International Journal of Engineering & Technology, 7 (4.30) (2018) 419-422

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Prediction in a Hybrid of Fuzzy Linear Regression with Symmetric Parameter Model and Fuzzy C-Means Method Using Simulation Data

Muhammad Ammar Shafi [1]*, Mohd Saifullah Rusiman [1*], Kavikumar Jacob [1], Nor Shamsidah Amir Hamzah[1], Norziha Che Him[1], Nazeera Mohamad[1]

[1] *Faculty of Applied Science and Technology, Universiti Tun Hussein Onn Malaysia, 86400, Pagoh Muar, Johor, Malaysia.*
**Corresponding author E-mail: ammar26121991@gmail.com*

Abstract

The objective of fuzzy linear regression model (FLRM) to predict the dependent variable and independent variables in vague phenomenon. In this study, several models such as fuzzy linear regression model (FLRM), fuzzy linear regression with symmetric parameter (FLWSP) and a hybrid model have been applied to be evaluated by 1000 rows in 1 simulation data. Moreover, the hybrid method was applied between fuzzy linear regression with symmetric parameter (FLRWSP) and fuzzy c-mean (FCM) method to get the effective prediction in a new model and best result in this study. To improve the accuracy of evaluating and predicting, this study employ two measurement error of cross validation statistical technique which are mean square error (MSE) and root mean square error (RMSE). The simulation result suggests that comparison among models using two measurement errors should be to determine the best results. Finally, this study notes that the new hybrid model of FLRWSP and FCM is verified to be a good model with the least value of MSE and RMSE measurement errors.

Keywords: *Fuzzy C-means; Fuzzy linear regression; Hybrid Model.*

1. Introduction

Recently, applied linear statistical models have been used in many fields such as medication, economy, social science and many more [1-3]. The main objective of linear statistical analysis is to predict the relationship of a respondent variable in terms of predictor variables in a linear function or multiple linear functions. From a linear regression approach, there are some assumptions that need to be fulfilled by reasearchers [4]. The assumptions are linear relationship, multivariate normality, no auto correlation and homoscedascity. Statistical linear regression model can be applied only if the dependent variables are continuous and distributed according to a statistical model. For fuzzy data, fuzzy membership function must be in line with fuzzy set theory [5, 6].

In constructing a fuzzy and vagueness models, there are three key characteristics of every system model such as complexity, credibility and uncertainty (vagueness) attempt to maximize its usefulness. The relationship is not fully covered by these three characteristics. On the upside, vagueness tends to reduce complexity and increase credibility of the resulting model when vagueness is the main characteristic in modelling. The vagueness is allowable to solve and estimate by developing methods for each modelling problem [7].

Vagueness in modelling are generally acceptable in publication of a seminal paper by Lotfi A. Zadeh in 1965 [8, 9]. In his paper, Zadeh introduced how to deal with vagueness information using a theory of fuzzy. The significance of Zadeh's paper was challenged not only in vagueness modelling theory but also the probability theory. After Zadeh's paper was introduced, modelling in fuzzy area gained more interest especially in prediction of vagueness phenomenon [1]. Tanaka et al (1982) explains fuzziness of respondents or fuzzy uncertainty of dependent variables in fuzzy regression model. Hence, there are three categories data of fuzzy regression model;

i) Non-fuzzy input and output
ii) Non-fuzzy input but fuzzy output
iii) Fuzzy input and output

Recently, there are one model of fuzzy commonly used by professional researchers such as FLRM. This study aims to provide improvements among fuzzy linear regression and fuzzy c-mean method by proposing a new RFLRM approach which is a combination by models. The FLRM, FLRSP and a new hybrid of fuzzy linear regression model have been applied to a simulation data. The comparison among these models was done using measurement errors such as MSE and RMSE [10-13].

2. Research Methodology

2.1. Fuzzy linear regression model (Tanaka, 1982)

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 by Hideo Tanaka in 1982. In the research, the main objective to estimate values are obtained as fuzzy sets which represent the fuzziness of the system structure,

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

while the conventional confidential interval is related to the observation errors. No assumptions are compulsory in fuzzy model.
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 expressed by fuzzy parameters [9].
Fuzzy output denoted as $Yi = (y_i, e_i)$, where y_i is a center and e_i is a width of fuzzy triangular diagram. The linear function of fuzzy linear regression as;

$$Y = A_0(\alpha_0, \varsigma_0) + A_1(\alpha_1, \varsigma_1)\, X_1 + \ldots + A_g(\alpha_g, \varsigma_g)\, x_g \quad (1)$$

