Efficiency Comparison of the Parameters Estimation by a Fuzzy Linear Regression Model

S Phanyaem

Department of Applied Statistics, Faculty of Applied Science, King Mongkut’s University of Technology North Bangkok, Bangkok, 10800 Thailand

Corresponding author: suvimol.p@sci.kmutnb.ac.th

Abstract. The fuzzy regression method can be used for fitting the fuzzy data for the construction of the simple linear regression model. This paper aims to study the performance and compare the adaptive fuzzy method and traditional fuzzy method by using the criteria of mean square error (MSE) and the coefficient of determine $R^2$. A Monte Carlo simulation with 5,000 iterations was performed to achieve the objective of research. In simulations, the independent variable has a gamma distribution with the parameters $\alpha = 4, 5, 6$ and $\beta = 1$. We carried out simulations to examine the performance of MSE from the adaptive fuzzy regression model and the traditional fuzzy regression model. The results showed that the transformation of independent variable with gamma distribution to be a normal distribution make the simple linear regression by the adaptive fuzzy method outperforms the traditional fuzzy method. In terms of MSE, the adaptive fuzzy method performed better than the traditional fuzzy method. In addition, the values of coefficient of determine from the adaptive fuzzy method and the traditional fuzzy method are slightly different.

1. Introduction

Regression analysis is a statistical method used for estimating the relationship between the dependent and other independent variables. The regression analysis model was used to forecast the value of the dependent variable. Linear regression analysis is the most popular forecasting method and is widely used in variety of fields such as medicine, business, industry, etc. However, in some situations the relationship of variables may have a nonlinear relationship model. For this reason, predictive methods have been developed by creating a linear regression model, the fuzzy method. Fuzzy linear regression analysis is one of the widely used methods to estimate the parameters in fuzzy data.

Fuzzy logic is a mathematical method for making decisions under ambiguity or ambiguity, similar to human cognitive logic. Fuzzy logic theory was proposed by Zadeh [1], which relied on a fuzzy set to convey uncertainty. By fuzzy theory, this set defines the membership value between 0 and 1. The degree of membership of the variable of interest is based on the membership function, which is of many types, such as the triangle function, trapezoidal function, sigmoid function and Gaussian functions, etc. The choice of member functions depends on the data of the variable.

Many researchers discussed the fuzzy methods for estimating the parameters in the case of linear regression models. For example, Buckley and Eslami [2], [3] suggested a new method to estimate fuzzy parameters using a set of confidence intervals based on a family of alpha-cut. Kim et al. [4] constructed the fuzzy linear regression model using the fuzzy least absolute deviation method. Arabpour and Tata
[5] suggested the fuzzy methods for estimating the parameters of linear regression models in the case of simple and multivariate linear regression models. They use the metric of distance between two triangular fuzzy numbers and generalize it to trapezoidal fuzzy numbers. The results showed that this method has a smaller the sum of errors of estimation than the traditional method. Chen and Hsueh [6] proposed a method for estimating the coefficient of regression model using the least square method based on the concept of distance. In addition, they suggested the confidence interval of linear regression model based on a few alpha-cuts. Rattanalertnusorn [7] constructed a fuzzy linear regression model using LR-fuzzy number and estimated the model parameters. Recently, Rattanalertnusorn et al. [8] extended the Chen and Hsueh method in fuzzy linear regression model in which the explanatory variable and response variable are trapezoidal fuzzy numbers.

Consequently, the aim of this paper is to propose the traditional fuzzy method and the adaptive fuzzy method for estimating the parameter of simple linear regression model using data transformation of independent variable based on gamma distribution to normal distribution. For the criteria for comparing the efficiency of parameter estimation methods will consider the mean square error (MSE) and the coefficient of determination $R^2$. The organization of this paper is as follows: In Section 1, the introduction of linear regression model is described. The methods for estimating the parameter of linear regression model are proposed in Section 2. In Section 3, we described the research procedure. In Section 4, the efficiency of parameter estimation by the adaptive fuzzy method is compared with the traditional fuzzy method. Finally, in Section 5 provides a conclusion.

