Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 12 (1): 56-63, 2012
ISSN 1812-5654 / DOI: 10.3923/jas.2012.56.63

Matrix Driven Multivariate Fuzzy Linear Regression Model in Car Sales

Lazim Abdullah and Nadia Zakaria
Department of Mathematics, University Malaysia Terengganu,
21030 Kuala Terengganu, Malaysia

Abstract: Fuzzy linear regression has been used in predicting analysis as to handle uncertainty variables. Many methods of fuzzy linear regressions were introduced but most of the methods associated with substantial complex computation procedures. The model of matrix-driven fuzzy linear regression was proposed as to overcome the computational risk and was successfully tested in a civil engineering application. This study extends the application of the model to investigate the relationship between variables impacting car sales volume. The variables of petroleum prices, population, Gross Domestic Product (GDP) and Gross National Product (GNP) are predicted with the response variable of car sales volume. Thirty years' time series data of the variables from various Malaysian agencies were fed into the models. It is found that the model successfully yield a fuzzy linear regression equation as to explain the relationship between predictors and response variable. It also notices that eighty eight percent variations in car sales volume attributed by price of petroleum, population, GNP and GDP. The model also successfully explained the contributions of left and right errors of fuzzy numbers of regression coefficients to the car sales volume. The fuzzy numbers that represent coefficients of regression certainly offer a new contribution to the relationships between the variable of car sales volume and the four predictors.

Key words: Car sales volume, matrix, fuzzy linear regression, error analysis, coefficient of determination

INTRODUCTION

Regression analysis is a powerful and comprehensive method for analyzing relationships between a response variable or popularly known as dependent variable and one or more explanatory variables or termed as independent variables. Inferential problems associated with regression model include the estimation of the model parameters and prediction of the response variable from knowledge of the explanatory variables. The regression method has been used in many day lives experiments. In spite of the widespread use of the multiple-regression method in many day life activities, there exists uncertainty in variables used in the multi-linear regression method. As to handle the uncertainty, fuzzy linear regression was introduced. Conventional regression cannot handle visual inspection results that are inherently non-crisp or linguistic. On the other hand, fuzzy regression provides an effective means for coping with such fuzzy data or linguistic variables. It was stemming from Zadeh (1965) novel thought that fuzzy uncertainty as ambiguity and vagueness. The fuzzy regression analysis was introduced more than thirty years ago by Tanaka *et al.* (1982). They explain fuzzy uncertainty of dependent variables with the fuzziness of response functions or regression coefficients in the linear regression model. Some years later, Sakawa and Yano (1992) and D'Urso and Gastaldi (2001) extend the idea of fuzzy linear regression. Basically, the fuzzy regression models can be classified in two classes. The first class is based on the possibilistic concept (Tanaka *et al.*, 1982; Tanaka and Lee, 1998; Chen, 2001; Lee and Chen, 2001) and the second class is the least-squares approach (D'Urso and Gastaldi, 2000, 2001; Xu and Li, 2001; Wu, 2003). Since then, there were many improvements of the fuzzy linear regression method and its applications as well. There has been handful of research using fuzzy linear regression to model real life applications but mostly concentrated on information technology management. Kahraman (2002), for example forecasts sale levels of computer equipment in Turkey using fuzzy linear regressions. Wu *et al.* (2009) describes an application of the fuzzy linear regression analysis for land-cover classification of Landsat TM data in remote sensing field.

In fuzzy regression, there is a tendency that the greater the values of independent variables, the wider the

