Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 8, Issue, 02, pp.27090-27093, February, 2016

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

A COMPARATIVE STUDY OF FUZZY LEAST SQUARE AND POSSIBILITY REGRESSION ANALYSIS AND ITS RELIABILITY

*Ubale, A. B., Sananse, S. L. and Chhaya Sonar

Department of Statistics, Dr. B. A. M. University, Aurangabad-431004(M.S) India

ARTICLE INFO

Article History:
Received 07th November, 2015
Received in revised form
20th December, 2015
Accepted 29th January, 2016
Published online 27th February, 2016

Key words:

Fuzzy least square, Possibility regression, Hybrid Standard error, Hybrid Correlation, Reliability measure.

ABSTRACT

Regression analysis is a powerful forecasting technique in various field. It is used in different fields viz. Agriculture, Economic, industries, Social Sciences. The fuzzy regression is a nonparametric method. When dataset is very small, uncertain, fuzzy way and qualitative in nature then fuzzy regression method is appropriate .It is widely used in management science and engineering field, also in this paper, we have applied two methods viz. fuzzy least square method and possibility regression analysis with numerical example. These two methods are compared by using hybrid correlation coefficient and reliability performance. On the basis of final results obtained, it is concluded that the fuzzy least square is better than possibilistic regression analysis.

Citation: **Ubale, A. B., Sananse, S. L. and Chhaya Sonar, 2016.** "A comparative study of fuzzy least square and possibility regression analysis and its reliability", *International Journal of Current Research,* 8, (01), xxxxx-xxxxx.

INTRODUCTION

The linear regression analysis is widely used as forecasting technique in most of the cases. The regression analysis explain the variation of response variable (Y) in the variation of explanatory variable (X). The general model, Y= f(X) where f(x) is the linear function. The linear regression analysis should follows Gaussian assumption it means that error terms has mean zero and unit variance. The linear regression model can be applied only if the dataset relation between X and Y are crisp. In fuzzy regression model, if a data is crisp or fuzzy independent variables and one or more dependent fuzzy variables are important in the model. This relationship can be studied by using fuzzy linear regression techniques. The fuzzy linear regression model is first proposed by (Tanaka *et al*., 1982), while the relation between the explanatory variables and response variable in data is introduced by a fuzzy function of which the distribution of the parameter is a possibility function proposed by (Tanaka, 1987). The deviations are between observed and estimated values were supposed to be due to the fuzziness of system structure then this structure was represented as a fuzzy linear function whose parameters are given by fuzzy sets. by (Diamond, 1988).

Corresponding author: Ubale, A. B.
Department of Statistics, Dr. B. A. M. University, Aurangabad-431004(M.S) India.

Literature Review

The possibilistic fuzzy regression model was proposed by Tanaka (1982), determined the regression coefficient and minimized total fuzziness in the given data. Diamond (1988) introduced fuzzy regression model to minimize sum of squares of differences for the center of fuzzy numbers and sum of squares of differences for spreads. Pierpaolo D'Urso, *et al*. (2000) introduced new approach of fuzzy linear regression analysis. They developed doubly linear adaptive fuzzy regression model, based on two linear models such as center regression model and a spread regression model. They observed that doubly linear adaptive fuzzy regression analysis had alternative methods for fuzzy linear regression analysis. Volker Krätschmer (2006) developed new fuzzy linear regression models. He had generalized the type of single ordinary equation in linear regression models by incorporated the physical vagueness of the involved items in the form of fuzzy data for the variables. Finally suggested that ordinary least-squares method was greater flexibility for modeling and estimation. Seung Hoe Choi, *et al*. (2008), proposed least absolute deviation estimators to construct fuzzy regression models. They had compared a fuzzy regression model and used the least absolute deviation estimators and another fuzzy regression model used the least square method. They were shown that the fuzzy least absolute deviation model was better performance than the fuzzy regression based on least squares method in case the data contains fuzzy outliers. Furkan Baseret *et al.* (2010) applied hybrid fuzzy least squares regression analysis to predict future claim costs by used the concept of London Chain Ladder (LCL) method. They had suggested that the hybrid fuzzy least-squares regression model was taken both randomness and fuzziness type of uncertainty into a regression model. Ubale and Sananse (2015) introduced brief research trends in application of fuzzy regression analysis in different field. Yun-His, *et al.* (2001) developed hybrid fuzzy least-squares regression by using weighted fuzzy-arithmetic mean and least-squares fitting criterion. They compared hybrid regression with the ordinary regression and other fuzzy regression methods.