2. Methods for Estimate the Parameter of Fuzzy Linear Regression Model

The objective of this paper is to compare the methods of estimation the parameters in simple linear regression model between traditional fuzzy method and adaptive fuzzy method. This section describes the statistical methods used to estimate parameters of simple linear regression model including the traditional fuzzy method and the adaptive fuzzy method.

2.1. The Simple Linear Regression Model

In the usual regression, the simple linear regression model is as follows:

$$y_i = \beta_0 + \beta_i x_i + \epsilon_i ; \ i = 1, 2, \ldots, n$$  \hspace{1cm} (1)

where $y$ is the response variable

$x$ is the explanatory variable.

The least square estimators of $\beta_0$ and $\beta_i$ are denoted by $\hat{\beta}_0$ and $\hat{\beta}_i$ respectively.

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$  \hspace{1cm} (2)

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^{n} x_i^2 - n \bar{x}^2}$$  \hspace{1cm} (3)

where $\bar{x}$ is the mean of independent variable

$\bar{y}$ is the mean of dependent variable.

2.2. Traditional Fuzzy Linear Regression Model

In this subsection, we let $y$ is the dependent variable and $x$ is the independent variable. More generally, a traditional fuzzy linear regression model can be written as:

$$y_i = \beta_0 + \beta_i (x_i - \bar{x}) + \epsilon_i ; \ i = 1, 2, \ldots, n$$  \hspace{1cm} (4)

where $\beta_0$ is termed as an intercept parameter.
\( \beta_i \) is termed as the slope parameter
\( \epsilon \) is termed as an error.

These parameters are usually called as regression coefficients. We assume that \( \epsilon_i \) are independent and identically distributed following a normal distribution with mean zero and constant variance \( \sigma^2 \). Next, we find the confidence interval for the intercept and slope parameters of linear regression model, we obtain the following estimates of the \( \beta_0 \) and \( \beta_1 \).

The \((1-\alpha)100\%\) confidence interval for an intercept parameter; \( \beta_0 \) as follows:

\[
\left[ \hat{\beta}_0 - t_{a/2,n-2} \sqrt{\hat{\sigma}^2 / (n-2)}, \hat{\beta}_0 + t_{a/2,n-2} \sqrt{\hat{\sigma}^2 / (n-2)} \right]
\]

where \( \hat{\beta}_0 \) is a point estimator of \( \beta_0 \)
\( \hat{\sigma}^2 \) is point estimator of \( \sigma^2 \)
\( t_{a/2,n-2} \) is \( \alpha / 2 \) the percent point of the \( t \)-distribution with \( n - 2 \) degrees of freedom.

The point estimator of \( \beta_0 \) is calculated by
\[
\hat{\beta}_0 = \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i ; \ i = 1,2,....
\]

and the point estimator of \( \sigma^2 \) is calculated by
\[
\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \left[ y_i - \hat{\beta}_0 - \hat{\beta}_1 (x_i - \bar{x}) \right]^2}{n} ; \ i = 1,2,....
\]

The \((1-\alpha)100\%\) confidence interval for the slope parameter; \( \beta_1 \) as follows

\[
\left[ \hat{\beta}_1 \pm t_{a/2,n-2} \sqrt{\frac{n \hat{\sigma}^2}{(n-2) \sum_{i=1}^{n} (x_i - \bar{x})^2}} \right]
\]

where \( \hat{\beta}_1 \) is a point estimator of \( \beta_1 \).

In this case, the point estimate of the parameter \( \beta_i \) is calculated by
\[
\hat{\beta}_i = \frac{\sum_{i=1}^{n} y_i (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}
\]

Therefore, the \((1-\alpha)100\%\) confidence interval of \( \sigma^2 \) is calculated by

\[
\left[ \frac{n \hat{\sigma}^2}{L(\lambda)}, \frac{n \hat{\sigma}^2}{R(\lambda)} \right]
\]

where \( L(\lambda) = [1-\lambda] \chi^2_{n-2,0.005} + \lambda n \)

\( R(\lambda) = [1-\lambda] \chi^2_{n-1,0.005} + \lambda n \).

