Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

Performance of Linear Programming Asymmetric Parameter Fuzzy Modelling Based on Statistical Error Measurement

Nor Kamariah Kamaruddin[1,*], Muhammad Ammar Shafi[1], Gusman Nawanir[2], Nur Azia Hazida Mohamad Azmi[1], Aliya Syaffa Zakaria[1], Zulfana Lidinillah[1]

[1] Department of Technology and Management, Faculty of Technology Management and Business, Universiti Tun Hussein Onn Malaysia, Malaysia
[2] Faculty of Economics and Business, Universitas Islam Riau, Jl. Kaharudin Nasution, No. 113, Kota Pekanbaru, Riau, Indonesia

ARTICLE INFO

Article history:
Received 8 August 2024
Received in revised form 9 December 2024
Accepted 7 April 2025
Available online 9 May 2025

Keywords:
Fuzzy modeling; comparison models; measurement error; performance of model

ABSTRACT

Modelling the relationship between a scalar answer and one or more explanatory factors using a linear technique is known as linear regression. The problem of using linear regression arises with the use of uncertain and imprecise data. Since the fuzzy set theory's concept can deal with data not to a precise point value (uncertainty data), this study applied the fuzzy linear regression with asymmetric parameter (FLRWAP) to 1000 row of simulation data. Five independent variables with different combination of variable types were considered. Other than that, the performance of the models such as the parameter, error and explanation for the model were included using two measurement statistical errors which is mean square error and root mean square error. FLRWAP found the results of least value of mean square error (MSE) and root mean square error (RMSE) is less than another model with 107.88 and 10.39 respectively.

1. Introduction

Regression analysis is widely used for prediction and forecasting, where its use substantially overlaps the field of machine learning. Regression analysis is also used to understand which one among the independent variables are related to the dependent variable and to explore the forms of these relationships. Hence, regression analysis can be used to infer causal relationships between the independent and dependent variables [1]. However, this can lead to illusions or false relationships, therefore caution about the data that applied is advisable.

Many techniques for carrying out regression analysis have been developed. Familiar methods such as linear regression, fuzzy linear regression and ordinary least squares regression are parametric, in that the regression function is defined in terms of a finite number of unknown parameters that are estimated from the data. Nonparametric regression refers to techniques that allow the regression function to lie in a specified set of functions which may be infinite-dimensional. The performance of regression analysis results depends on the form of the data generating process

* *Corresponding author.*
E-mail address: nkamariah@uthm.edu.my

https://doi.org/10.37934/ard.130.1.126133

and how it relates to the regression approach being used. Since the data generating process is generally good or no missing values, the process or results regression analysis on making assumptions will be acceptable. These assumptions are sometimes testable if a sufficient quantity of data is available. Regression models for prediction are often useful even when the assumptions are moderately violated, although they may not perform optimally.

However, this may happen in many applications, especially with small effects or questions of causality based on observational data [2]. However, regression models are very sensitive to outliers. An outlier is a data point that differs significantly from other observations. The variability in the measurement may indicate experimental error and an outlier can cause serious problem in regression analysis. A researcher found another linear model that is not focus on outliers such as support vector machine (SVM) model. Support vector machine is widely applied to classifying something into a group of objects. In machine learning, SVM is supervised learning models with associated learning algorithms that analyse data for classification and regression analysis. There is a lot of using a support vector machine versus artificial neural network to find the minimum errors and the sigmoid function in both [3]. Vladimir N. Vapnik and Alexey Ya. Chervonenkis were the persons who develop the original SVM algorithm in 1963 and extend the algorithm to non-linear classifier by applying the kernel trick to maximum-margin hyperplanes in 1992.

Studied in University of California at Berkeley introduced the paper on fuzzy sets in 1964 [4]. Among the contents described in the paper are the idea of grade membership was born, sharp criticism from academic community and waste of government funds. Moreover, on 1965 until 1975 continued to broaden the foundation of fuzzy set theory [4]. The concept fuzzy set theory provides a fuzzy multistage decision making, fuzzy similarity relations, fuzzy restrictions and 4 linguistic hedges. Fuzzy logic can be interpreted in a wider sense as theory of fuzzy sets. As such two objectives, fuzzy logic alleviates difficulties in developing and analysing complex systems encountered by conventional mathematical tools and observing that human reasoning can utilize concepts and knowledge that do not have well-defined and sharp boundaries. Such for examples are tall of human, the lighter of lamp and else.