Finally they suggested that hybrid fuzzy regression model satisfied a limiting behavior that the fuzziness decreases, the equations were similar to the results of the ordinary regression. RuoningXu, *et al.* (2001), developed fuzzy multivariable linear regression by least squares and also the methods was similar to the traditional least-squares method. They had introduced a fuzzy analogue by a distance defined on a fuzzy number space, and fuzzy multivariable least square linear regression model. They showed that the model had a unique solution and the solution was given in an analytic expression.

Research Methodology

Fuzzy linear regression analysis is first proposed by Tanaka (1982), and applied for many researches. The model of Hsiao-Fan Wang *et al.* (2000) is given below.

$$\tilde{Y} = \tilde{A}_0 X_0 + \tilde{A}_1 X_1 + \cdots + \tilde{A}_N X_N = \tilde{A}X \quad \ldots\ldots(1)$$

where $X = [X_0, X_1, \ldots. X_N]^T$ is a vector of independent variables, $\tilde{A} = [\tilde{A}_0, \tilde{A}_1, \ldots. \tilde{A}_N]^T$ is a vector of fuzzy coefficient presented in the form of triangular fuzzy numbers denoted by $\tilde{A}_j = (\alpha_j, c_j)$ with its membership function described as

$$\mu_{\tilde{A}_j}(a_j) = \begin{cases} 1 - \frac{|\alpha_j - a_j|}{c_j} & , \alpha_j - c_j \le a_j \le \alpha_j + c_j, \forall j = 1,2,\ldots,N \\ 0, otherwise & \end{cases} \ldots(2)$$

Where $\alpha_j$the central is value and c_j is the spread value.

Therefore the eq (1) can be written as

$$\tilde{Y}_i = (\alpha_0, c_0) + (\alpha_1, c_1)X_1 + (\alpha_2, c_2)X_2 + \cdots + (\alpha_N, c_N)X_N$$

The above fuzzy regression analysis assumes the crisp input and output data, while the relation between the input and output data is defined by a fuzzy function of which the distribution of the parameter is a possibility function (Tanaka, 1987). By applying the Extension Principle (Zadeh, 1975), it derives the membership function of fuzzy number $\tilde{Y}_i$ as shown in (3) and each value of dependent variable can be estimated as a fuzzy number $\tilde{Y}_i = (Y_i^{L}, Y_i^{h=1}, Y_i^{U})$ $i = 1,2,\ldots,M$ where the lower bound of $\tilde{Y}_i$ is $Y_i^{L} = \sum_{j=0}^{N}(\alpha_j - c_j)X_{ij}$, the central value of $\tilde{Y}_i$ is $Y_i^{h=1} = \sum_{j=0}^{N}\alpha_j X_{ij}$ and the upper bound of $\tilde{Y}_i$ is $Y_i^{U} = \sum_{j=0}^{N}(\alpha_j + c_j)X_{ij}$.

