International Journal of Engineering & Technology, 7 (4.33) (2018) 78-82

SPC

International Journal of Engineering & Technology

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

Research paper

Multivariate Matrix for Fuzzy Linear Regression Model to Analyse The Taxation in Malaysia

Noor Hidayah Mohamed Isa*, Mahmod Othman, Samsul Ariffin Abdul Karim

Fundamental and Applied Sciences Department, Universiti Teknologi Petronas, 32610 Seri Iskandar, Perak, Malaysia
**Corresponding author E-mail: noor_16001610@utp.edu.my*

Abstract

A multivariate matrix is proposed to find the best factor for fuzzy linear regression (FLR) with symmetric triangular fuzzy numbers (TFNs). The goal of this paper is to select the best factor influence tax revenue among four variables. Eighteen years' data of the variables from IndexMundi and World Bank Data. It is found that the model is successfully explained between independent variables and response variable. It is notices that $(HR = 0.66)$ sixty-six percent of the variance of tax revenue is explained by Gross Domestic Product, Inflation, Unemployment and Merchandise Trade. The introduction of multivariate matrix for fuzzy linear regression in taxation is a first attempt to analyses the relationship the tax revenue with the independent variables.

Keywords: *Multivariate; Matrix; Fuzzy Linear Regression; Tax Revenue.*

1. Introduction

Tax is a way of collecting revenue from citizens, companies or other entities. It is considered as the most important sources to ensure the economy of a country grow and it is also one of the oldest phenomena that provides essential services with the cost. Taxes have an important function to determine the direction of social system and economic of a country. The purposes of the taxes are to develop the economy for every country to give the protection for the local product and others.

In Malaysia, the tax system was introduced by British Colonial government in the year of 1948 [10]. The taxation in Malaysia is important to finance government expenditure that contributes significantly to the Gross Domestic Product (GDP) of the country. In addition, the tax payment is different for every population. In Malaysia, every person who had stayed in this country for more than 182 days, will be subjected to the payment of the taxes regardless of nationally. Besides, the people who stayed less than 182 days are also subject to payment of taxes but it will be on a different scale.

However, there are several types of taxes which are direct taxes and indirect taxes. The role of collection of direct taxes is carried out by the Inland Revenue Board (IRB) and for the indirect taxes is the Royal Malaysian Customs Department (KDM). Direct tax can be defined as tax that cannot be transferred to another person which means the person must directly pay to the government. Another type of tax is indirect taxes, which is collected as an intermediary from the customer such as sales tax [10]. The total tax revenue is always increasing and decreasing almost every year. The changes of value of tax revenue are affected by a few factors.

Statistical regression and fuzzy regression model are two types of models with different philosophies [18]. In statistical regression, difference between the actual values and the estimates value are due to random errors. However, in fuzzy regression model the differences are attributed to the vagueness of the model structure [18].

In spite of the widespread use of the statistical regression in daily life activities, there exists uncertainty in variables. To handle the uncertainty, fuzzy linear regression model was introduced. As conventional statistical regression is unable to handle on subjective judgment results which can be non-crisp or linguistic. It is difficult to make a good prediction on a problem involving relationship of factors since the relationship does not usually come out with good prediction. Fuzzy linear regression model is used to obtain an appropriate linear relation between a response variable and independent variables.

Fuzzy linear regression model is able to cope with fuzzy data or linguistic variables. It was stemming from [19] thought fuzzy that was able to deal with ambiguity. The fuzzy regression model divided into two classes, firstly is based on the possibility concept and the second class is the least-squares approach [6-7]. There were many improvements of the fuzzy linear regression model and its applications as well. The most of existing paper on the fuzzy regression model have used the least squares method to construct the fuzzy regression model. However, the least squares method is so sensitive to outliers.

In [16] firstly proposed the fuzzy linear regression model and it is becoming popular among of researchers. This model is restricted to symmetric triangular fuzzy numbers but in [5] developed a new model which is fuzzy least-squares regression model to overcome this limitation. For their studies, the regression coefficients are derived from nonlinear programming problem that requires considerable computations. Then, in [15] introduced a matrix-driven fuzzy linear regression model as some of their efforts to upgrade computational efficiency. The method has been successfully tested to engineering study of estimating bridge performance and car sales volume and also dimension of health related quality of life [11-12].