Fuzzy logic is a form of many-valued logic which can be any real number or point number between 0 and 1. In contrast with traditional logic theory, binary sets have two valued logic which is the truth value that ranges in degree between 0 and 1 (true and false). The truth value of fuzzy logic may be at range between completely true and completely false. Furthermore, if linguistic variables are used, specific function degree are also managed [5]. There are various other methods in fuzzy, including the Fuzzy Predator-Prey System [6], the Fuzzy Delphi Method [7], and the Interval-Valued Fuzzy Bézier Curve Interpolation Model [8], which have been widely applied in recent research. These methods have been widely applied to many fields such as aerospace, automotive, business, chemistry industry, financial and medical. It allows getting the approximate values and numbers as well as incomplete and ambiguous data in all fields of fuzzy data. Fuzzy is able to solve incomplete data using controlling and decision-making part.

Other than that, [9] was the first person that developed fuzzy linear regression both the research as well as statistic. In his study, he concerned with the application of fuzzy linear function to a regression analysis in a vague phenomenon. Usually in regression model, deviations between the observed values and the estimated values are supposed to be due to observation errors. It assumed that these deviations of system parameter depend on vagueness of the system structure. The data considered input and output relations whose vagueness the systems structure [10].

There is significant advantage of fuzzy model in analysis which can be used without any assumptions. If the error of data is not normally distributed, then the data still can be used. It is difference with another regression analysis in statistic. Fuzzy logic provides a fundamental

mathematical framework for dealing with vagueness. In addition, it also manages uncertainty and ambiguity in expert evaluations, enabling consensus in intricate decision-making processes [7]. Other significant advantages that fuzzy model effectively handles challenges with high-degree polynomials and demonstrates accuracy in managing uncertainty in mathematical models [6].

Fuzzy regression analysis gives a fuzzy functional relationship between the dependent and independent variables in a vagueness environment. Linear regression is recommended as initial analysis before fuzzy regression analysis to make the 5 more significant decisions in fuzzy data. The input of the fuzzy data may be crisp or fuzzy. There are two types of fuzzy regression models such as Tanaka's linear programming approach and the fuzzy least-squares approach. Several methods have been presented to estimate fuzzy regression models. The first model fuzzy regression was proposed [9] for linear cases by focusing on the extension principle [11].

Fuzzy regression used in complex systems such as in industry, economy, finance, marketing, and ecology function in the real world and it is more imprecision. Such systems require decisions based on human thinking and judgmental and involve human–machine interactions. In such environments, human often not be able to obtain exact numerical data about the system. The nature of information about the complex systems with vagueness is frequently fuzzy. In general, fuzzy regression seems to be intuitively more adequate for real life problems. Therefore, fuzzy regression analysis is more effective for modeling of complex systems. The pioneering work in this field reported that the authors used Zadeh‟s extension principle, A-level procedure, interval arithmetic, and linear programming techniques to develop a fuzzy linear regression analysis. Minimization of these distances in the fuzzy number space with respect to the unknown parameters of regression models leads to solving systems of equations [12].

The fuzzy linear regression method is characterised by its ease of use and transparency in calculation, which distinguishes it from classical regression. However, it is important to note that the differences between the fuzzy linear regression method and classical regression are not significant. Moreover, the findings presented in this study offer empirical evidence in favour of the applicability of fuzzy linear regression prediction, particularly in the context of handling fuzzy data. The application of the fuzzy linear regression method offers enhanced accuracy in finding the least measurement error values. This method surpasses traditional regression analysis by providing more reliable predictions for imprecise data.

This research presents a study on the performance of linear programming asymmetric parameter fuzzy modelling based on statistical error measurement. The primary goal of this study is to find the results of the least value of mean square error (MSE) and root mean square error (RMSE) in order to get precise results. The entirety of this article is organized as follows: Part I focuses on significant studies on underpinning theory and potential determinants, while Part II describes the research methodologies. Section III contains the results and discussion, while Section IV concludes the paper and makes suggestions for future research.

2. Methodology

2.1 Material

This study focused on 1000 row of data as sample for the simulation data. The data including the dependent variable in continuous number and five independent variables. The results were analyzed using SPSS, Microsoft Excel and MATLAB. The flowchart of the study as in Figure 1.