Corresponding Author: Lazim Abdullah, Department of Mathematics, University Malaysia Terengganu, 21030 Kuala Terengganu, Malaysia

width of the estimated dependent variables. This causes a decrease in the accuracy of the fuzzy regression model constructed by the least square method. Choi and Buckley (2008) suggest the least square absolute deviation estimators to construct the fuzzy regression model and investigate the performance of the fuzzy regression models with respect to a certain error measure. Simulation studies and examples show that their model produces less error than the fuzzy regression model studied by many authors that used the least squares method when the data contains fuzzy outliers. Bargiela *et al*. (2007) has proposed that an iterative algorithm for multiple regression with fuzzy variables. While using the standard least squares criterion as a performance index, he creates the regression problem as a gradient-descent optimization. Thus far, most of the existing papers 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 that it could be greatly affected by small number of outliers. Since outliers in the response variable represent model failure, there has been an increased interest in robust estimation procedures which are insensitive to some outliers, applied to the regression analysis. Like to an ordinary regression, robust methods in order to estimate the fuzzy regression coefficients are needed. In the regression analysis, it was well known that the least squares estimators perform poorly in case the data contains some outliers and/or depart from a normal distribution. Also, in the fuzzy regression model that used the least squares method, there is a tendency that the larger the values of independent variables, the wider the spreads of the estimated dependent variables.

However, many of the existing fuzzy regression models require substantial computations due to the complicated fuzzy arithmetic. The regression model proposed by Tanaka *et al*. (1982) is quite popular and useful; however, this model is restricted to symmetric triangular fuzzy numbers. To overcome this limitation, Chang and Lee (1996) developed a fuzzy least-squares regression model. In their model, the regression coefficients are derived from a nonlinear programming problem that requires considerable computations. Also, the notion of least squares method or best fit line incorporates the optimization functional associated with the problems. It is true that the best fit line mechanism in fuzzy linear regression confronts with high computational risk especially in solving optimization problems. Therefore, Pan *et al*. (2009) introduced a matrix-driven multivariate fuzzy linear regression as part of their efforts to improve computational efficiency. The method has been successfully tested to engineering study of estimating bridge performance. The matrix driven fuzzy regression has come out to determine reliable bridge maintenance and rehabilitation strategy. The matrix-driven fuzzy regression presents a multiple fuzzy linear regression using matrix algebra. The model is capable of dealing with a mixture of fuzzy data and crisp data. Moreover, the approach is intuitive and easy to implement as compared to other related fuzzy regression models. Motivated by the low computational risk of the method and silent attempt of the method in modeling parameters of car business peripherals, this paper intends to extend the method to offer the relationships between car sales volume and variables that impact on sales. Specifically the aim of this paper is to simplify the matrix-driven fuzzy linear regression into a nine-step computation procedure and test it to car sales volume variable and its related variables. The test includes obtaining equation of the fuzzy linear equation, estimating error terms between the actual car sales volume and predictive values and estimating correlation coefficients between variables.

MATRIX-DRIVEN MULTIVARIATE FUZZY LINEAR REGRESSION MODEL

For the purpose of clarity and self-explained, the method proposed by Pan *et al*. (2009) is retrieved. The general matrix-driven multivariate fuzzy linear regression is given as follows.

With k crisp independent variables and one fuzzy dependent variable, the estimated fuzzy linear regression can be expressed as:

$$\hat{Y}_i = (a_0, c_{0,L}, c_{0,R}) + (a_1, c_{1,L}, c_{1,R})X_1 + (a_2, c_{2,L}, c_{2,R})X_2 + \ldots + (a_k, c_{k,L}, c_{k,R})X_k \quad (1)$$

where, $(a_0, c_{0,L}, c_{0,R})$ is the fuzzy intercept coefficient; $(a_1, c_{1,L}, c_{1,R})$ is the fuzzy slope coefficient for X_1; $(a_2, c_{2,L}, c_{2,R})$ is the fuzzy slope coefficient for X_2 and $(a_k, c_{k,L}, c_{k,R})$ is the kth fuzzy slope coefficient. The estimated $\hat{Y}_i$ at a particular μ value is given by:

$$\mu_{\hat{Y},L} = [a_0 - (1-\mu)\, c_{0,L}] + [a_1 - (1-\mu)\, c_{1,L}]X_1 + \ldots + [a_k - (1-\mu)\, c_{k,L}]\, X_k = (a_0 + a_1 + \ldots + a_k) - (1-\mu)(c_{0,L} + c_{1,L}\, X_1 + c_{2,L}\, X_2 + \ldots + c_{k,L}\, X_k) \quad (2)$$