Thus, the traditional fuzzy linear regression model is defined by
\[ y_i = \hat{\beta}_0 + \hat{\beta}_1 (x_i - \bar{x}) \]  

(11)

where \( \hat{\beta}_0 \) is the estimator of an intercept parameter using traditional fuzzy method

\( \hat{\beta}_1 \) is the estimator of a slope parameter using traditional fuzzy method.

**2.3. Adaptive Fuzzy Linear Regression Model**

In this study, an adaptive fuzzy model was investigated as an alternative method for approximating a simple linear regression model. The adaptive fuzzy linear regression model was originally proposed to estimate the regression model (see Arabpour and Tata [5]). The adaptive fuzzy linear regression model is as follows:

\[ y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\varepsilon}_i; \quad i = 1, 2, \ldots, n \]  

(12)

where \( \hat{\beta}_0 \) is the intercept parameter of adaptive fuzzy linear regression model

\( \hat{\beta}_1 \) is the slope parameter of adaptive fuzzy linear regression model.

In the present paper, we use the adaptive fuzzy theory for estimating the parameters of linear regression model. Thus, the adaptive fuzzy estimators; \( \hat{\beta}_0^* \) and \( \hat{\beta}_1^* \) can be calculated by (13) and (14), respectively.

\[ \hat{\beta}_0^* = (\hat{\beta}_{00}, \hat{\beta}_{01}, \hat{\beta}_{0r}) \]  

(13)

\[ \hat{\beta}_1^* = (\hat{\beta}_{10}, \hat{\beta}_{11}, \hat{\beta}_{1r}) \]  

(14)

where \( \hat{\beta}_0^* \) is the intercept estimator of adaptive fuzzy linear regression model

\( \hat{\beta}_1^* \) is the slope estimator of adaptive fuzzy linear regression model.

Therefore, the adaptive fuzzy estimators of the intercept parameter \( \hat{\beta}_0 \) are calculated as:

\[ \hat{\beta}_{0l} = \bar{y}_l - \hat{\beta}_1 \bar{x}_l, \quad \hat{\beta}_{0m} = \bar{y}_m - \hat{\beta}_1 \bar{x}_m, \quad \hat{\beta}_{0r} = \bar{y}_r - \hat{\beta}_1 \bar{x}_r \]  

(15)

And the adaptive fuzzy estimators of the slope parameter \( \hat{\beta}_1 \) are calculated as:

\[ \hat{\beta}_{0l} = \frac{\sum_{i=1}^{n_l} x_{il} y_{il} - n \bar{x}_l \bar{y}_l}{\sum_{i=1}^{n_l} x_{il}^2 - n \bar{x}_l^2}, \quad \hat{\beta}_{0m} = \frac{\sum_{i=1}^{n_m} x_{im} y_{im} - n \bar{x}_m \bar{y}_m}{\sum_{i=1}^{n_m} x_{im}^2 - n \bar{x}_m^2}, \quad \hat{\beta}_{0r} = \frac{\sum_{i=1}^{n_r} x_{ir} y_{ir} - n \bar{x}_r \bar{y}_r}{\sum_{i=1}^{n_r} x_{ir}^2 - n \bar{x}_r^2} \]  

(16)

**3. The Procedure of Construction the Linear Regression Model**

In this section, we used the R program for generating the data. There are operating procedures for construction the traditional fuzzy linear regression model as follows:

**Step 1:** Simulate the fuzzy dataset of independent variables \( (x_i, x_m, x_r) \) and the dependent variable \( (y_l, y_m, y_r) \).