$$\mu(Y_i) = \begin{cases} 1 - \frac{|Y_i - X^t\alpha|}{c^t|X|} & X \neq 0, \\ 1 \quad X = 0, Y \neq 0 \ \forall i = 1,2,\ldots,M. & \\ 0, \quad X = 0, Y = 0 & \end{cases} \quad \ldots\ldots(3)$$

the degree of the fitting of estimated fuzzy linear model

$\tilde{Y} = \tilde{A}X$ to the given data $Y_i = (y_i, e_i)$, is measured by following $\overline{h_1}$ which maximizes h subject to $Y_i^h \subset Y_i^{*h}$, where

$$Y_i^h = \{y|\mu(Y_i) \ge h\}$$

$$Y_i^{*h} = \{y|\mu(Y_i^*) \ge h\}$$

Where Y_i^* is fitted value and the constraints require that each observation Y_i has at least h degree of belonging to (Tanaka *et al.*, 1982). The problem is to find out the fuzzy parameters $\tilde{A}_j = (\alpha_j, c_j)$, to determine the fuzzy parameters while minimizing the total sum of the spreads of the estimated values for a certain h level, Tanaka and Watada (1988) and Tanaka *et al.* (1989) formulated a linear programming problem as follows:

$$Min \sum_{i=1}^{m}\sum_{j=0}^{N}(c_j|X_{ij}|)$$

s. t.

$$\sum_{j=0}^{N}\alpha_j X_{ij} + (1-h)\sum_{j=0}^{N}c_j|X_{ij}| \ge Y_i + (1-h)e_i \quad i1,2,\ldots,M$$

$$\sum_{j=0}^{N}\alpha_j X_{ij} - (1-h)\sum_{j=0}^{N}c_j|X_{ij}| \le Y_i(1-h)e_i \quad i = 1,2,\ldots,M \quad \ldots(4)$$

$$c_j \ge 0, a \in R, X_{i0} = 1, (0 \le h \le 1)$$

The multidimensional least –square fitting with fuzzy data is based on distance criteria and estimating fuzzy parameters (Ruoning Xu and Chulin Li, 2001) is used in numerical example. Suppose that the observation data are denoted as $(\tilde{y}_i, X_{i1}, \ldots, X_{iN}), (i = 1,2,\ldots,M; M > N)$, where $\tilde{y}_i \epsilon N$ $(i = 1,2,\ldots,M)$and $X_{ij} \epsilon R$ $(i = 1,2,\ldots,M, j = 1,2,\ldots,N)$ without loss of generality, we can assume that $X_{ij} > 0$

We will consider the following model

$$\tilde{Y} = \tilde{A}_0 + \tilde{A}_1 X_1 + \tilde{A}_2 X_2 + \cdots + \tilde{A}_N X_N, \ \tilde{A}_j \epsilon N, \ (j = 0,1,2,\ldots,N) \ (5)$$

Which is to be fitted to the observation data in the sense of best fit with respect to the distance $\tilde{d}$. Namely, consider the least-squares optimization problem.

$$minimize\ M(\tilde{A}_0, \tilde{A}_1, \ldots, \tilde{A}_N) = \sum_{i=1}^{M}\tilde{d}^2(\tilde{A}_0 + \tilde{A}_1 X_{i1} + \cdots + \tilde{A}_N X_{iN}, \tilde{y}_i) \quad (6)$$

Denoting $\tilde{y}_i = (\tilde{y}_i, S_i)$ $((i = 1,2,\ldots,M)$ and $\tilde{A}_j = (a_j, \sigma_j)$ $j = 1,2,\ldots,N$ We have

$$\tilde{A}_0 + \tilde{A}_1 X_{i1} + \cdots + \tilde{A}_N X_{iN} = (a_0 + a_1 X_{i1} + \cdots + a_N X_{iN}, \sigma_0 + \sigma_1 X_{i1} + \cdots + \sigma_N X_{iN}) \quad (7)$$

The least square problem (6) can be written as

$$minimize\ M(\tilde{A}_0, \tilde{A}_1, \dots, \tilde{A}_N) = \sum_{i=1}^{M}(a_0 + a_1X_{i1} + \dots + a_NX_{iN,} - Y_i)^2 \frac{1}{2}\sum_{i=1}^{M}(\sigma_0 + \sigma_1X_{i1} + \dots + \sigma_NX_{iN,} - s_i)^2 \quad \dots (8)$$