From previous research, the successfully and the low computational risk of the model and silent attempt to perform the method

in modeling of the parameter of the taxation in Malaysia. Thus, this paper intends to extend the fuzzy linear regression model to identify the most factors among four variables that affects the changes of the tax revenue The introduction of multivariate matrix for fuzzy linear regression in taxation is a first attempt to analyses the relationship the tax revenue with the independent variables. The independent variables are gross domestic product (GDP), inflation, unemployment and merchandise trade while dependent variable is tax revenue. This could help the government to focus on those variables to ensure the tax revenue collection is always constant and stable.

2. Methodology

In this section, the multivariate matrix for fuzzy linear regression model for taxation is introduced. The model based on matrix-driven are discussed. The succeeding section presents the data collection, method of data analysis, multivariate matrix fuzzy linear regression model and the fuzzy coefficient of determination is given in the final section.

2.1. Data ccollection

There are two types of data which are primary and secondary data. In this study, the secondary data is used to solve the problem. The data is retrieved from the website of The World Bank Data [17] and IndexMundi [9] in year 1996 until 2013 for factors that selected which are GDP, inflation, unemployment and merchandise trade. The four factors were identified as important and can support the success of the taxation analysis.

2.2. Method of data analysis

The general form of the model can be written as in (1).

$$\tilde{Y}_i = f(X, A) = \tilde{A}_0 + \tilde{A}_1 X_{i1} + ... + \tilde{A}_n X_{ik} \quad (1)$$

The function $f(X, A)$ is consider which is mapped from X into Y with the elements of X denoted by $x_i = (x_{i0}, x_{i1}, \ldots, x_{ik})$, ith which represents the independent or input variables of the model. The dependent variable or response variables are denoted as Y_i in Y, where $\tilde{A} = (\tilde{A}_0, \tilde{A}_1, ..., \tilde{A}_n)$ are the model regression coefficients. If $\tilde{A}_j (j = 0,1,\ldots n)$ are given as fuzzy sets, the model $f(X, A)$ is called a fuzzy model.

$\tilde{Y}_i$ is the estimated fuzzy response variable or non- fuzzy output data. $\tilde{A}_j$ is fuzzy coefficient in terms of symmetric fuzzy numbers, which can determine by solving FLR model [8]. The fuzzy components were assumed as triangular fuzzy numbers (TFNs).

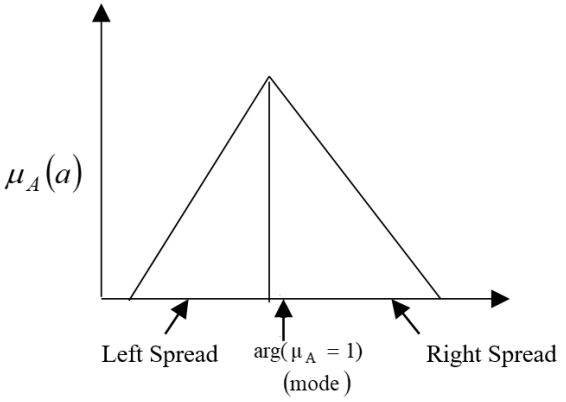

Fig. 1: Fuzzy Coefficient

Figure 1 shows the fuzzy coefficient of TFN that involve its centre or modal value, left and right spreads. Observed data is represented by asymmetrical TFN that defined by a triplet $\tilde{Y}_i = (a_j, c_j^L, c_j^R)$. Moreover, if left and right are equal, the function left and right are equivalent, then the TFN is known as a symmetrical TFN (STFN). In this study, the data that was used in is symmetrical triangular fuzzy numbers.