**Step 2:** Checking the pattern of relationship between independent and dependent variables.

**Step 3:** Estimating the mean of the independent variables \( (x_i, x_m, x_r) \).

**Step 4:** By the traditional fuzzy linear regression model, we estimate the regression coefficient \( \hat{\beta}_0^* \) and \( \hat{\beta}_1^* \), after that we create a traditional fuzzy linear regression model: \( y_i = \hat{\beta}_0^* + \hat{\beta}_1^* (x_i - \bar{x}) \).

**Step 5:** Substitute the independent variable \( (x) \) in the traditional fuzzy linear regression equation: \( y_i = \hat{\beta}_0^* + \hat{\beta}_1^* (x_i - \bar{x}) \) to estimate the variable based on \( (y) \).
Step 6: We show the effectiveness of traditional fuzzy method by calculate the mean square error \((MSE)\) and the determination coefficient of the traditional fuzzy linear regression equation \((R^2)\) and repeat the experiment \((M)\) equal to 5,000 times.

The algorithm of adaptive regression model as follows:

Step 1: Firstly, we simulate the fuzzy data including the predictor variable \(\left(x_l, x_m, x_r\right)\) and the response variable \(\left(y_l, y_m, y_r\right)\).

Step 2: To check the relationship between the predictor variable and the response variable.

Step 3: To identify the parameter of linear regression model using the traditional method and adaptive method.

Step 4: To construct the linear regression model and put a predictor value \((x_i)\) into linear regression model, we will get the prediction value of \((y_i)\).

Step 5: Finally, we compute a mean square error and a coefficient of determination of adaptive fuzzy linear regression model.

4. Numerical Results

In this section, a numerical study is performed in order to study the performance of the estimators for simple linear regression model. Here, Monte Carlo simulation are done via R Program. We used the MC simulation to generate the fuzzy data and performed based on traditional fuzzy and adaptive fuzzy methods. For each situation of the corresponding estimators have been replicated 5,000 times. The variables used in this study are independent and dependent variables. To determine the independent variable to have a gamma distribution, the parameter values for gamma distribution are selected as \((\alpha, \beta) = (4, 1), (5, 1)\) and \((6, 1)\). The number of sample sizes are selected as \(n = 30, 40, 50, \ldots, 150\) respectively. In this paper, we use the mean square error \((MSE)\) and the coefficient of determination \((R^2)\) are the criteria for comparing the efficiency of a linear regression model between the conventional fuzzy method and the improved fuzzy method. The mean square error can be represented

\[
MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n-2}
\]  

(17)

where \(y_i\) is the observation value
\(\hat{y}_i\) is the predictive value.

The \(MSE\) value is large, indicating that the estimated regression equation is less appropriate.

The \(MSE\) value is small, indicating that the estimated regression equation is very appropriate.

The \(MSE\) value is 0, indicating that all values \(y_i\) are on the regression equation line.

The coefficient of determination can be calculated

\[
R^2 = \frac{\sum_{i=1}^{n}x_iy_i - n\bar{x}\bar{y}}{\left(\sum_{i=1}^{n}x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^{n}y_i^2 - n\bar{y}^2\right)}
\]  

(18)

where \(\bar{x}\) is the mean of independent variable
\(\bar{y}\) is the mean of dependent variable.

The value \(R^2\) is between 0 and 1. In the case of \(R^2\) approach 1, this means that the regression model can explain the value of the dependent variable well or independent variables in the regression equation can be described the change of the dependent variable was good, since the dependent and independent
variables are high correlation. Whereas they approach 0, the regression equation does not explain the value of the dependent variable or the independent variable in that regression equation poorly described changes in the dependent variable, since the dependent variable and independent variables have low correlation. Simulation results are given in Tables 1 to 3 and explored in some figures. In Table 1, show the results of the comparison of $MSE$ values and $R^2$ of their approximation of simple linear regression parameters, traditional and adaptive fuzzy methods, with gamma distribution ($\alpha = 4, \beta = 1$). In Table 2, show the results of the comparison of $MSE$ values and $R^2$ of their approximation of simple linear regression parameters, traditional fuzzy methods and adaptive fuzzy methods, with gamma distribution data ($\alpha = 5, \beta = 1$). In Table 3, show the results of the comparison of $MSE$ values and $R^2$ of their approximation of simple linear regression parameters, traditional fuzzy methods and adaptive fuzzy methods, with gamma distribution data ($\alpha = 6, \beta = 1$).

