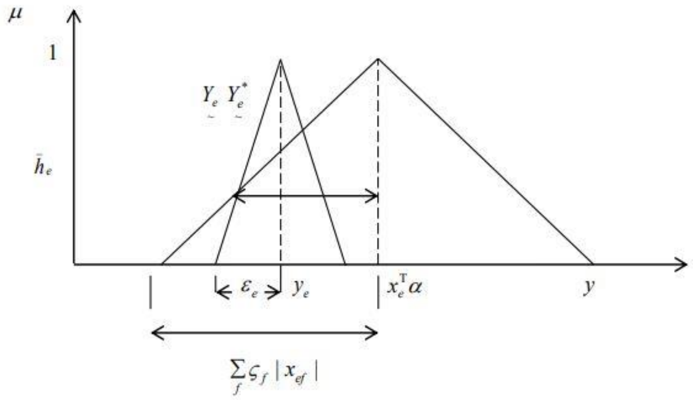

Fig. 1. Degree of fitting of Y_e^* to a given fuzzy data Y_e

iii. The vagueness of the fuzzy linear model is defined by:

$$JJ = \varsigma_1 + \ldots + \varsigma_g \tag{4}$$

The problem was elucidated by acquiring fuzzy parameters A* which minimized *JJ* subject to $\bar{h}$e $\geq H$ for all e , where H was selected by the decision maker as the degree of fit of the fuzzy linear model. The $\bar{h}$e can be acquired by utilizing:

$$\bar{h}e = 1 - \frac{|y_e - x_e^T \alpha|}{\sum_f \varsigma_f |x_{ef}| - \varepsilon_e} \tag{5}$$

Tanaka *et al.,* [10] model estimated the fuzzy parameter $\boldsymbol{A_e^*} = (\boldsymbol{\alpha_e}, \boldsymbol{\varsigma_e})$, which are the solutions of the following linear programming problem:

$$\min_{\alpha,\varsigma} = \boldsymbol{\varsigma_1} + \ldots + \boldsymbol{\varsigma_g}$$

Subject to $\boldsymbol{\varsigma} \geq \boldsymbol{0}$ and

$$\alpha^T x_e + (1-H)\sum_f \varsigma_f |x_{ef}| \geq y_e + (1-H)\varepsilon_e$$
$$-\alpha^T x_e + (1-H)\sum_f \varsigma_f |x_{ef}| \geq -y_e + (1-H)\varepsilon_e \tag{6}$$

The best fitting model for the given data may be obtained by solving the conventional linear programming problem in Eq. (6). The number of constraints, 2 *N*, was generally substantially greater than the number of variables, g. As a result, solving the dual problem of Eq. (6) was easier than solving the primal problem of Eq. (6).

The fuzzy linear regression model (FLRM) can be stated as:

$$Y = A_0\ (\alpha_0\ , \varsigma_0\ ) + A\ (\alpha_1, \varsigma_1)\ x + \ldots + A\ (\alpha_g, \varsigma_g)\ x_g \tag{7}$$

2.3 Methodology

Based on the figure above, fuzzy linear regression was applied in this paper as a method to evaluate the data. There were six degree of fitting H-values from 0.0 until 0.5 has been applied to analyse the centre (a_i) and width (c_i) of each variable using fuzzy linear regression method. The estimated parameter for simulation data will be extracted from the width and centre of each variable. The next step was calculating the errors of each H-value using the standard statistical measurement errors which are mean square error (MSE) and root mean square error (RMSE). The lowest value of error will be chosen as the best prediction model for fuzzy linear regression model in the research.