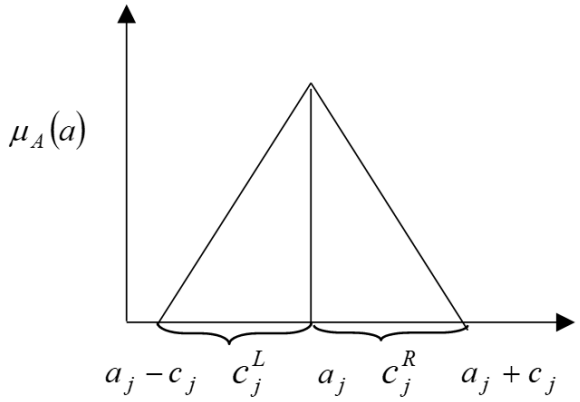

Fig. 2: Symmetrical Fuzzy Parameter

Since the data was a symmetrical TFNs, thus considered $c_j^L = c_j^R = c_j$, Figure 2 shows the membership function for the symmetrical fuzzy regression coefficients, $A_j = (a_j, c_j)$. The response variable represented by symmetric $\tilde{Y}_i = (a_j, c_j)$, whereas a_j is the centre point that represents the original value of data and c_j represents for the spread.

Then, the values of a_j and c_j are substitute in (3) of tax revenue, $\tilde{Y}$. The value of the spread, c_j is obtained by calculating in (2)

$$c_{n,j} = \frac{(Y_i - \hat{Y})}{2} \quad (2)$$

where $Y_i (i = 1,\ldots 18)$ stands for the value of tax revenue and $\hat{Y}$ stands for the mean of the data value of Y_i. In this method, four independent variables and one fuzzy response variable. Then, the fuzzy linear regression is estimated in (3),

$$\tilde{Y}_i = \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \quad (3)$$

where $\beta_0 = a_0, c_{0,j}$ is the fuzzy intercept coefficient, $\beta_1 = a_1, c_{1,j}$ $\beta_2 = a_2, c_{2,j}$ $\beta_3 = a_3, c_{3,j}$ and $\beta_4 = a_4, c_{4,j}$ are the fuzzy slope coefficient for X_i, $i = 1,2,3,4$. X_1 represent GDP, X_2 represent inflation, X_3 represent unemployment, X_4 represent merchandise trade.

2.3. Multivariate matrix fuzzy linear regression model

The matrix algebra is employed to solve the method of fuzzy regression analysis. The variables are illustrated into the multivariate matrix. This model is taken from [15] that was simplified into several steps.