and:

$$\mu_{\hat{Y},L} = (a_0 + a_1 + ... + a_k) - (1-\mu)(c_{0,R} + c_{1,R}\ X_1 + c_{2,R}\ X_2 + ... + c_{k,R}\ X_k) \quad (3)$$

where, a_0, a_1,..., a_k are the estimated coefficients of $\hat{Y}_i$ at $\mu = 1$; $c_{0,L}+c_{1,L} X_1$ and $c_{0,R}+c_{0,R} X_1$ are the left fuzzy width and the right fuzzy width for X_2; $c_{0,L}+c_{2,L} X_2$ and $c_{0,R}+c_{2,R} X_2$ are the left fuzzy width and the right fuzzy width for X_2; $(1-\mu)(c_{0,L}+c_{1,L} X_1)$ and $(1-\mu)(c_{0,R}+c_{2,R} X_2)$ are the left fuzzy width for X_1 at a given μ value; $(1-\mu)(c_{0,L}+c_{2,L} X_2)$ and $(1-\mu)(c_{0,R}+c_{2,R} X_2)$ are the left fuzzy width and the right fuzzy width for X_2 at a given μ value.

Observed data represented by asymmetrical triangular fuzzy numbers can be expressed as:

$$\tilde{Y}_i = (y_i, (1-\mu)\, e_{i,L}, (1-\mu)\, e_{1,R}) \quad (4)$$

If $\tilde{Y}_i$ is symmetric triangular fuzzy number, $e_{i,L} = e_{i,R}$. The method of least square is used to find that particular regression line ($\tilde{Y}_i$) where the sum of squared deviations of the data points ($\tilde{Y}_i$) above or below it is minimized. To facilitate the fuzzy regression analysis, matrix algebra is employed. The general fuzzy linear model can be expressed in the following matrix form:

$$\tilde{Y} = X\tilde{\beta} \quad (5)$$

Where:

$$\tilde{Y} = \begin{bmatrix} (y_1, (1-\mu)e_{i,L}, (1-\mu)\, e_{1,R}) \\ (y_2, (1-\mu)e_{2,L}, (1-\mu)\, e_{2,R}) \\ (y_n, (1-\mu)e_{n,L}, (1-\mu)\, e_{n,R}) \end{bmatrix} \quad (6)$$

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \quad (7)$$

$$\hat{\tilde{\beta}} = [\hat{\tilde{\beta}}_0 \quad \hat{\tilde{\beta}}_1 \quad \vdots \quad \hat{\tilde{\beta}}_k] = \begin{bmatrix} (a_0, (1-\mu)c_{0,L}, (1-\mu)\, c_{0,R}) \\ (a_1, (1-\mu)c_{1,L}, (1-\mu)\, c_{1,R}) \\ (a_n, (1-\mu)c_{n,L}, (1-\mu)\, c_{n,R}) \end{bmatrix} \quad (8)$$

In the above equations, matrices $\tilde{Y}$ and X are the data matrices associated with response variable and predictor variables, respectively. Matrix $\hat{\tilde{\beta}}$ contains the least squares estimates of the regression coefficients. To obtain the regression parameters, Eq. 5 can be transformed by:

$$(X'X)\hat{\tilde{\beta}} = X'\tilde{Y} \quad (9)$$

where, x' is the transpose matrix of x.

By matrix operations, the regression coefficients can be derived as follows:

$$\hat{\tilde{\beta}} = (X'X)^{-1} X'\tilde{Y} \quad (10)$$

where, $(X'X)^{-1}$ is the inverse matrix of X'X.

The fitted fuzzy regression equation can be developed based on the estimated regression coefficients. After establishing the regression equation, it is of interest to measure the quality or reliability of the fitted regression equation. The fuzzy coefficient of determination $(HR)^2$ is used to interpret the proportion of the total variation in Y explained by the regression line which is defined by:

$$(HR)^2 = \frac{\sum_{i=1}^{n}(\hat{\tilde{Y}}_i - \bar{\tilde{Y}})^2}{\sum_{i=1}^{n}(\tilde{Y}_i - \bar{\tilde{Y}})^2} \quad (11)$$

where, $\bar{\tilde{Y}}$ is the mean of fuzzy data $\tilde{Y}$.