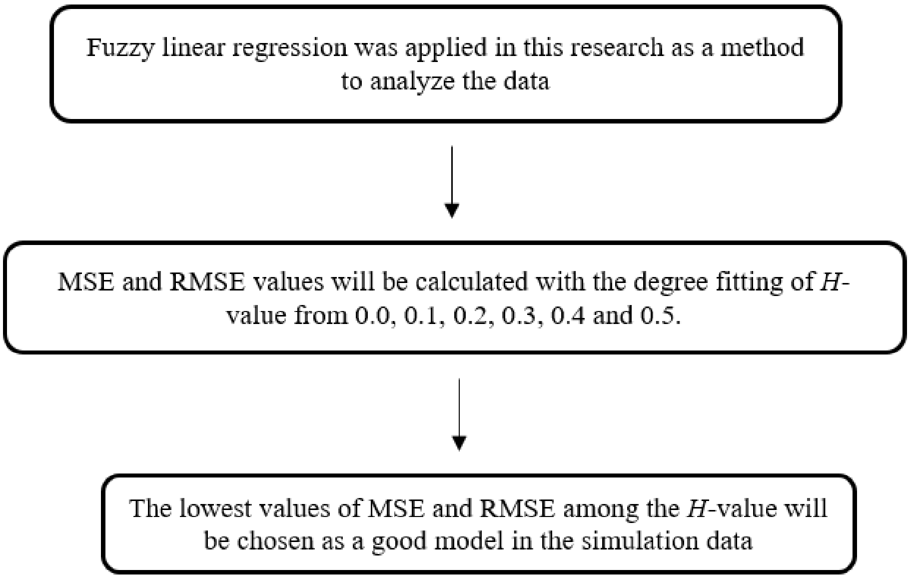

Fig. 2. Step of analysing the data using the fuzzy linear regression model

3. Results and Discussion

Fuzzy linear regression machine model provided 15 rows of data as simulation data. This model was used to study and analyse six predictor variables such as A_1,A_2, A_3,A_4,A_5,A_6. Microsoft Excel and Matlab were applied to obtain the results. The common measurement error of cross-validation technique which are MSE (mean square error) and RMSE (root mean square error), will then be used to acquire the errors of fuzzy linear regression model. The degree of fitting H-values which are 0.0, 0.1, 0.2, 0.3, 0.4 and 0.5 will be compared by calculating the centre, a_i and width, c_i of six variables for each H-value. a_i is center of the fuzzy parameter while c_i is the fuzziness of parameter (width). The centre and the width will then be used to calculate the MSE and RMSE. The degree of fitting H-values with the smallest error value of MSE and RMSE will be the best value of the fuzzy regression model of this simulation data.

Fuzzy linear regression (FLR), proposed by Tanaka *et al.,* [10], was used for predicting manufacturing income. This model was compared by six degree of fitting *H*-values simultaneously. The centre, a_i and width, c_i of six variables for each *H*-value has been portrayed in the tables below.

Table 1
H-value of 0

Variable	Centre (a_i)	Width (c_i)
A_1	1.078	0
A_2	1.078	0
A_3	180.850	18.403
A_4	126.003	0
A_5	-1.803	0
A_6	-187.292	0

The estimated model parameter for simulation data is stated as below:

Ŷ = (1.078, 0) x1 + (1.078, 0) x2 + (180.850, 18.403) x3 + (126.003, 0) x4 + (-1.803, 0) x5 + (-187.292, 0) x6

Table 2
H-value of 0.1

Variable	Centre (a_i)	Width (c_i)
A_1	942.098	0
A_2	1.108	0
A_3	178.834	19.529
A_4	126.182	0
A_5	-1.769	0
A_6	-176.339	0

The estimated model parameter for simulation data is indicated as follows:

Ŷ = (942.098, 0) x1 + (1.108, 0) x2 + (178.834, 19.529) x3 + (126.182, 0) x4 + (-1.769, 0) x5 + (-176.339, 0) x6

Table 3
H-value of 0.2

Variable	Centre (a_i)	Width (c_i)
A_1	806.618	0
A_2	1.137	0
A_3	176.816	20.936
A_4	126.361	0
A_5	-1.736	0
A_6	-165.388	0

The estimated model parameter for simulation data is stated as below:

Ŷ = (806.618, 0) x1 + (1.137, 0) x2 + (176.816, 20.936) x3 + (126.361, 0) x4 + (-1.736, 0) x5 + (-165.388, 0) x6

Table 4
H-value of 0.3

Variable	Centre (a_i)	Width (c_i)
A_1	671.138	0
A_2	1.167	0
A_3	174.799	22.746
A_4	126.539	0
A_5	-1.702	0
A_6	-154.437	0

The estimated model parameter for simulation data is stated as follows:

Ŷ = (671.138, 0) x1 + (1.167, 0) x2 + (174.799, 22.746) x3 + (126.539, 0) x4 + (-1.702, 0) x5 + (-154.437, 0) x6

Table 5
H-value of 0.4

Variable	Centre (a_i)	Width (c_i)
A_1	535.659	0
A_2	1.196	0
A_3	172.781	25.159
A_4	126.718	0
A_5	-1.668	0
A_6	-143.485	0

The estimated model parameter for simulation data is indicated as below:

Ŷ = (535.659, 0) x1 + (1.196, 0) x2 + (172.781, 25.159) x3 + (126.718, 0) x4 + (-1.668, 0) x5 + (-143.485, 0) x6

Table 6
H-value of 0.5

Variable	Centre (a_i)	Width (c_i)
A_1	400.179	0
A_2	1.226	0
A_3	170.764	28.536
A_4	126.897	0
A_5	-1.634	0
A_6	-132.533	0

The estimated model parameter for simulation data is stated as below:

Ŷ = (400.179, 0) x1 + (1.226, 0) x2 + (170.764, 28.536) x3 + (126.897, 0) x4 + (-1.634, 0) x5 + (-132.533, 0) x6

FLR model analyses performance using two statistical measurement errors, including MSE and RMSE. The performance of the two methods would also be assessed using the degree of fit (*H*-value) in Table 7. The lowest error values determine the best model in fuzzy linear regression.

