Advances in Fuzzy Sets and Systems

Published Online: February 2016
http://dx.doi.org/10.17654/FS021010007
Volume 21, Number 1, 2016, Pages 7-20 **ISSN: 0973-421X**

FORECASTING FUZZY LINEAR REGRESSION MODEL USING TIME SERIES DATASET

B. Onoghojobi and N. P. Olewuezi

Department of Statistics
Federal University of Technology
Owerri, Nigeria
e-mail: ngolewe@yahoo.com

Abstract

In this paper, we introduce the use of fuzzy linear regression in prediction and forecasting for a dataset. Using the basic idea underlying fuzzy regression gives computational efficiency of forecasting solutions.

1. Introduction

Predicting the future value is the primary issue when dealing with statistics and time series dataset. Classical statistical theory contains two types of uncertainty. These are uncertainty due to randomness and uncertainty due to fuzziness. Uncertainty due to fuzziness results from lack of relative distinction due to cognitive source, such as human estimation errors, abstract representation of a system and not enough data due to cost and other restrictions [8, 13].

A fuzzy type of conventional regression analysis has been proposed to evaluate the functional relationship between input and output variables in a

Received: March 23, 2015; Revised: July 8, 2015; Accepted: July 31, 2015

2010 Mathematics Subject Classification: Primary 62A86; Secondary 62B86.

Keywords and phrases: regression analysis, fuzzy set, fuzzy regression, function estimation, forecasting.

fuzzy environment. Indeed, unlike statistical regression modeling based on probability theory, fuzzy regression is based on possibility theory and fuzzy set theory [1].

A fuzzy linear regression takes the general form [4]:

$$Y = \tilde{A}_0 + \tilde{A}_1 x_1 + \cdots + \tilde{A}_n x_n, \tag{1}$$

where Y is the fuzzy output, $\tilde{A}_i$, $i = 0, 1, 2, ..., n$, is a fuzzy coefficient, and $x = (x_1, ..., x_n)$ is an n-dimensional non-fuzzy input vector [4, 14].

If equation (1) is indexed with time it becomes a fuzzy time series model. The concept of 'fuzzy' variable was first proposed by Zadeh [12]. He proposed that fuzzy set can be applied to represent data which has the characteristic of vagueness. A fuzzy linear regression model was first introduced by Tanaka et al. [11]. They formulated a linear regression model with fuzzy response data, crisp predictor data and fuzzy parameters as a mathematical programming problem. Subsequently Dongale et al. [4] studied fuzzy linear regression (FLR) problem with crisp explanatory variables and fuzzy response variables. They formulated FLR problem as linear programming model to determine regression coefficients as fuzzy numbers, where objective function minimizes total spread of fuzzy regression coefficients subject to constraint, that the support of estimated values is needed to cover support of their associated observed values for certain pre-specified level [2, 7].

Recent works on fuzzy such as application philosophy of fuzzy regression by Campobasso et al. [3] stated that the uncertainties and its prediction normally tend to be complex phenomena. Application of fuzzy using Monte Carlo has been established [6]. The regression algorithm based on fuzzy rough set, consisting of fuzzy partition, fuzzy approximation and estimation of regression value had been established [10]. In order to analyze the freight volume prediction problem of city under the influence of various factors, triangular fuzzy linear model and trapezoidal fuzzy linear regression model were, respectively, set up [9]. ST-decomposition method was developed to compute the parameters of fuzzy linear regression based on the least square approach [15].

Also, intriguingly, Berry-Stoyzle et al. [2] presented a test procedure to explicitly examine whether an independent variable has a clear functional relationship with the dependent variable in a specific regression model, or whether their relationship is fuzzy. Fuzzy logic have been applied to ARIMA model because of using finite number of data and calculating upper and lower limits to indicate best and worst status of investigating variables of interest in policy making for stock index [5].

The main contribution of this paper is the reformation of classical linear regression model as a fuzzy regression model for prediction and forecasting by using the mode for a symmetric triangular fuzzy number.

In Section 2, we provided a background on classical statistical regression. In Section 3, we explained how to forecast using fuzzy linear regression model. Section 4 offers some practical considerations of forecasting using fuzzy linear regression models and provides a numerical example of the application.

2. Preliminary Definition of Terms

Crisp data/value: Crisp data is also called precise data. A crisp value is a single value without any ambiguity.

Symmetric triangular fuzzy number (STFN): As indicated, the salient features of the TFN are its mode, its left and right spreads, and its support. When the two spreads are equal, the TFN is known as a symmetrical TFN (STFN).