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 coefficient presented in form of triangular fuzzy number. In FLR, there are fitting model can be fine by the data given and solving the linear programming problem. Other than that, the fuzzy parameter can be fined by following linear programming problem:

$$\begin{aligned} &\alpha^t xi + (1-H)\sum_j c_j \mid x_{ij} \mid \geq y_i + (1-H)e_i \\ &-\alpha^t xi + (1-H)\sum_j c_j \mid x_{ij} \mid \geq -y_i + (1-H)e_i \end{aligned} \quad (2)$$

3.1. Fuzzy linear regression with symmetric parameter (Zolfaghari, 2014)

Fuzzy linear regression with symmetric parameter (FLRWSP) is one of the most model uses by professional researcher in fuzzy phenomena. Fuzzy linear regression with symmertric parameter represent some conditions in vagueness and unclear. In the study, the researcher applied fuzzy linear regression to determine the quality of food products especially fried donut. From the science and engineering point of view, the theory of model is useful for conceptual framework and results that can be directly applied in models of systems using fuzzy approach and recent development in fuzzy logic [14].
If $\tilde{A}_i$ $(i = 0, 1, \ldots, n)$ is symmetrical fuzzy number and x_i is crisp real number, and will be a triangular fuzzy number and defined as, $\hat{Y}=(f^c(x), f^3(x))$ when $f^c(x)$ is the mode and $f^3(x)$ is spread of triangular fuzzy number.
The model of FLRWSP can be write as follow:

$$\begin{aligned} f^s(x) &= s_0 + s_1 x_1 + \ldots + s_n x_n \\ f^c(x) &= a_0 + a_1 x_1 + \ldots + a_n x_n \end{aligned} \quad (3)$$

Target function is defined in symmetric condition of triangular fuzzy number as follows:

$$\begin{aligned} &(1-h)s_0^L + (1-h)\sum_{i=1}^{n}(s_i^L \mid x_{ji} \mid) - a_0 - \sum_{i=1}^{n}(a_i x_{ji}) \geq -y_j \\ &(1-h)s_0 + (1-h)\sum_{i=1}^{n}(k_i s_i^L \mid x_{ji} \mid) + a_0 + \sum_{i=1}^{n}(a_i x_{ji}) \geq -y_j \end{aligned} \quad (4)$$

4.1. Fuzzy c-means methods

Fuzzy c-means (FCM) is a clustering method that allows one set of data belongs to more than one cluster. Dunn (1973) was developed this method and then improved by Bezdek (1981). Fuzzy c-means method also frequently used in pattern recognition and then this algorithm is based on minimization fuzzy c-means toward the following objective function or criterion such as:

$$J = \sum_{q=1}^{N}\sum_{r=1}^{C} u^2_{qr} d^2_{qr} \quad (5)$$

Where z is any real number greater than 1, μ_{qr} is the membership values, d_{qr} represent as the distance according to Euclidean. N is the number of objects and C is the number of clusters. The index q $(q=1,\ldots, N)$correspond to object number q and the index $r(r=1,..,C)$ to cluster number r. In case of Euclidean distance the algorithm for minimising J can be summarized by the following steps.

1) Randomly select cluster centers 'c'. Choose the termination tolerance between 0 and 1, then fuzziness exponent, $z > 1$.

2) Update distance, d_{qr} for given μ_{qr} by computing the weighted average for each group and the Euclidean distance as,

$$d^2_{qr} = \| x_q - v_r \|^2 \qquad v_r = \frac{\sum_{q=1}^{N} u_{qr}{}^2 X_q}{\sum_{q=1}^{N} u_{qr}{}^z} \quad (6)$$

3) Update membership values as,

$$u_{qr} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{qr}}{d_{qk}}\right)^{\frac{2}{z-1}}}, \text{ for } z = 1 \quad (7)$$

4) Calculate the objective or criterion J and make a iteration in order to minimize the objective function. The iteration repeated for k = 1,2,.., ∞, then stop the iteration, else repeated step 2.

5.1. Structure and procedure of a hybrid model

The hybrid is defined as a combination of both fuzzy linear regression model proposed by Zolfaghari (2014) and fuzzy c-means method. There are steps to produce the hybrid in Figure 1:

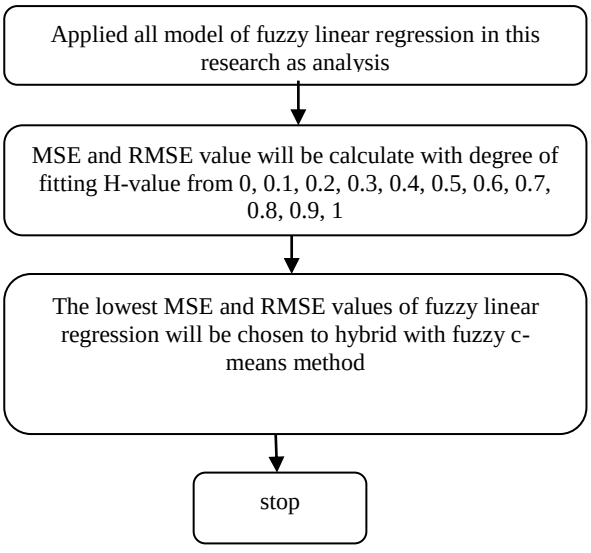

Fig. 1: Steps of hybrid model

3. Results

This study used primary data and the data consists of 1000 as respondents for the models. Dependent variable or outcome is continuous variable. Actually, there were five independent variables in binary and continuous values. The software used to get the results was Microsoft excel, matlab and social science package (SPSS). The comparison among these models was done using cross validation statistical technique which can be known as MSE and RMSE,. However, this comparison is to find the best model in predicting based on MSE and RMSE.

2.1. Fuzzy linear regression

FLR model proposed by Tanaka (1982) used to predict manufacturing income. This model evaluates by two measuring performance such as MSE and RMSE. This measuring performance of the two methods could also be evaluated by degree of fitting (H-Value) in Table 1. The smallest error values become the best model in FLR.

Table 1: MSE and RMSE values

H	MSE values	RMSE values
0	76272.83	276.175
0.1	55041.97	234.610
0.2	9069.75	95.235
0.3	125851.3	354.755
0.4	82564.3	287.340
0.5	103601.9	321.872
0.6	29958.89	173.086
0.7	25023.84	158.189
0.8	32887.64	181.348
0.9	50716.69	225.203
1	1145.919	33.851

The FLR model with H-value = 1 as below:

$\hat{Y}$ = (-2.943, 0) x1 + (5.925, 0) x2 + (4.778, 0) x3 + (-1.410, 0) x4 + (6.131, 0) x5

3.1 Fuzzy linear regression with symmetric parameter

FLRWSP model was proposed by Zolfaghari (2014). This model evaluates by two measuring performance such as MSE and RMSE. This measuring performance of the two methods could also be avaluated by fegree of fitting (H-Value). The smallest error values become the best model in FLR. This model was applied for the study in simulation data. The result for measurement error was in Table 2.

Table 2: MSE and RMSE values

H	**MSE values**	**RMSE values**
0	717.760	26.791
0.1	875.538	29.589
0.2	988.420	31.439
0.3	1116.306	33.411
0.4	1388.772	37.266
0.5	1668.34	40.845
0.6	1972.35	44.411
0.7	2316.383	48.128
0.8	2674.885	51.719
0.9	3642.536	60.353
1	1148.919	33.895

The FLRWSP model with H-value = 0 as below:

$\hat{Y}$ = (-0.397, 0) x_1 + (7.293, 0) x_2 + (3.281, 1.380) x_3 + (1.282, 0) x_4 + (2.887, 0) $x5$

3.1. A hybrid of reflective fuzzy linear regression and fuzzy c-means methods

There is no assumption needs for hybrid model of FLRWSP and FCM. Data was analysed by using Matlab software to find FCM clustering. Furthermore, the independent variables x_3, x_5, x_2, x_4 and x_1 are chosen since it has the highest correlation value. The data will be divided into two cluster, cluster 1 and cluster 2.The correlation values among x_3, x_5, x_2, x_4 and x_1 as below Table 3:

Table 3: Correlation values y vs x

Correlation	Value	Significant
Y vs x_3 (1)	0.682	0.000
Y vs x_5 (2)	0.654	0.000
Y vs x_2 (3)	0.547	0.000
Y vs x_4 (4)	0.265	0.000
Y vs x_1 (5)	0.006	0.852

Based on five correlation chosen, dependent, y vs independent ,x_1 is the best cluster among others. MSE final = (MSE cluster 1 + MSE cluster 2) / 1000. The amount of respondents in the simulation data taken with 1000. The results as follow Table 4:

Table 4: MSE Values

Correlation	MSE Cluster 1	MSE Cluster 2	Final MSE
Y vs x3 (H=0.0)	817.6843553	485.5865002	1.3033
Y vs x5 (H=0.0)	378.5410514	499.5570361	0.8781
Y vs x2 (H=1.0)	555.5880901	1256.925833	1.8125
Y vs x4 (H=0.0)	828.0517756	658.4090994	1.4865
Y vs x1 (H=0.2)	23.96675899	739.1967409	0.7632

Cluster 1 (Y vs X1) for hybrid model

Cluster 1 (Y vs X1) for the hybrid model between FLRWSP and FCM used 5 data as respondent in the analysis. The parameter of the model cluster 1 as follow:

$\hat{Y}$ = (4.176, 0) x1 + (18.720, 0) x2 + (1.339, 1.380) x3 + (-1.578, 0) x4 + (-7.463, 0) x5

Cluster 2 (Y vs X1) for hybrid model

Cluster 2 (Y vs X1) for the hybrid model between FLRWSP and FCM used 995 data as respondent in the analysis. The parameter of the model cluster 1 as follow:

$\hat{Y}$ = (-1.28, 0) x1 + (3.57, 0) x2 + (3.79, 1.380) x3 + (2.05, 0) x4 + (3.32, 0) x5

4. Conclusion

The simulation results of this study suggest that y-respondent variable was continuous. There are FLRWSP and FCM were applied using hybrid method. A new model was approach to represent for a vague phenomenon. Based on two models of measurement errors, a hybrid between FLRWSP and FCM is the greatest model can be used in any of researcher field. Moreover, analysis among MLR, FLRM, FLRWSP and hybrid FLRWSP and FCM can be concluded that the best model with the smallest value of measurement errors MSE and RMSE. The summary of the models in Table 5.

Table 5: MSE and RMSE values in each model

Model of linear regression	MSE	RMSE
MLR	271.664	16.482
FLR	1145.919	33.851
FLRWSP	717.760	26.791

Hybrid	0.7632	0.874

The hybrid of fuzzy linear regression and fuzzy c-means method can be applied in various fields in future especially for vagueness information. Although only fuzzy linear regression and fuzzy c-means are discussed in this study, another model can be used by the same approach.

Acknowledgement

The authors would like to express gratefully heartfelt thanks to the Universiti Tun Hussein Onn Malaysia and Office for Research, Innovation, Commercialization and Consultancy Management (ORICC) for the financial support under the TIER 1 research grant (H232).

References

[1] Shafi MA, Rusiman MS (2015), The use of fuzzy linear regression models for tumor size in colorectal cancer in hospital of Malaysia. *Applied Mathematical Sciences* 9, 2749-2759.
[2] Shafi MA, Rusiman MS, Nor ME, Khamis A, Abdullah SNS, Azmi MS, Abidin MSZ, Ali M (2018), The factors that influence job satisfaction among royal Malaysian customs department employee. *Journal of Physics: Conference Series* 995, 12-16.
[3] Shafi MA, Rusiman MS, Hamzah NSA, Nor ME, Azmi NAHM, Latip MFA, Azman AH (2018), The analysis of morphometric data on rocky mountain wolves and artic wolves using statistical method. *Journal of Physics: Conference Series* 995, 12-15.
[4] Michael HK, Christopher JN, John N, William Li (2004), Applied linear statistical models. *Fifth edition* MC GRAW HILL.
[5] Tanaka H, Uejima S & Asai K (1982), Linear regression analysis with fuzzy model. *IEEE Transactions On Systems, Man and Cybernetics* 12, 903-907.
[6] Tanaka H (1987), Fuzzy data analysis by possibilities linear models. *Fuzzy Set Syst* 24, 363-375.
[7] George JK, Bo Y (2008), Fuzzy sets and fuzzy logic. *Prentice hall of india, new delhi*, 110001.
[8] Zadeh LA (1965), Fuzzy Sets. *Inform Control* 8, 338-358.
[9] Zadeh LA (1975). The Concept Of A Linguistic Variable And Its Application To Approximate Reasoning. Information Sciences, Part I : No. 8, pp. 199-249 ; Part II : No. 8, pp. 301-357 ; Part III : No. 9, pp. 43-80
[10] Bin MAS, Bin MSR, Yusof NSHC (2014), Determinants status of patient after receiving treatment at intensive care unit: a case study in Johor Bahru. *Computer, Communications, and Control Technology (I4CT), 2014 International Conference on*, 80 – 82.
[11] Shafi MA, Rusiman MS, Hamzah NSA, Sufahani SF, Khamis A, Azmi NAHM (2018), Fuzzification of quantitative data to predict tumour size of colorectal cancer. *Far East Journal of Mathematical Sciences* 103, 951-959.
[12] Shafi MA, Rusiman MS, Razali M, Asyikin SN, Mohamad M (2015), The effective model of linear regressions for colorectal cancer stages in general hospital: a case study in Kuala Lumpur. *Australian Journal of Basic and Applied Sciences* 9, 7-10.
[13] Shafi MA, Rusiman MS, Hamzah NSA, Nasibov EN, Azmi NAHM (2016), The use of linear statistical model to predict tumour size of colorectal cancer. *Journal of Science and Technology* 8.
[14] Sadat ZZ, Mohebbeat M & Marzieh N (2014). Application of Fuzzy Linear Regression Method for Sensory Evaluation Of Fried Donut. *Applied Soft Computing* 22, 417-423.