Table 7
MSE and RMSE value

H	MSE	RMSE
0.0	9.993	3.161
0.1	78.861	8.869
0.2	78.651	8.869
0.3	68.125	8.254
0.4	57.600	7.589
0.5	47.07	6.860

Based on the table above, *H*-value of 0.0 has the lowest measurement error, which are mean square error value of 9.993 and root mean square error value of 3.161 compared to other *H*-values. The value of fuzzy parameter is very low and has a value even though the value is 0. The obtained values in fuzzy parameters are shown in Table 1, which includes six predictor variables. The fuzzy mean value of the simulation data can be explained by $\boldsymbol{A_3}$ with the highest fuzzy parameter = 180.850.

4. Conclusion

The purpose of the fuzzy linear regression model (FLRM) is to determine the best prediction model with the lowest measurement error. The results of fuzzy parameter shown that *H*-value of 0.0 is a good prediction model for simulation data in the fuzzy linear regression model as it has the lowest measurement error among other *H*-values. Table 7 displays the summary evaluation of fuzzy linear regression with mean square error and root mean square error. The mean square error for *H*-value of 0.0 was 9.993, and the root mean square error was 3.161. Fuzzy linear regression can be found in various domains in future applications, particularly for inaccurate data. Although only fuzzy linear regression is presented in this paper, another model can be applied by the same approach.

Acknowledgments
This research was supported by the Ministry of Higher Education (MOHE) through the Fundamental Research Grant Scheme (FRGS/1/2021/STG06/UTHM/03/1).

References
[1] Pérez, Lisset Denoda, Gladys Casas Cardoso, J. L. Martínez, Laureano Rodríguez Corvea, and Emilio González Rodríguez. "Fuzzy Linear Regression Models: A Medical Application." *Departamento de Computación, Centro de Estudios de Informática. Facultad Matemática, Física y Computación, Universidad, All content following this page was uploaded by Lisset Denoda on* 9 (2015).

[2] Gkountakou, Fani, and Basil Papadopoulos. "The use of fuzzy linear regression and ANFIS methods to predict the compressive strength of cement." *Symmetry* 12, no. 8 (2020): 1295. https://doi.org/10.3390/sym12081295

[3] Kang, Hao, Ting Ai, Qiuyan Tian, and Qiaoling Zhang. "The Advantage of Fuzzy Regression Analysis and the Establishment of Uml Model." In *2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, pp. 2218-2220. IEEE, 2018. https://doi.org/10.1109/IMCEC.2018.8469757

[4] Jeon, Jihye. "The strengths and limitations of the statistical modeling of complex social phenomenon: Focusing on SEM, path analysis, or multiple regression models." *International Journal of Economics and Management Engineering* 9, no. 5 (2015): 1634-1642.

[5] Shafi, Muhammad Ammar, and Mohd Saifullah Rusiman. "The use of fuzzy linear regression models for tumor size in colorectal cancer in hospital of Malaysia." *Applied Mathematical Sciences* 9, no. 56 (2015): 2749-2759. https://doi.org/10.12988/ams.2015.5175

[6] Shafi, Muhammad Ammar, Mohd Saifullah Rusiman, Shuhaida Ismail, and Muhamad Ghazali Kamardan. "A hybrid of multiple linear regression clustering model with support vector machine for colorectal cancer tumor size prediction." *International Journal of Advanced Computer Science and Applications* 10, no. 4 (2019). https://doi.org/10.14569/IJACSA.2019.0100439

[7] Shafi, Muhammad Ammar, Mohd Saifullah Rusiman, and S. Abdullah. "Application of fuzzy linear regression with symmetric parameter for predicting tumor size of colorectal cancer." *Mathematics and Statistic (1)* (2021): 36-40. https://doi.org/10.13189/ms.2021.090106

[8] Škrabánek, P., and J. Marek. "Models used in fuzzy linear regression." In *Proceedings of the 17th Conference on Applied Mathematics—APLIMAT*, pp. 955-964. 2018.

[9] Sorkheh, Karim, Ahmad Kazemifard, and Shakiba Rajabpoor. "A comparative study of fuzzy linear regression and multiple linear regression inagricultural studies: a case study of lentil yield management." *Turkish Journal of Agriculture and Forestry* 42, no. 6 (2018): 402-411. https://doi.org/10.3906/tar-1709-57

[10] Asai, H. T. S. U. K., S. Tanaka, and K. Uegima. "Linear regression analysis with fuzzy model." *IEEE Trans. Systems Man Cybern* 12 (1982): 903-907. https://doi.org/10.1109/TSMC.1982.4308925

[11] Zadeh, L. A. "Fuzzy sets. Information and Control, Vol. 8, pp. 338-353." (1965). https://doi.org/10.1016/S0019-9958(65)90241-X

[12] Zolfaghari, Zahra Sadat, Mohebbat Mohebbi, and Marzieh Najariyan. "Application of fuzzy linear regression method for sensory evaluation of fried donut." *Applied Soft Computing* 22 (2014): 417-423. https://doi.org/10.1016/j.asoc.2014.03.010