Membership function (MF): A membership function is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. The input space is sometimes referred to as universe of discourse - a fancy name for a simple concept. For example, a set of tall people, in this case, the universe of discourse is all potential heights; say from 3 feet to 9 feet and the word tall will correspond to a curve that defines the degree to which any person is tall.

Regression analysis: In statistics, regression is an approach for modeling the relationship between a scalar dependent variable Y and one or more explanatory variables denoted *X*. In regression, data are modeled using predictor functions and unknown model parameters are estimated from the data. If we use *Y* to represent the dependent variable *X* and the independent variable, this relationship could be described as the regression of *Y* on *X*, and in the simplest case, this is assumed to be a straight line. The slope of the line depends on whether the correlation is positive or negative.

Ordinary least square: In statistics, least squares are a method for estimating the unknown parameters in a regression model. This method minimizes the sum of squared vertical distances between the observed response in the data set and the responses predicted by the linear approximation.

Forecasting: This is the process of making statements about events whose actual outcome (typically) have not yet been observed. A common example might be estimation of some variables of interest at some specified future date.

Classical statistical regression

Classical statistical regression takes the form

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + e_i, \quad i = 1, 2, ..., m, \tag{2}$$

where the dependent (response) variable, Y_i, the independent (explanatory) variables, X_{ij}, and the coefficients (parameters), β_j, are crisp values, and ε_i is a crisp random error term with $E(\varepsilon_i) = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma^2$, and $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0, \forall i, j, i \neq j$.

Fuzzy linear regression

The regression analysis dealing with fuzzy data is usually called *fuzzy regression analysis*. There are two motivations for developing fuzzy regression analysis. The first motivation results from the realization that it is not often realistic to assume that a crisp function of a given form can be used to represent the relationship between the given variables. Fuzzy relationship which is even though less precise seems intuitively more realistic. The second motivation results from the fact that the nature of the data in many cases has inherent characteristics of uncertainty.

Although statistical regression has many applications, it is problematic if:

1. The data set is too small.
2. There is difficulty verifying that the error is normally distributed.
3. If there is vagueness in the relationship between the independent and dependent variables.
4. If there is ambiguity associated with the event.
5. If the linearity assumption is inappropriate.

These are the situations fuzzy regression was meant to address.

In contrast to the classical statistical linear regression, fuzzy regression takes the general form in equation (1).

The major contrast here is in dealing with errors as fuzzy variables in fuzzy regression modeling, and in dealing with errors as random residuals in classical statistical linear regression models.

3. Forecasting Using Fuzzy Linear Regression Models

One of the objectives of model building is to provide forecast of the future values. After research and conceptualization of fuzzy regression models, we verify the assumptions of the model and we apply the model using a data set and use it in predicting future possible values.

This was done using the Tanaka's approach of the possibilistic regression model. Here we calculate the regression equation for the upper and lower bounds denoted Y^U and Y^L, respectively. The method assumes that the fuzzy numbers must be symmetric triangular fuzzy numbers, whereby the support of the membership function has equal spreads. After verifying that this assumption holds, it implies that the modes of the membership function fall midway between the boundary lines. $Y_i^{h=1}$ is the mode of the

MF and if a SFTN is assumed, $Y_i^{h=1} = (Y_i^U + Y_i^L)/2.$ Given the parameters, $(Y^U, Y^L, Y^{h=1})$ which characterize the fuzzy regression model, the *i*th data pair (x_i, y_i), is associated with the model parameters $(Y_i^U, Y_i^L, Y_i^{h=1})$.

Given two triangular fuzzy $\tilde{X}$ and $\tilde{Y}$ their sum is closed under addition

$$\tilde{Z} = \tilde{X} + \tilde{Y} = (x_L + y_L, x_n + y_n + x_R + y_R)$$

which represents a triangular fuzzy number.

The average of $\tilde{X}$ and the square distance between $\bar{X}$ and $\bar{Y}$ can also be computed as

$$\bar{X} = \sum \frac{\tilde{X}}{n} = \left(\frac{2x\frac{R}{i}}{n}, \sum \frac{x\frac{L}{i}}{n}, \sum \frac{x\frac{h}{i}}{n} \right),$$

$$d(\tilde{X}, \tilde{Y})^2 = d\{(x^R, x^L, x^h), (y^U, y^L, y^h)\}$$

$$= (x^h - y^h)^2 + (x^L - y^L)^2 + (x^R - y^R)^2, \text{ respectively.}$$

We can define the total sum of squares as

$$\text{Total sum of square} = \sum d(\tilde{Y}_h, \tilde{Y})^2$$

$$= \text{Reg SS} + \text{Res SS} + nd\,(\bar{Y} * \bar{Y})^2 + \eta,$$

where

$$\eta = -2\sum [\bar{y} \cdot (y^h - y_n^*) + \bar{y}(y_{Lh} - y_{Lh}^*) + \bar{y}_R(y_{Rh} - y_{Ri}^*)]. \qquad (3)$$

We use this model to predict the financial leverage of the company in some subsequent years.