The above expression can be represented by the following expression:

$$(HR)^2 = \frac{\sum_{i=1}^{n}(a_0 + a_1 X_i - \bar{Y})^2 + (1-\mu)\sum_{i=1}^{n}(c_{0,L} + c_{1,L} X_i - \bar{e}_L)^2 + (1-\mu)\sum_{i=1}^{n}(c_{0,R} + c_{1,R} X_i - \bar{e}_R)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2 + (1-\mu)\sum_{i=1}^{n}(e_{i,L} - \bar{e}_L)^2 + (1-\mu)\sum_{i=1}^{n}(e_{i,R} - \bar{e}_R)^2} \quad (12)$$

where, $\bar{e}_L$ and $\bar{e}_R$ are the mean of left fuzzy width and mean of right fuzzy width, respectively. Likewise, the fuzzy correlation coefficient (HR) is the root of $(HR)^2$ which can evaluate the strength of the linear relationship between predictor variables and response variables.

The ith residual is the different between the observed value Y_i and the corresponding fitted value $\hat{\tilde{Y}}_i$. This residual is denoted by e_i and is defined in general as follows:

$$e_i = Y_i - \hat{\tilde{Y}}_i$$

In this case, it needs to distinguish between the model error term $\varepsilon_i = Y_i - E\{Y_i\}$ and the residual:

$$e_i = Y_i - \hat{Y}_i$$

The appropriate sum of squares, denoted by SSE, is:

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

where, SSE stands for sum squared errors or residual sum of squares. The sum of squares SSE has n-2 degrees of freedom associated with it. Two degrees of freedom are lost because both β_0 and β_1 had to be estimated in obtaining the estimated means $\hat{Y}_i$. Hence the appropriate mean square, denoted by MSE or s^2, is:

$$s^2 = MSE = \frac{SSE}{n-2} = \frac{\sum Y_i - \hat{Y}_i}{n-2} = \frac{\sum e_i^2}{n-2}$$

where, MSE stands for error mean square or residual mean square.

The normality assumption for the error terms is justifiable in many situations because the error terms frequently represent the effects of factors omitted from the model that affect the response to some extent and that vary at random without reference to the variable X. Also, there might be random measurement errors in the recording of Y. In so far, as these random effects have a degree of mutual independence, the composite error term e_i representing all these factors would tend to comply with the central limit theorem and the error term distribution would approach normality as the number of factor effects becomes large.

Fuzzy linear regression equation, coefficient of determination and correlation coefficient are the main modeling measures that would be obtained from the car sales data.

VARIABLES OF THE MODEL

Owning a private car in the present day is an indispensable component of all aspects of modern life. Cars help people increase in mobility and is no longer viewed as a status symbol. The importance of owning cars gives a good indicator to car sales volume as demand for owning cars increases. There are many variables that can be related with volume of car sales volume. One of the must have variables in car sales volume is prices of petroleum. Prices of petroleum are depending on world economy. Petroleum pricing is always unpredictable depending to many other economy parameters. Apart from prices of petroleum, total population in a country is also contributes to car sales volume. Other than that, Gross Domestic Product (GDP) is another variable in predicting car sales volume. GDP is used as a basic measure of a country's overall economic output. The market value of all final goods and services made within the borders of a country in a year. It is often positively correlated known as standard of living. Another important variable that can be counted in explaining car sales volume is the Gross National Product (GNP). This measure reflects total market value of all the goods and services produced by an action during a specified period. GNP is a measure of a country's economic performance or what its citizens produced (i.e., goods and services) and whether they produced these items within its borders. These variables are chosen based on the variables used in several researches in the past. Khan and Willumsen (1986), for example used gross national product per head, the purchase and registration tax per vehicle, the ownership tax per vehicle, import duty per vehicle, the regular fuel price and population density were selected as causal variables in car ownership at a developing countries. Prevedouras and An (1998) include percentile form of the consumer price index, GDP, unemployment rate, railway passenger mileage and roadway mileage as variables in time-series regression model in Asian countries. Lam and Tam (2002) used eight variables affecting car ownership were chosen for analysis, including annual GDP, annual passenger trips on public transport, annual railway passenger kilometers, the average annual license fee per private car, the average first registration tax per private car, average petrol price per liter, population and population density, all selected as explanatory variables. Hook and Replogle (1996) indicate that growth in vehicle ownerships and use is often seen as an inevitable outcome of increasing GDP and incomes. Notwithstanding the other variables, the present paper limits to the four explanatory variables (predictors) of the model. The explanatory variables are prices of petroleum, population, GDP and GNP. The impact of these predictors are simultaneously employed to the matrix-driven fuzzy linear regression where the volume of car sales as response variable.