If the above problem has a solution then the solution must satisfy the following equation

$$\frac{\partial M(\tilde{A}_0, \tilde{A}_1, \dots, \tilde{A}_N)}{\partial a_j} = 0 \ and\ \frac{\partial M(\tilde{A}_0, \tilde{A}_1, \dots, \tilde{A}_N)}{\partial \sigma_j} = 0 \quad (j = 0, 1, 2, \dots, N).$$

It follows that

$$\sum_{i=1}^{m}(a_0 + a_1X_{i1} + \dots + a_NX_{iN} - y_i)x_{ij} = 0$$

$$\sum_{i=1}^{m}(\sigma_{0+}\sigma_1 X_{i1} + \dots + \sigma_N X_{iN} - s_i)x_{ij} = 0 \quad (j = 0, 1, 2, \dots, N) \dots (9)$$

Where denoting $x_{i0} = 1\ (i = 1,2, \dots, m)$. Namely,

$$a_0\sum_{i=1}^{m} x_{i0}x_{ij} + a_1\sum_{i=1}^{m} x_{i1}x_{ij} + \dots + a_N\sum_{i=1}^{m} x_{iN}x_{ij} = \sum_{i=1}^{m} y_ix_{ij}$$
$$\sigma_0\sum_{i=1}^{m} x_{i0}x_{ij} + \sigma_1\sum_{i=1}^{m} x_{i1}x_{ij} + \dots + \sigma_N\sum_{i=1}^{m} x_{iN}x_{ij} = \sum_{i=1}^{m} s_ix_{ij} \quad (j = 0,1, \dots N). \quad \dots (10)$$

Let

$$A = \begin{pmatrix} \sum_{i=1}^{m} x_{i0}x_{i0} & \sum_{i=1}^{m} x_{i1}x_{i0} & \cdots & \sum_{i=1}^{m} x_{iN}x_{i0} \\ \sum_{i=1}^{m} x_{i0}x_{i1} & \sum_{i=1}^{m} x_{i1}x_{i1} & \cdots & \sum_{i=1}^{m} x_{iN}x_{i1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{m} x_{i0}x_{iN} & \sum_{i=1}^{m} x_{i1}x_{iN} & \cdots & \sum_{i=1}^{m} x_{iN}x_{iN} \end{pmatrix}$$

$$a = (a_0, a_1, \dots, a_N)^T,\ \ y = \left(\sum_{i=1}^{m} y_ix_{i0}, \sum_{i=1}^{m} y_ix_{i1}, \dots, \sum_{i=1}^{m} y_ix_{iN}\right)^T,$$

$$\sigma = (\sigma_0, \sigma_1, \dots, \sigma_N)^T,\ \ y = \left(\sum_{i=1}^{m} s_ix_{i0}, \sum_{i=1}^{m} s_ix_{i1}, \dots, \sum_{i=1}^{m} s_ix_{iN}\right)^T.$$

Then the equations (10) can be expressed as

$$Aa = y,\ \ A\sigma = s \quad \dots (11)$$
$$a = A^{-1}y,\ \sigma = A^{-1}s$$

Goodness of Fit

After estimating the regression coefficient by using fuzzy least square and possibilistic regression analysis. The Reliability and hybrid correlation coefficient, hybrid standard error of the hybrid regression equation is used (Yun-Hsi O. and Chang, 2001). For observed fuzzy data, the mean ($\bar{\bar{Y}}$) is calculated by using weighted fuzzy arithmetic as a crisp number. And, the standard deviation ($S_{\tilde{Y}}$) is calculated by using the following definition:

$$S_{\tilde{Y}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{\bar{Y}}_i)^2}$$