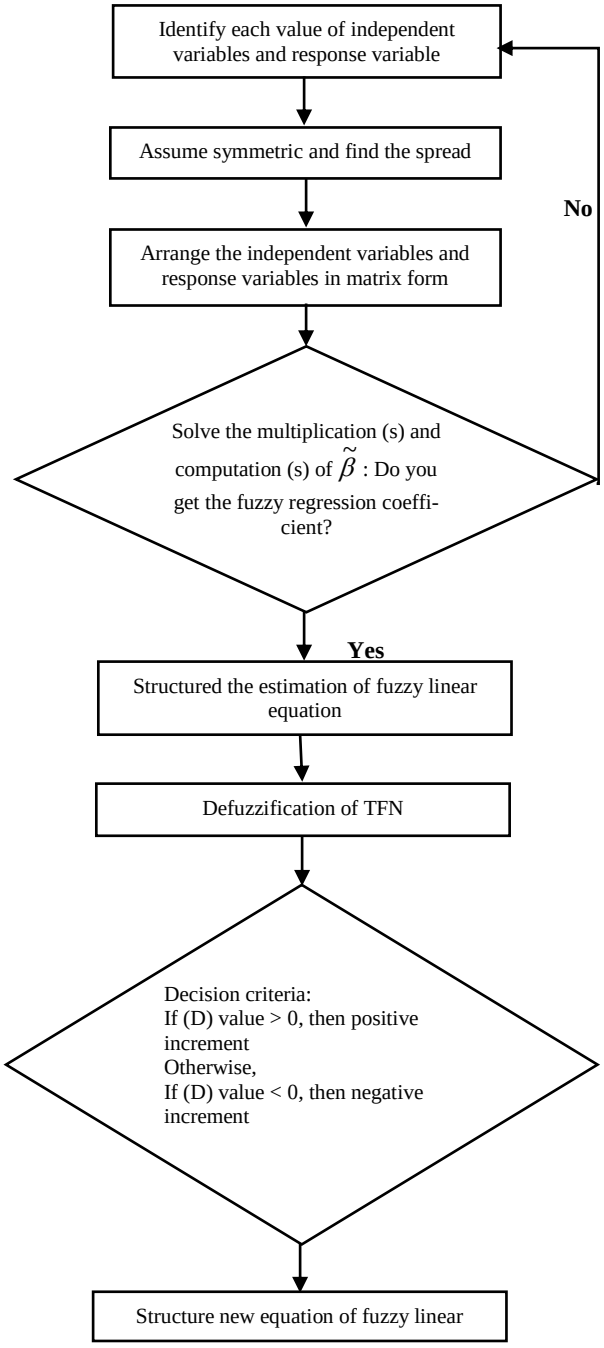

Fig. 3: Framework of multivariate matrix for fuzzy linear regression

In this study, a new framework of multivariate matrix for fuzzy linear regression is proposed by modifying the work done by [15] by simplifying several steps and the data are considered as symmetrical TFNs. The framework is shown in Figure 3. The general model expressed in the matrix as in (4) and the explanation is shown in (5)-(7),

$$\tilde{Y} = X\tilde{\beta} \tag{4}$$

where

$$X = \begin{bmatrix} 1 & X_{11} & X_{21} & X_{31} & X_{41} \\ 1 & X_{12} & X_{22} & X_{32} & X_{42} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{118} & X_{218} & X_{318} & X_{418} \end{bmatrix} \tag{5}$$

$$\tilde{Y} = \begin{bmatrix} a_1, c_{1,j} \\ a_2, c_{2,j} \\ \vdots \\ a_{18}, c_{18,j} \end{bmatrix} \tag{6}$$

and

$$\tilde{\beta} = \left[\hat{\tilde{\beta}}_0 \hat{\tilde{\beta}}_1 : \hat{\tilde{\beta}}_4\right] = \begin{bmatrix} \left(a_0, (1-\mu)c_{0,j}\right) \\ \left(a_1, (1-\mu)c_{1,j}\right) \\ \vdots \\ a_4, (1-\mu)c_{4,j} \end{bmatrix} \tag{7}$$

In (5)-(7), data matrices related with response variable and independent variables. In this study, matrix $\tilde{\beta}$ include the least squares estimates of the regression coefficients. In (4) can be transformed as shown in (8).

$$\left(X'X\right)\hat{\tilde{\beta}} = X\tilde{Y} \tag{8}$$

$$X'X = \begin{bmatrix} n & \sum_{i=1}^{n} x_{x1i} & \sum_{i=1}^{n} x_{2i} & \cdots & \sum_{i=1}^{n} x_{ki} \\ \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i}x_{2i} & \cdots & \sum_{i=1}^{n} x_{1i}x_{ki} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} x_{ki} & \sum_{i=1}^{n} x_{ki}x_{1i} & \sum_{i=1}^{n} x_{ki}x_{2i} & \cdots & \sum_{i=1}^{n} x_{ki}^2 \end{bmatrix} \tag{9}$$

$$X'Y = \begin{bmatrix} g_0 = \sum_{i=1}^{n} y_i \\ g_1 = \sum_{i=1}^{n} x_{1i}y_i \\ \vdots \\ g_k = \sum_{i=1}^{n} x_{ki}y_i \end{bmatrix} \tag{10}$$

where X' is the transpose of matrix X .Then, by matrix operations, the regression coefficients derived as shown is in (11),

$$\hat{\tilde{\beta}} = \left(X'X\right)^{-1}X'\tilde{Y}, \tag{11}$$

where $\left(X'X\right)^{-1}$ is the inverse matrix of $X'X$. The fitted fuzzy regression equation was formed from the estimated regression coefficients.

Since the estimated regression coefficients is in triangular fuzzy numbers, they need to be defuzzified (D) using in (12) shown in Figure 4.

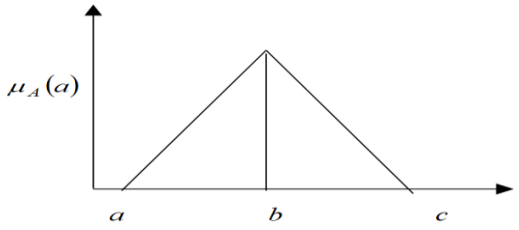

Fig. 4: Triangular fuzzy number

$$D = \frac{a + 4b + c}{6} \tag{12}$$

Decision criteria:

If (D) value > 0, then positive increment
Otherwise,
If (D) value < 0 then negative increment