IMPLEMENTATIONS

Thirty years of time series data from the year 1980 to 2009 are gathered to test the model. Data of the response variable and predictors are retrieved from Malaysian Ministry of Domestic Affairs and World Bank database. As to fit with the model, the four predictors are labeled as X_1, X_2, X_3 and X_4 where, the X_1 represents for the prices of petroleum, X_2 represents total population in Malaysia, X_3 represents GNP and X_4 represents GDP. Sales of car is the response variables and denoted as $\tilde{Y}$. These variables are fed into the matrix-driven fuzzy linear regression model.

The model of Pan *et al*. (2009) is simplified into a nine-step computation procedure.

The proposed computation steps that specifically employed for the case of car sales volume and its related variables are given as follows:

Step 1: Identify value each of the explanatory variables (predictors) X_1, X_2, X_3, X_4 and arrange in matrix form:

$$X=\begin{bmatrix}1 & 0.89 & 13763441 & 31 & 24.9\\ 1 & 0.89 & 14106332 & 33.6 & 25.2\\ 1 & 0.89 & 14466535 & 40.4 & 27.2\\ 1 & 0.89 & 14846704 & 43.8 & 30.7\\ 1 & 0.89 & 15249772 & 48.7 & 34.6\\ 1 & 0.89 & 15677187 & 49.5 & 31.8\\ 1 & 0.89 & 16130844 & 51.4 & 28.2\\ 1 & 0.89 & 16608660 & 55.9 & 32.2\\ 1 & 0.89 & 17103098 & 64 & 35.3\\ 1 & 0.89 & 17603827 & 72.4 & 38.8\\ 1 & 1.1 & 18103341 & 83.2 & 4.1\\ 1 & 1.1 & 18597361 & 93.5 & 49.1\\ 1 & 1.1 & 19087376 & 104 & 59.2\\ 1 & 1.1 & 19578569 & 117 & 66.9\\ 1 & 1.1 & 20079056 & 131 & 74.5\\ 1 & 1.1 & 20593952 & 147 & 88.8\\ 1 & 1.1 & 21125065 & 165 & 101\\ 1 & 1.1 & 21668014 & 178 & 106\\ 1 & 1.1 & 2214316 & 167 & 72.2\\ 1 & 1.1 & 22752310 & 177 & 79.1\\ 1 & 1.2 & 23273615 & 194 & 93.8\\ 1 & 1.3 & 23774848 & 202 & 92.8\\ 1 & 1.32 & 24258296 & 217 & 101\\ 1 & 1.33 & 24728210 & 239 & 110\\ 1 & 1.42 & 25191441 & 263 & 125\\ 1 & 1.52 & 25652985 & 287 & 138\\ 1 & 1.62 & 26113731 & 319 & 156\\ 1 & 1.92 & 26549518 & 251 & 187\\ 1 & 2.7 & 36992577 & 371 & 195\\ 1 & 2.05 & 25715819 & 382 & 201\end{bmatrix}$$