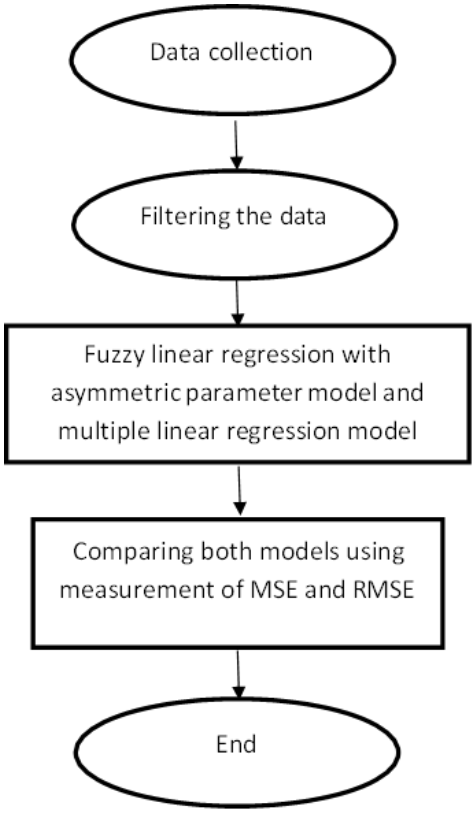

Fig. 1. Flowchart of the study

2.2 Method

2.2.1 Fuzzy linear regression

Statistical analysis is versatile and can be used in any of fields especially with regards of the method of linear regression. Fuzzy linear regression is a fuzzy type of regression analysis in which some elements of the model are represented by fuzzy number. FLRM was an approach explored [13]. In the research, the main objective to estimate values are obtained as fuzzy sets which represent the fuzziness of the system structure, while the conventional confidential interval is related to the observation errors. No assumptions are compulsory in fuzzy model. Fuzzy sets are an effective method for resolving data inaccuracy and ambiguity. Rather than merely classifying a symptom as "present" or "absent," a fuzzy set can be used to quantify the degree to which a patient exhibits a particular symptom. The proposed alternative is preferable to the traditional method for assigning binary values.

The data input and output data whose vagueness is derived from the existence of fuzzy parameters. In the model, the deviations among data are explained as the vagueness of the system structure ex-pressed by fuzzy parameters. To formulate a fuzzy linear regression model, the following Eq. (1) were assumed to hold. The data can be represented by a fuzzy linear model:

$$Y_e^* = A_1^* x_{e_1} + \cdots + A_g^* x_{e_g} \triangleq x_e \tag{1}$$

Where, A_g is Fuzzy parameter, x_e is the variable of fuzzy parameter and Y_e^* is equation of the fuzzy as in Eq. (2).

$$Y_e^* = A^* x_e \tag{2}$$

Where, the given data ^{Y}e = *(ye, εe)* was measured by the following index h_e, which maximizes h subject to $Y_e^h \subset Y_e^{*h}$, where, $Y_e h = \{y | \mu_{Y_e}(y) \geq h\}$ and $Y_e^* = \{y | \mu_{Y_e}(y) \geq h\}$, which are 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 *Y1, …, YN* is defined by min *f* [*hf*] as in Eq. (3). The vagueness of the fuzzy linear model is defined by:

$$JJ = \varsigma_1 + \cdots + \varsigma_g \tag{3}$$

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 applying Eq. (4) and (5):

$$\bar{h}_e = 1 - |ye - xeT\alpha| \tag{4}$$

$$\int f \varsigma f |xef| - \varepsilon e \tag{5}$$

Tanaka [13] model estimated the fuzzy parameter $A_e^* = \alpha_e, \varsigma_e$, which are the solutions of the following linear programming problem in Eqs. (6) – (8).

$$\min_{\alpha,\varsigma} = \varsigma_1 + \cdots + \varsigma_g \tag{6}$$

Subject to $\varsigma \geq 0$ and

$$\alpha^T x_e + (1 - H) \int \varsigma f \left| x_{e_f} \right| \geq y_e + (1 - H)\varepsilon_e \tag{7}$$

$$-\alpha^T x_e + (1 - H) \int \varsigma f \left| x_{e_f} \right| \geq -y_e + (1 - H)\varepsilon_e \tag{8}$$

The best fitting model for the given data may be obtained by solving the conventional linear programming problem above. The number of constraints, 2 *N*, was generally substantially greater than the number of variables, *g*. As a result, solving the dual problem above was easier than solving the primal problem. Fuzzy output denoted as $Y_i = (y_i, e_i)$ where y_i is a center and e_i is a width of fuzzy triangular diagram. The linear function of fuzzy linear regression is shown in Eq. (9).

$$Y = A_0(\alpha_0, \varsigma_0) + A_1(\alpha \quad X_1 + \cdots + A_g(\alpha, \quad x_g \tag{9}$$