2.4. Fuzzy coefficient of determination

The fuzzy coefficient of determination $(HR)^2$ is represent the amount of variance in the response variable explained by independent variables and the range is 0 to 1 which is defined by:

$$(HR)^2 = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{\tilde{Y}}\right)^2}{\sum_{i=1}^{n}\left(\tilde{Y}_i - \bar{\tilde{Y}}\right)^2} \tag{13}$$

where $\bar{\tilde{Y}}$ stands for mean of fuzzy data, $\tilde{Y}_i$ while $\hat{Y}$ is the estimation value. Besides, in (13) can represent by the following expression in (14) with the $\mu = 0$.

$$(HR)^2 = \frac{\sum_{i=1}^{n}\left(a_0 + a_1 X_1 - \bar{Y}\right)^2 + (1-\mu)\sum_{i=1}^{n}\left(c_{0,j} + c_{1,j} X_1 - \bar{e}_j\right)^2}{\sum_{i=1}^{n}\left(Y_1 - \bar{Y}\right)^2 + (1-\mu)\sum_{i=1}^{n}\left(e_{i,j} - \bar{e}_j\right)^2} \tag{14}$$

$\bar{e}_j$ represents the mean of spread fuzzy width. The fuzzy correlation coefficient (HR) is the multiply *HR* times *HR* to get $(HR)^2$ value. It can interpret how strong of the linear relationship between response variable and independent variables.
Next, e_i is the deviation between actual value Y_i and the estimated value $\hat{Y}_i$. It is defined as general in (15),

$$e_i = Y_i - \hat{Y}_i. \tag{15}$$

There are three types of errors that were computed which are Sum Square Regression (SSR), Sum Square Errors (SSE) and Total of Sum Square (SST). The Equation of SSR, SSE and SST are shown in (16)-(18),

$$SSR = \sum_{i=1}^{n}\left(\tilde{Y}_i - \bar{\tilde{Y}}_i\right)^2 \tag{16}$$

$$SSE = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n} e_i^2, \tag{17}$$

$$SST = SSR + SSE \tag{18}$$

Fuzzy linear regression equation, fuzzy correlation coefficient and fuzzy coefficient of determination are the main modelling measures that were obtained from the tax revenue data.

3. Results and Discussion

Eighteen years of time series data from the year 1996 to 2013 are gathered to test the model. Data of the response variable and independent variables are retrieved from The World Bank Data [17] and IndexMundi [9]. The four predictors are labelled as X_1, X_2, X_3 and X_4 where X_1 represents for the GDP, X_2 represents for inflation, X_3 represents for unemployment and X_4 represents for the merchandise trade. Tax Revenue is the response variable denoted as $\tilde{Y}$. These variables are fed into the multivariate matrix for fuzzy linear regression model. The model of [15] is simplified into a several step computation procedures shown in Figure 3.

3.1. Estimating fuzzy linear regression equation

Step 1: Identify value each of the independent variables X_1, X_2, X_3, X_4 and the value of response variable Y_i .
Step 2: Assume symmetric and find the spread.
Step 3: Arrange the value of each independent variables and response variable in matrix form as given in (5)-(6).

$$X = \begin{bmatrix} 1 & 100.85 & 3.5 & 2.5 & 155.42 \\ 1 & 100.17 & 2.7 & 2.4 & 157.76 \\ 1 & 72.18 & 5.3 & 3.2 & 182.39 \\ 1 & 79.15 & 2.7 & 3.4 & 188.79 \\ 1 & 93.79 & 1.5 & 3 & 182.39 \\ 1 & 92.78 & 1.4 & 3.5 & 174.46 \\ 1 & 100.85 & 1.8 & 3.5 & 172.47 \\ 1 & 110.2 & 1 & 3.6 & 170.6 \\ 1 & 124.75 & 1.5 & 3.5 & 185.81 \\ 1 & 143.53 & 3 & 3.5 & 178.32 \\ 1 & 162.69 & 3.6 & 3.3 & 178.98 \\ 1 & 193.55 & 2 & 3.2 & 166.44 \\ 1 & 230.99 & 5.4 & 3.3 & 154.13 \\ 1 & 202.25 & 0.6 & 3.7 & 138.93 \\ 1 & 255.02 & 1.7 & 3.4 & 142.44 \\ 1 & 297.95 & 3.2 & 3.1 & 139.47 \\ 1 & 314.44 & 1.7 & 3 & 134.82 \\ 1 & 323.28 & 2.1 & 3.1 & 134.32 \end{bmatrix} \quad \tilde{Y} = \begin{bmatrix} 19.38 & 1.7764 \\ 19.75 & 1.9614 \\ 19.73 & 1.9514 \\ 14.09 & 0.8686 \\ 13.67 & 1.0786 \\ 17.79 & 0.9814 \\ 17.48 & 0.8264 \\ 15.5 & 0.1636 \\ 15.2 & 0.3136 \\ 14.83 & 0.4986 \\ 14.52 & 0.6536 \\ 14.3 & 0.7636 \\ 14.67 & 0.5786 \\ 14.94 & 0.4436 \\ 13.33 & 1.2486 \\ 14.79 & 0.5186 \\ 15.61 & 0.1086 \\ 15.31 & 0.2586 \end{bmatrix}$$