Step 2: Identify the value of response variable $\tilde{Y}$:

$$\tilde{Y}=\begin{bmatrix}97262 & 77809.6 & 106988.2\\ 100935 & 80748 & 11028.5\\ 102447 & 81957.6 & 112691.7\\ 108314 & 86651.2 & 119145.4\\ 109915 & 87932 & 120906.5\\ 94999 & 75999.2 & 104498.9\\ 67847 & 54277.6 & 74631.7\\ 48996 & 39196.8 & 53895.6\\ 71592 & 87273.6 & 78751.2\\ 109357 & 87485.6 & 120292.7\\ 165861 & 132688.8 & 182447.1\\ 181877 & 145501.6 & 36364.7\\ 145084 & 116067.2 & 159592.4\\ 167928 & 134342.4 & 1847208\\ 200435 & 160348 & 220478.5\\ 285792 & 228633.6 & 257212.8\\ 364788 & 291830.4 & 401266.8\\ 404837 & 323869.6 & 4453207\\ 163851 & 131080.8 & 180236.1\\ 288547 & 230837.6 & 317401.7\\ 343173 & 274538.4 & 3774903\\ 396381 & 371704.8 & 436019.1\\ 434954 & 347963.2 & 478449.4\\ 405745 & 324596 & 446319.5\\ 487605 & 390084 & 536365.5\\ 552316 & 441852.8 & 607547.6\\ 490768 & 392614.4 & 539844.8\\ 487176 & 389740.8 & 535893.6\\ 458115 & 438492 & 602926.5\\ 397619 & 318095.2 & 437380.9\end{bmatrix}$$

Step 3: Compute the multiplication of $X'\ Y_i$:

$$X'Y_i=\begin{bmatrix}7.824516\times10^{6}\\ 1.10155349399999995\times10^{7}\\ 1.78328866941695000\times10^{14}\\ 1.6968269792000000\times10^{9}\\ 8.65718363100000024\times10^{8}\end{bmatrix}$$

Step 4: Compute the multiplication of $X'e_L$:

$$X'e_L=\begin{bmatrix}6.25961279999999982\times10^{6}\\ 8.81242795199999958\times10^{6}\\ 1.42663093553356000\times10^{14}\\ 1.35746158335999990\times10^{9}\\ 6.92574690480000019\times10^{8}\end{bmatrix}$$

Step 5: Compute the multiplication of $X'\ e_R$:

$$X'e_R = \begin{bmatrix} 1.004896400000000 \times 10^7 \\ 1.37028801140000000 \times 10^7 \\ 2.24489368650984500 \times 10^{14} \\ 2.0373124472000003 \times 10^9 \\ 1.05039725717000008 \times 10^8 \end{bmatrix}$$

Step 3 to 5 are prerequisites prior to computing coefficients of the regression model.

The following steps (Step 6 to 8) are executed to find the regression coefficient and fuzzy coefficient for the case $\alpha = 0$.

Step 6: Compute mid value $\tilde{\beta} = (X'X)^{-1}.X'Y_i$:

$$\tilde{\beta} = \begin{bmatrix} -2.8768622 \times 10^5 \\ -49339.007 \\ 0.024089433 \\ -159.7509 \\ 1743.1885 \end{bmatrix}$$

Step 7: Compute left $\tilde{\beta} = (X'X)^{-1}.X'e_L$:

$$\tilde{\beta} = \begin{bmatrix} -2.3014900 \times 10^5 \\ -39471.205 \\ 0.0019271551 \\ -127.8010 \\ 1394.5511 \end{bmatrix}$$

Step 8: Compute right $\tilde{\beta} = (X'X)^{-1}.X'e_R$:

$$\tilde{\beta} = \begin{bmatrix} -1.34256990 \times 10^6 \\ -63678.884 \\ 0.093122953 \\ -7599.1158 \\ 10901.7521 \end{bmatrix}$$

Step 9: Structure the estimated fuzzy linear regression equation

Finally the equation can be written as:

$$\hat{\tilde{Y}} = (-2.3014900 \times 10^5, -287686.22, -1342569.9) + (-39471.21, -49339.01, 63678.88)X_1 + (0.00193, 0.0241, 0.0931)X_2 + (127.8, 159.751, 7599.12)X_3 + (1394.55, 1743.19, 10901.75)X_4$$

It can be seen that the slope of X_1 is (-39471.21, -49339.01, 63678.88), indicating that regardless of total population of Malaysia (X_2), based on GNP (X_3) and GDP (X_4). The estimated decreases of -49339.01 in the condition on index, decreases of -39471.21 in the lower bound (left interval) and increases of 63678.884 in the upper bound (right interval).