The value of $S_{\tilde{Y}}$ is a measure of dispersion or spread of data. Hybrid correlation coefficient (*HR*) is used to evaluate the linearity assumption of the hybrid linear regression model. A hybrid standard error of estimate (HS_e) is used to measure the goodness of fit between the hybrid regression model and the observed fuzzy data. Referring to the reliability measures for ordinary regression in (Ayyub and McCuen, 1996) the HR and HS_e are defined by using the weighted fuzzy arithmetic as follows:

Hybrid correlation coefficient

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

Hybrid standard error of estimates

$$HS_e = \sqrt{\frac{1}{n-p-1}\sum_{i=1}^{M}\left(\hat{Y}_i - \bar{Y}_i\right)^2}$$

In the above formula in which $n-p-1$ is the degrees of freedom for the regression analysis. The values of HS_e is smaller the better goodness of fit and better accuracy in the prediction. If HS_e is close to or greater than $S_{\tilde{Y}}$, the regression analysis has not been successful. Other regression methods may be needed in order to provide a better goodness of fit. Also since $S_{\tilde{Y}}$, is constant and independent from regression methods used the ratio $\frac{HS_e}{S_{\tilde{Y}}}$ is normalize measure of the goodness of fit. Therefore, HS_e and $\frac{HS_e}{S_{\tilde{Y}}}$ is used to evaluate the effectiveness of various regression methods.

Numerical Examples for Comparison of Studies

Portland cement industries such as, heat evolved in calories (per gram), amount of tricalcium aluminate, amount of tricalcium silicate, amount of tetracalcium alumno ferrite, of cement data used for this study. The data of Portland cement (Ruoning Xu and Chulin Li, 2001) is depicted in Table 1.

Table 1. Portland cement data

$\tilde{Y} = (y_i, s_i)$	X_{i1}	X_{i2}	X_{i3}
(78.5,6.9)	7	26	6
(74.3,6.4)	1	29	15
(104.3,9.4)	11	56	8
(87.6,7.8)	11	31	8
(95.5,8.6)	7	52	6
(109.2,9.9)	11	55	9
(102.7,9.3)	3	71	17
(72.5,6.2)	1	31	22
(93.1,8.3)	2	54	18
(115.9,10.6)	21	47	4
(83.8,7.4)	1	40	23
(113.3,10.6)	11	66	9
(109.4,9.9)	10	68	8

$\tilde{Y}$: Heat evolved in calories per gram of cement;

X_{i1} : amount of tricalcium aluminate;

X_{i2} : amount of tricalcium silicate;

X_{i3} : amount of tetracalcium alumno ferrite;

X_{i1}, X_{i2}, X_{i3} are measured as percent of the weight of the clinkers from which the cement was made.

By using the equation (4) we solved using Tora software as h=0.00 we have fitted equation is

$$\tilde{Y} = (49.05,1.6) + (2.03,.69)X_1 + (.57,0.00)X_2 + (.4,.41)X_3$$

And also equation (4) we solved by using Tora software as h=0.5 we have fitted equation is

$$\tilde{Y} = (46.78,0.00) + (2.05,0.95)X_1 + (.60,0.00)X_2 + (0.41,0.61)X_3$$

By using the formula (11) least-square equation is fitted.

$$\tilde{Y} = (47.905,3.79) + (1.704,0.1704)X_1 + (.6560,0.0668)X_2 + (.2660,0.0234)X_3$$

Table 2. Reliability measures for various regression models

Regression method	Regression equation	HR	HS_e	$\frac{HS_e}{S_{\tilde{Y}}}$
A.Fuzzy Least square regression	$\tilde{Y} = (47.905,3.79) + (1.704,0.1704)X_1 + (.6560,0.0668)X_2 + (.2660,0.0234)X_3$	.74	2.31	0.14
B.Possibility regression analysis (Tanaka's model) *h=0.0*	$\tilde{Y} = (49.05,1.6) + (2.03,.69)X_1 + (.57,0.00)X_2 + (.4,.41)X_3$	.74	3.60	.23
C.Possibility regression analysis (Tanaka's model) *h=0.5*	$\tilde{Y} = (46.78,0.00) + (2.05,0.95)X_1 + (.60,0.00)X_2 + (0.41,0.61)X_3$	.76	4.15	.26