Step 4: The computation of centre, a_j and spread c_j of each β , based on section B shown in Table 1.

Table 1: Fuzzy Regression Coefficient

	β_0	β_1	β_2	β_3	β_4
a_j	39.9099	-0.0294	0.4396	-1.1453	-0.1016
c_j	6.4466	-0.0068	0.1375	-0.5839	-0.018

Step 5: Apply Fuzzy linear regression in (3).

$$\tilde{Y}_i = (39.9099, 6.4466) + (-0.0294, -0.0068)X_1 + (0.4396, 0.1375)X_2 + (-1.1453, -0.5839)X_3 + (-0.1016, -0.018)X_4$$

Step 6: Defuzzification of TFN into crisp output in (12)

Table 2: Defuzzification of Triangular Fuzzy Number

a	b	c	D
27.0167	33.4633	39.9099	33.4633
-0.0158	-0.0226	-0.0294	-0.0226
0.1646	0.3021	0.4396	0.3021
0.0225	-0.5614	-1.1453	-0.5614
-0.0656	-0.0836	-0.1016	-0.0836

Based on the decision criteria from section C referring Table 2, the (D) value greater than zero means the positive increment that ef-

fect the value of $\tilde{Y}$ and influence the model. However, if the (D) value less than zero means negative increment of the model. Thus, any increment of independent variables by 1 unit, the (D) value will increase or decrease by that value.

Step 7: Reform equation of fuzzy linear regression based on step 5 and 6.

$$\tilde{Y}_i = 33.4633X_0 + (-0.0226X_1) + 0.3021X_2 + (-0.5614X_3) + (-0.0836X_4)$$

3.2. Estimating for fuzzy coefficient of determination

The estimating of fuzzy coefficient of determination shown in Table 3.

Table 3: Estimating Coefficient of Determination

	Center	Spread	Total
SSE	25.1493	1.4959	26.6452
SSR	47.6603	4.2166	51.8770
SST	72.8096	5.7126	78.5222
HR^2			0.6606
HR			0.8128

Table 4: Interpretation of Correlation Coefficients [4]

Range of Correlation Coefficients	Degree of Correlation
0.8 – 1.00	Very strong positive
0.6 – 0.79	Strong positive
0.4 – 0.59	Moderate positive
0.2 – 0.39	Weak positive
0 – 0.19	Very weak positive
0 – (-0.19)	Very weak positive
(-0.20) – (-0.39)	Weak negative
(-0.40) – (-0.59)	Moderate negative
(-0.60) – (-0.79)	Strong negative
(-0.80) – (-1.00)	Very strong negative

Table 3 shows the result for fuzzy coefficient of determination, which is 66.06% This means that 66.06% of variance of tax revenue is explained by the independent variables. The other 33.94% of total variation in tax revenue remains unexplained. The fuzzy correlation coefficient, thus yields $0.8128=\sqrt{0.6606}$. According to Table 4, signifying a strong positive linear correlation between the tax revenue and independent variables.