Where, $X = (\alpha_i, c_i$ is a vector of independent variables and $A = [A_0, A_1 \ldots . A_g]$ is a vector of fuzzy coefficients presented in the form of a triangular fuzzy number. In FLR, there are fitting models can be fine by the data given and solving the linear programming problem. Other than that, the fuzzy parameter can be found by following the linear programming problem as shown in Eqs. (10) and (11):

$$\alpha^t x_i + (1 - H) \int_j^{\cdot} c_j |x_{ij}| \geq y_i + (1 - H)e_i \tag{10}$$

$$-\alpha^t x_i + (1 - H) \int_j^{\cdot} c_j |x_{ij}| \geq -y_i + (1 - H)e_i \tag{11}$$

2.2.2 Fuzzy linear regression with asymmetric parameter

The issue with fuzzy linear regression asymmetric parameter is there is no proper shape of triangular fuzzy. If $A_i(i = 0,1, \ldots, n)$ is assumed as asymmetrical fuzzy number and xi is crisp real number, then the output of $\widehat{Y}_i = A0 + A1x_i1 + +A_jx_{ij}$ will be triangular fuzzy number [14]. Fuzzy parameter can be determined according to formulation and solution of fuzzy and linear programming of fuzzy linear regression asymmetric parameter and is shown in Eqs. (12) and (13). The membership function of asymmetrically triangular fuzzy number is shown in Figure 2.

$$(1-h)S_0^L + (1+h)\int\left(S_i^L\left|x_{ji}\right|\right) - a_0 - \sum\left(a_ix_{ij}\right) \geq -y_i \quad (12)$$

$$(1-h)S_0 + (1+h)\int\left(S_i\left|x_{ji}\right|\right) + a_0 + \sum\left(a_ix_{ij}\right) \geq -y_i \quad (13)$$

Where, S_0^L, S_i^L are spread left of TFN, S_0, S_i are spread of TFN, α_0, α_i are mode of TFN, and h is degree of TFN.

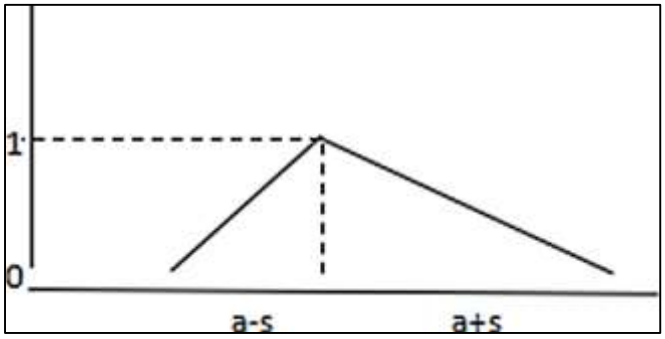

Fig. 2. Membership function of asymmetrically triangular fuzzy number (TFN)

3. Results and Discussion

This study used simulation data consisting of 1000 rows of data as sample, with five variables while continuous number is the dependent variable. The error of the model was done using cross-validation statistical techniques which are MSE and RMSE. This study also was applied multiple linear regression (MLR) and fuzzy modeling [14] as an asymmetric triangular fuzzy towards fuzzy structure. The model was applied in this study to analyze the data on 1000 rows of simulation data. Five of predictor variables were used against dependent variable. The degree of fitting H = 0.5 is also applied in this model based on the reliability and consistency of triangular fuzzy number membership function. The results of analysis are showed Table 1.

Table 1
Results of statistical error of model

Model	*H*	MSE	RMSE
MLR	-	306.257	17.50
FLRWAP	0.5	107.88	10.39

The results of fuzzy parameter found by values of α_i is a center of fuzzy parameter and c_i is fuzziness of parameter (width). The value of width parameter fuzzy is very low close to 0. The obtained values in fuzzy parameters and five predictor variables were included. The fuzzy mean value

of dependent variable can be explained by independent variable (IV) 3 with highest fuzzy parameter = 0.024.

The dispersion can be explained applying the fuzziness parameter. Moreover, the independent variables 1 and 5 show negative values and depends on the strong correlations among the variables. The dependent variable is inversely proportional between IV 1 and IV 5. The summary evaluation of FLRWAP with mean square error and root mean square error are shown in Table 1. Mean square error for this model was 107.88 and root mean square error was 10.39 and describing the accuracy of prediction on model performance. There were 5 determinants to estimate fuzzy linear regression with asymmetric parameter model as below:

$$\hat{Y} = 6.533e\text{-}11 + (-0.003, 0)\ IV1 + (0.001, 0)\ IV2 + (0.024, 0)\ IV3 + (0.015, 0)\ IV4 + (-0.013, 0)\ IV5$$

4. Conclusions

The purpose of this study was to determine the best prediction model with the lowest measurement error between fuzzy linear regression with asymmetric parameter and multiple linear regression models. The results of FLRWAP show that a H-value of 0.5 is a good prediction model for simulation data in the fuzzy linear regression model as it has the lowest measurement error among another model. Table 1 displays the summary evaluation of multiple linear regression and fuzzy linear regression with mean square error and root mean square error. Fuzzy linear regression with asymmetric parameter can be found in various domains in future applications, particularly for inaccurate data. Although only FLRWAP is presented in this paper, another model can be applied by the same approach.