**Table 1.** Values of $MSE$ and $R^2$ of Traditional Fuzzy and Adaptive Fuzzy Methods for the Parameter Values of Gamma Distribution ($\alpha = 4, \beta = 1$).

| Sample size $n$ | Traditional Fuzzy method | Adaptive Fuzzy method |
|-----------------|---------------------------|-----------------------|
| $MSE$           | $R^2$                     | $MSE$                 | $R^2$                 |
| 30              | 0.3369                    | 0.9974                | 0.3339                | 0.9967                |
| 40              | 0.3353                    | 0.9963                | 0.3335                | 0.9975                |
| 50              | 0.3349                    | 0.9924                | 0.3333                | 0.9928                |
| 60              | 0.3343                    | 0.9974                | 0.3323                | 0.9968                |
| 70              | 0.3341                    | 0.9945                | 0.3330                | 0.9969                |
| 80              | 0.3337                    | 0.9932                | 0.3324                | 0.9943                |
| 100             | 0.3334                    | 0.9913                | 0.3327                | 0.9925                |
| 110             | 0.3342                    | 0.9935                | 0.3331                | 0.9947                |
| 120             | 0.3335                    | 0.9974                | 0.3325                | 0.9969                |
| 130             | 0.3328                    | 0.9959                | 0.3318                | 0.9968                |
| 140             | 0.3315                    | 0.9963                | 0.3307                | 0.9972                |
| 150             | 0.3299                    | 0.9978                | 0.3271                | 0.9989                |

**Table 2.** Values of $MSE$ and $R^2$ of Traditional Fuzzy and Adaptive Fuzzy Methods for the Parameter Values of Gamma Distribution ($\alpha = 5, \beta = 1$).

| Sample size $n$ | Traditional Fuzzy method | Adaptive Fuzzy method |
|-----------------|---------------------------|-----------------------|
| $MSE$           | $R^2$                     | $MSE$                 | $R^2$                 |
| 30              | 0.3363                    | 0.9980                | 0.3321                | 0.9982                |
| 40              | 0.3351                    | 0.9983                | 0.3318                | 0.9983                |
| 50              | 0.3343                    | 0.9985                | 0.3302                | 0.9987                |
| 60              | 0.3345                    | 0.9983                | 0.3325                | 0.9985                |
| 70              | 0.3347                    | 0.9981                | 0.3322                | 0.9985                |
| 80              | 0.3327                    | 0.9989                | 0.3301                | 0.9991                |
| 100             | 0.3331                    | 0.9988                | 0.3311                | 0.9989                |
| 110             | 0.3315                    | 0.9985                | 0.3291                | 0.9986                |
| 120             | 0.3311                    | 0.9993                | 0.3289                | 0.9998                |
| 130             | 0.3319                    | 0.9990                | 0.3286                | 0.9995                |
| 140             | 0.3309                    | 0.9991                | 0.3273                | 0.9993                |
| 150             | 0.3219                    | 0.9995                | 0.3212                | 0.9998                |
Table 3. Values of $MSE$ and $R^2$ of Traditional Fuzzy and Adaptive Fuzzy Methods for the Parameter Values of Gamma Distribution ($\alpha = 6, \beta = 1$).