From Table 2 it is revealed than the hybrid correlation coefficient was observed in fuzzy least square regression (0.74), Possibility regression analysis (h=0.0) (0.74) and Possibility regression analysis (h=0.5) (0.76). The hybrid standard error was observed for fuzzy least square regression (2.31), Possibility regression analysis (3.60), Possibility regression analysis (4.15). The ratio of $\frac{HS_e}{S_{\tilde{Y}}}$ is minimum in fuzzy least square method than possibility regression analysis (h=0 and h=0.5). On the basis of final results, it is concluded that fuzzy least square method is better than possibility regression analysis.

Conclusion

In this paper we have compared fuzzy least square regression and possibility regression analysis with the numerical example. The fuzzy least square can produce better prediction models in comparison to possibilistic regression models. Therefore it is proposed that the fuzzy linear regression used in fuzzy data and also used to similar type of data.

REFERENCES

Ayyub B.M., R. McCuen, 1996. Numerical Methods for Engineers, Prentice-Hall, New York.

Choi, S. H., Buckley, J. J., 2008. "Fuzzy Regression Using Least Absolute Deviation Estimators", Soft Computing, vol.12,No.3, pp.257-263.

Diamond P. 1988. Fuzzy least squares, *Inform. Sci*., 46, 141-157.

Draper N.R., H. Smith, 1980. Applied Regression Analysis, Wiley, New York.

D'urso, P. 2000. "Linear Regression Analysis For Fuzzy/Crisp Input And Fuzzy/ Crisp Output Data", Computational Statistics & Data Analysis, Vol.42, No.(1-2), pp.47-72.

Furkan Baser, Aysen Apaydin, 2010. "Calculating Insurance Claim Reserves With Hybrid Fuzzy Least Squares Regression Analysis," Gazi University Journal Of Science, vol.23,No.2, pp.163-170.

Hsiao-Fan Wang, Ruey-Chyn Tsaur, 2000. "Insight of a fuzzy regression model", Fuzzy Sets and Systems 112, 355-369.

Ruoning Xu, Chulin Li, 2001. "Multidimensional least-squares fitting with a fuzzy model", Fuzzy Sets and Systems 119, 215-223.

Tanaka H. 1987. Fuzzy data analysis by possibilistic linear models, Fuzzy Sets and Systems 24, 363-375.

Tanaka H., I. Hayashi, J. Watada, 1989. Possibilistic linear regression analysis for fuzzy data, *Eur. J. Operat. Res*., 40, 389–396.

Tanaka H., J. Watada, 1988. Possibilistic linear systems and their application to the linear regression model, Fuzzy Sets Syst. 27, 275–289.

Tanaka H., S. Uejima, K. Asai, 1982. Fuzzy linear model, fuzzy linear regression model, IEEE Trans. System Man Cybernet. 12, 903-907.

Ubale A.B., S. L. Sananse, 2015. "Fuzzy Regression Model and Its Application: A Review," *International Journal of Innovative Research in Science, Engineering and Technology*, Vol. 4, Issue 11, November 2015.

Volker Krätschmer, 2006. "Least-Squares Estimation In Linear Regression Models With Vague Concepts", Fuzzy Sets And Systems, vol. 157, pp.2579 – 2592.

Yun-Hsi O. Chang, 2001. "Hybrid fuzzy least-squares regression analysis and its reliability measures", Fuzzy Sets and Systems 119, 225-246

Zadeh, L.A. 1975. The concept of a linguistic variable and its application to approximate reasoning I, Inform. Sci. 8, 199-249

****