4. Conclusion

In this paper, a multivariate matrix is proposed to select the best factor influence tax revenue among four variables with symmetric triangular fuzzy numbers. By using symmetric triangular fuzzy numbers, the computations based on multivariate matrix fuzzy linear regression can be performed more accurately. Furthermore, the model is simple and easier to apply by using matrix algebra.
The method was evaluated using fuzzy coefficient of determination (HR^2). As mention in the result, we can see that the (HR^2) was strongly positive relationship between tax revenue and predictor variables. Therefore, the application of this method should be able to provide more accurate in predicting the tax revenue.

Acknowledgement

I would like to thanks to Mahmod Othman and Samsul Ariffin Abdul Karim, for valuable support and guidance throughout my studies. This research is supported by UTP Graduate Assistantship (GA) scheme.

References

[1] Al-Ghandoor, A., & Samhouri, M. (2009). Electricity consumption in the industrial sector of Jordan: Application of multivariate linear regression and adaptive neuro-fuzzy techniques. Jordan Journal of Mechanical and Industrial Engineering, 3(1), 69-76.
[2] Ahmad, A. (2015). Rakyat perlu cuba faham GST. Sinar Harian, http://www.sinarharian.com.my/wawancara/rakyat-perlu-cuba-faham-gst-1.368964.
[3] Jaffri, A. A., Tabassum, F., & Asjed, R. (2015). An empirical Investigation of the relationship between trade liberalization and tax revenue in Pakistan. Pakistan Economic and Social Review, 53(2), 317-330.
[4] Chowdhury, K. A., Debsarkar, A., & Chakrabarty, S. (2015). Novel methods for assessing urban air quality: Combined air and noise pollution approach. Journal of Atmospheric Pollution, 3(1), 1-8.
[5] Chang, P. T., & Lee, E. S. (1996). A generalized fuzzy weighted least-squares regression. Fuzzy Sets and Systems, 82(3), 289-298.
[6] D'Urso, P., & Gastaldi, T. (2000). A least-squares approach to fuzzy linear regression analysis. Computational Statistics and Data Analysis, 34(4), 427-440.
[7] D'Urso, P., & Gastaldi, T. (2001). Linear fuzzy regression analysis with asymmetric spreads. In S. Borra, R. Rocci, M. Vichi, & M. Schader (Eds.), Advances in Classification and Data Analysis. Berlin: Springer, pp. 257-264.
[8] Chen, F., Chen, Y., Zhou, J., & Liu, Y. (2016). Optimizing h value for fuzzy linear regression with asymmetric triangular fuzzy coefficients. Engineering Applications of Artificial Intelligence, 47, 16-24.
[9] Index Mundi. (2017). https://www.indexmundi.com/blog/.
[10] Inland Revenue Board of Malaysia. (2015). http://www.hasil.gov.my/.
[11] Abdullah, L., & Zakaria, N. (2012). Matrix driven multivariate fuzzy linear regression model in car sales. Journal of Applied Sciences(Faisalabad), 12(1), 56-63.
[12] Abdullah, L., & Jamal, N. J. M. (2015). The relationship between dimensions of health related quality of life and health conditions among elderly people: A fuzzy linear regression approach. Modern Applied Science, 10(2), 1-10.
[13] Abdullah, L., & Khalid, N. D. (2014). Prediction of carbon dioxide emissions using fuzzy linear regression model: A case of developed and developing countries. Journal of Sustainability Science and Management, 9(1), 69-77.
[14] Mahmood, H., & Chaudhary, A. R. (2013). Impact of FDI on tax revenue in Pakistan. Pakistan Journal of Commerce and Social Sciences, 7(1), 59-69.
[15] Pan, N. F., Lin, T. C., & Pan, N. H. (2009). Estimating bridge performance based on a matrix-driven fuzzy linear regression model. Automation in Construction, 18(5), 578-586.
[16] Tanaka, H., Uejima, S. & Asai, K. (1982). Linear regression regression analysis with fuzzy model. IEEE Transaction on System, Man and Cybernetics, 12(6), 903-907.
[17] World Bank Data. www.worldbankdata.org.
[18] Liu, X., & Chen, Y. (2013). A systematic approach to optimizing value for fuzzy linear regression with symmetric triangular fuzzy numbers. Mathematical Problems in Engineering, 2013, 1-9.
[19] Zadeh, L. A. (1965). Fuzzy sets. Information and Control, 8, 338-353.