| Sample size | Traditional Fuzzy method | Adaptive Fuzzy method |
|-------------|--------------------------|-----------------------|
| $n$         | $MSE$ | $R^2$ | $MSE$ | $R^2$ |
| 30          | 0.3359 | 0.9982 | 0.3347 | 0.9979 |
| 40          | 0.3348 | 0.9979 | 0.3335 | 0.9980 |
| 50          | 0.3345 | 0.9978 | 0.3336 | 0.9982 |
| 60          | 0.3343 | 0.9982 | 0.3326 | 0.9980 |
| 70          | 0.3338 | 0.9985 | 0.3329 | 0.9988 |
| 80          | 0.3332 | 0.9983 | 0.3325 | 0.9985 |
| 100         | 0.3328 | 0.9986 | 0.3321 | 0.9989 |
| 110         | 0.3335 | 0.9984 | 0.3326 | 0.9987 |
| 120         | 0.3335 | 0.9983 | 0.3325 | 0.9980 |
| 130         | 0.3331 | 0.9987 | 0.3324 | 0.9985 |
| 140         | 0.3325 | 0.9989 | 0.3313 | 0.9988 |
| 150         | 0.3302 | 0.9992 | 0.3298 | 0.9994 |

Figure 1. Value of $MSE$ of traditional fuzzy and adaptive fuzzy methods for the parameter $\alpha = 4, \beta = 1$ of gamma distribution.

Figure 2. Value of $MSE$ of traditional fuzzy and adaptive fuzzy methods for the parameter $\alpha = 5, \beta = 1$ of gamma distribution.

Figure 3. Value of $MSE$ of traditional fuzzy and adaptive fuzzy methods for the parameter $\alpha = 6, \beta = 1$ of gamma distribution.
From Table 1 to Table 3, it was also found that the mean square error ($MSE$) values from the simple linear regression model by the adaptive fuzzy method were less than the traditional fuzzy method for all situations of sample size values. The $MSE$ for different values of the gamma parameters are shown in Figures 1 - 3. Figures 1-3 show that the adaptive fuzzy method less mean square error than the traditional fuzzy method in all situations. However, the values of coefficient of determination ($R^2$) from the adaptive fuzzy method and the traditional fuzzy methods are slightly different.

5. Conclusion
In this paper, we compare the performances of traditional and adaptive fuzzy estimation methods under simple linear regression models. The performances of regression estimators are compared by two criterions, the mean square error ($MSE$) and the coefficient of determination ($R^2$). The Monte-Carlo simulation study suggests the use of adaptive fuzzy estimation in case of data follows a gamma distribution. The results showed that the transformation of independent variable with gamma distribution to be a normal distribution make the simple linear regression by the adaptive fuzzy method outperforms the traditional fuzzy method. Finally, the main conclusion one should draw from the paper’s results is that adaptive fuzzy method is better than traditional fuzzy method in term of minimum $MSE$ and high coefficient of determine.

6. References
[1] Zadeh LA 1965 Fuzzy sets Infor. Control. 8 pp 338-353
[2] Buckley JJ and Eslami E 2003a Uncertain probabilities I: the discrete case Soft Comput. 7 pp 500-505
[3] Buckley JJ and Eslami E 2003b Uncertain probabilities II: the continuous case Soft Comput. 7 pp 500-505
[4] Kim KI, Kim DH and Choi SH 2005 Least absolute deviation estimator in fuzzy regression J. Appl. Math. & Comput. 18 pp 649-656
[5] Arabpour AR and Tata M 2008 Estimation the parameter of a fuzzy linear regression model Iran. J. Fuzzy Syst. 5 pp 1–19
[6] Chen LH and Hsueh CC 2009 Fuzzy regression models using the least-squares method based on the concept of distance IEEE Trans. Fuzzy Syst. 17 pp 1259-72
[7] Rattanalertnusorn A 2016 Fuzzy linear regression model based on alpha-cuts approach Proc. Int. Conf. on Applied Statistics 13-15 July 2016 (Phuket) pp 79-85
[8] Rattanalertnusorn A, Thongteeraparp A and Bodhisuwan W 2014 Parameter estimation of fuzzy linear regression model: The extension of Chen and Hsueh J. Appl. Sci. 14 pp 631-640

Acknowledgments
This research was funded by King Mongkut’s University of Technology North Bangkok Contract no. KMUTNB- 61-DRIVE-025. The authors thank the editor and the reviewer for careful reading of the research article and constructive comments that greatly improved this paper.