In future studies, other researchers can compare other models with the fuzzy linear regression model to prove that it is the most accurate model with the least measurement errors. The researchers also can add more variables and rows of data in the data analysis. Moreover, many statistical methods in quantitative analysis can be conducted especially the fuzzy linear regression with asymmetric parameter is more popular and effective in predicting model compared to other fuzzy methods. Fuzzy theory and its application are widely applied to various fields, especially in medicine. This study contributes to the application of fuzzy set theory. The fuzzy linear regression with asymmetric parameter is proposed in this study and proves to be the best model with the highest accuracy or the model with prediction in the uncertainty data.

Acknowledgement

This research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) through Tier 1 (vot Q424).

References

[1] Kutner, Michael H., N. J. Nachtsheim, and W. Li, "*Applied Linear Statistical Model*." 2004: Mc Graw Hill. 197-209.
[2] Hammerton, Gemma, and Marcus R. Munafò. "Causal inference with observational data: the need for triangulation of evidence." *Psychological medicine* 51, no. 4 (2021): 563-578. https://doi.org/10.1017/S0033291720005127
[3] Jack, L. B., and A. K. Nandi. "Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms." *Mechanical Systems and Signal Processing* 16, no. 2-3 (2002): 373-390. https://doi.org/10.1006/mssp.2001.1454
[4] Zadeh, L.A. "Fuzzy set." *Information Control* 8 (1965): 338-353. http://dx.doi.org/10.1016/S0019-9958(65)90241-X
[5] Imeni, Mohsen. "Fuzzy logic in accounting and auditing." *Journal of fuzzy extension and applications* 1, no. 1 (2020): 66-72. https://doi.org/10.22105/jfea.2020.246647.1002

[6] Zulkefli, Nor Atirah Izzah, Yeak Su Hoe, and Nor Afifah Hanim Zulkefli. "Fuzzy predator-prey systems by extended runge-kutta method with polynomial interpolation technique." *Semarak International Journal of Machine Learning* 1, no. 1 (2024): 11-19. https://doi.org/10.37934/sijml.1.1.1119

[7] Kamarudin, Nurzatulshima, Nik Mawar Hanifah Nik Hassan, Mohd Mokhtar Muhamad, Othman Talib, Haryati Kamarudin, Norhafizan Abdul Wahab, Aidatul Shima Ismail, Haza Hafeez Borhan, and Nazihah Idris. "Unveiling collaborative trends in Fuzzy Delphi Method (FDM) research: A Co-authorship bibliometrics study." *International Journal of Computational Thinking and Data Science* 2, no. 1 (2024): 1-20. https://doi.org/10.37934/CTDS.2.1.120

[8] Sahrom, Nor Ashikin, and Mohammad Izat Emir Zulkifly. "Interval-Valued Fuzzy Bézier Curve Interpolation Model and its Visualization." *Journal of Advanced Research in Computing and Applications* 36, no. 1 (2024): 61-69. https://doi.org/10.37934/arca.36.1.6169

[9] 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

[10] Nowaková, Jana, and Miroslav Pokorný. "Fuzzy linear regression analysis." *IFAC Proceedings Volumes* 46, no. 28 (2013): 245-249. https://doi.org/10.3182/20130925-3-CZ-3023.00079

[11] Ebrahimnejad, Ali, and Naser Amani. "Fuzzy data envelopment analysis in the presence of undesirable outputs with ideal points." *Complex & Intelligent Systems* 7, no. 1 (2021): 379-400. https://doi.org/10.1007/s40747-020-00211-x

[12] Srinivasan, Arumugam, and G. Geetharamani. "Linear programming problem with interval type 2 fuzzy coefficients and an interpretation for ITS constraints." *Journal of Applied Mathematics* 2016, no. 1 (2016): 8496812. https://doi.org/10.1155/2016/8496812

[13] Tanaka, Hideo. "Fuzzy data analysis by possibilistic linear models." *Fuzzy sets and systems* 24, no. 3 (1987): 363-375. https://doi.org/10.1016/0165-0114(87)90033-9

[14] 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