Likewise, the estimated slope coefficient X_2, is can be shown by (0.001927, 0.02409, 0.09312) which signifies the increasing or decreasing value that regardless of X_1 and X_3 for each additional X_2, $\tilde{Y}$ increases by 0.02409, with increases by 0.001927 for the lower bound and increases by 0.09312 in the upper bound, respectively.

Similarly, the estimated slope for the coefficient of X_3 which is regarding to GNP is (-127.801, -159.751, 7599.12) X_3 -159.7509, denotes that regardless of X_1 and X_2 and for each additional X_3, an estimated decreases of in $\tilde{Y}$ by -159.7509, with decreases by -127.8010 in the lower bound and 7599.1158 increases of in the upper bound. The result produced by the model indicate that price of petroleum is the most important variable for reducing volume of car sales in this case study. This information is useful for car market players to identify which variables should be given more attention in order to enhance car sales performance.

The second part analyzes errors as a result of modeling and predicting. As to ascertain stability of model, the estimated equation is further investigated using error analysis.

Three type of errors are computed viz. Sum Square Regression (SSR), Sum Square Errors (SSE) and Total of Sum Square (SST).

The SSR for response variable, *Y* and errors are computed and obtained as:

- For Y, SSR = 7.0127×10^{11}
- For e_L, SSR = 4.49×10^{11}
- For e_R, SSR = 8.94×10^{11}

The SSE is computed with the similar fashion.
It is found that:

- For Y, SSE = 9.47×10^{10}
- For e_L, SSE = 6.06×10^{10}
- For e_R, SSE = 2.48×10^{12}

The SST is computed using the relation SSR+SSE = SST.

- For Y, SST = 7.96×10^{11}
- For e_L, SST = 5.09×10^{11}
- For e_R, SST = 3.37×10^{12}

The error analysis can be used to validate the model based on correlation measure. This correlation is termed as coefficient of determination (HR) aiming to see the contribution of predictors to the overall car sales volume.

The $(HR)^2$ is used to interpret the proportion of the total variation in Y explained by the regression line and another way to express the level of prediction accuracy.

$$\text{Coefficient of determination } (HR)^2 = SSR/SST$$

Therefore:

- For Y $(HR)^2 = 0.881043715$
- For e_L, $(HR)^2 = 0.881044383$
- For e_R, $(HR)^2 = 0.265$

The model shows that nearly 88% of the total variation in car sales volume can be explained by the four predictors. Accordingly, 88 and 26% of the total variation e_L and e_R, respectively can be explained by flexibility of fuzzy numbers in the four predictors.

The strength of correlation between predictors and response variable are measured using correlation coefficient. The fuzzy correlation coefficient of the model is 0.938639289 $(=\sqrt{0.881043715})$, signifying a strong positive linear correlation between the car sales volume and the predictors. Therefore, these four variables provide a reliable prediction model of car sales performance.

CONCLUSIONS

One of the main concerns in explaining the associations between predictors and response variable is opting the appropriate linear model which can reduce computation risk and giving flexible interpretations. This study has proposed a nine-step wise computation based on matrix driven fuzzy linear regression. The matrix driven multivariate fuzzy linear regression based model has been tested to the data of response variable of car sales volume and its four predictors. The model shows that volume of car sales is mainly affected by price of the petroleum, total population, gross domestic product and gross national product. In order to explain the effects of these variables on car sales volume, a fuzzy multiple-regression model is used. The major reason for applying fuzzy regression is to overcome the inter correlation problem associated with the independent variables. In this study, price of petroleum, population, gross domestic product, gross national product are selected as explanatory variables or predictors. The results show that, by applying a multi-variable approach to fuzzy regression, the model provides not only provide a crisp output in fuzzy numbers but also offer a coefficient of determination of the car sales volume model. This paper concludes that the simplified computation of matrix- driven fuzzy linear regression has successfully tested to explain the relationship between car sales volume and the four dependent variables.

REFERENCES

Bargiela, A., W. Pedrycz and T. Nakashima, 2007. Multiple regression with fuzzy data. Fuzzy Sets Syst., 158: 2169-2188.

Chang, P.T. and E.S. Lee, 1996. A generalized fuzzy weighted least-squares regression. Fuzzy Sets Syst., 82: 289-298.

Chen, Y.S., 2001. Outliers detection and confidence interval modification in fuzzy regression. Fuzzy Sets Syst., 119: 259-272.

Choi, S.H. and J.J. Buckley, 2008. Fuzzy regression using least absolute deviation estimators. Soft Comput.: Fusion Found. Methodol. Appl., 12: 257-263.

D'Urso, P. and T. Gastaldi, 2000. A least-squares approach to fuzzy linear regression analysis. Comput. Stat. Data Anal., 34: 427-440.

D'Urso, P. and T. Gastaldi, 2001. Linear Fuzzy Regression Analysis with Asymmetric Spreads. In: Advances in Data Science and Classification, Borra, S., R. Rocci, M. Vichi and M. Schader (Eds.). Springer, Heidelberg, pp: 257-264.

Hook, W. and M. Replogle, 1996. Motorization and non-motorized transport in Asia: Transport system evolution in China, Japan and Indonesia. Land Use Policy, 13: 69-84.

Kahraman, C., 2002. An application of fuzzy linear regression to the information technology in Turkey. Int. J. Technol. Manage., 24: 330-339.

Khan, M.A. and L.G. Willumsen, 1986. Modelling car ownership and use in developing countries. Traffic Eng. Control, 27: 554-560.

Lam, W.H.K. and M.L. Tam, 2002. Reliability of territory-wide car ownership estimates in Hong Kong. J. Trans. Geogr., 10: 51-60.

Lee, H.T. and S.H. Chen, 2001. Fuzzy regression model with fuzzy input and output data for manpower forecasting. Fuzzy Sets Syst., 119: 205-213.

Pan, N.F., T.C. Lin and N.H. Pan, 2009. Estimating bridge performance based on a matrix-driven fuzzy linear regression model. Autom. Constr., 18: 578-586.

Prevedouras, P.D. and P. An, 1998. Automobile ownership in Asian countries: Historical trends and forecasts. Inst. Transp. Eng. J., 68: 24-29.

Sakawa, M. and H. Yano, 1992. Fuzzy Linear Regression and its Application. In: Studies in Fuzziness, Fuzzy Regression Analysis, Kacprzyk, J. and M. Fedrizzi (Eds.)., 3rd Edn., Vol. 1, Omnitech Press, Warsaw, Poland, pp: 61-80.

Tanaka, H. and H. Lee, 1998. Interval regression analysis by quadratic programming. IEEE Trans. Fuzzy Syst., 6: 473-481.

Tanaka, H., S. Uejima and K. Asis, 1982. Linear regression analysis with fuzzy model. IEEE Trans. Syst. Man. Cybernet., 12: 903-907.

Wu, H., I. Okutani and J. Xu, 2009. Identification of band reflectance and land-cover classification using fuzzy linear regression analysis. Proceedings of the 6th International Conference on Fuzzy Systems and Knowledge Discovery, August 14-16, 2009, IEEE., India, pp: 114-118.

Wu, H.C., 2003. Fuzzy estimates of regression parameters in linear regression models for imprecise input and output data. Comput. Stat. Data Anal., 42: 203-217.

Xu, R. and C. Li, 2001. Multidimensional least-squares fitting with a fuzzy model. Fuzzy Sets Syst., 119: 215-223.

Zadeh, L.A., 1965. Fuzzy sets. Inform. Control, 8: 338-358.