4. Numerical Computation

This section focuses on the preliminary analysis of data on the financial leverage and returns on equity of an investment company, using the fuzzy regression model which takes the general form as in (1).

The fuzzy regression model was applied using a data set and was used to forecast. The model equation was used to forecast for four years. This was done using the possibilistic model of fitting fuzzy linear regression models.

Construction of the upper and lower bounds

Using the time series data set from 2003 to 2010 we filled a straight line through two or more data points in such a way that it bounds the data points from above. Here, these points are determined heuristically and ordinary least square is used to compute the parameters of the line labeled y^h, which takes the values $y = 39 + 0.75x$, as shown in Figure 4.1 below:

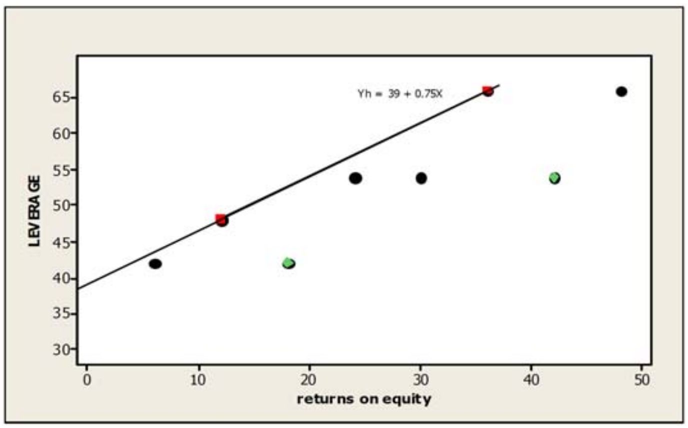

Figure 4.1. Construction of the upper bound.

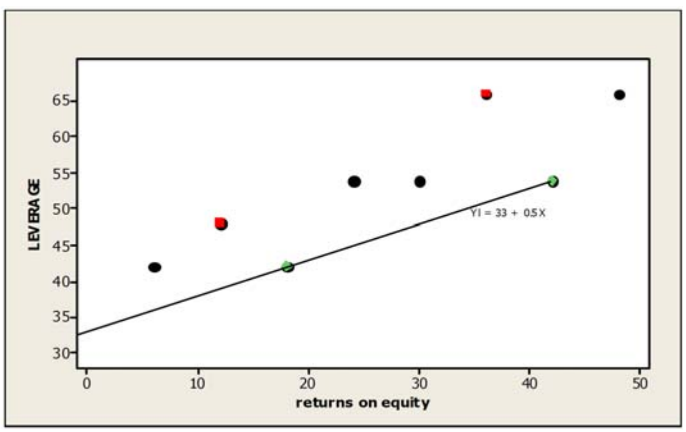

Figure 4.2. Construction of the lower bound.

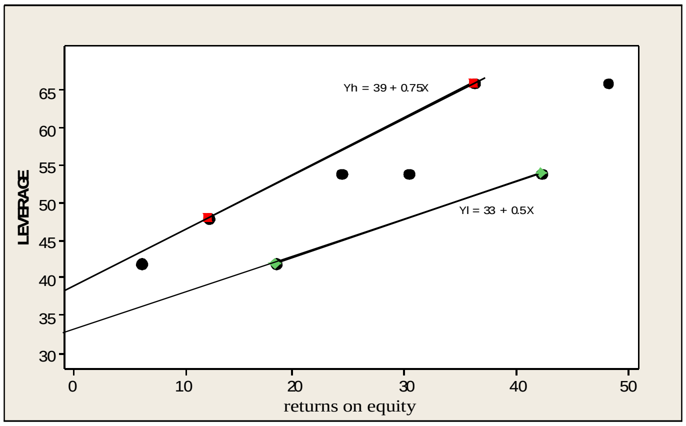

Figure 4.3. Construction of the upper and lower bounds.

Similarly, we fit a second straight line through two or more data points in such a way that it bounds the data points from below. As shown in Figure 4.2, the fitted line in this case is labeled Y^L and takes the values $Y = 33 + 0.5x$.

Estimating the model

Assuming, for the purpose of this example, that STFN are used for the MFs, the modes of the MFs fall midway between the boundary lines as indicated in Figure 4.4.

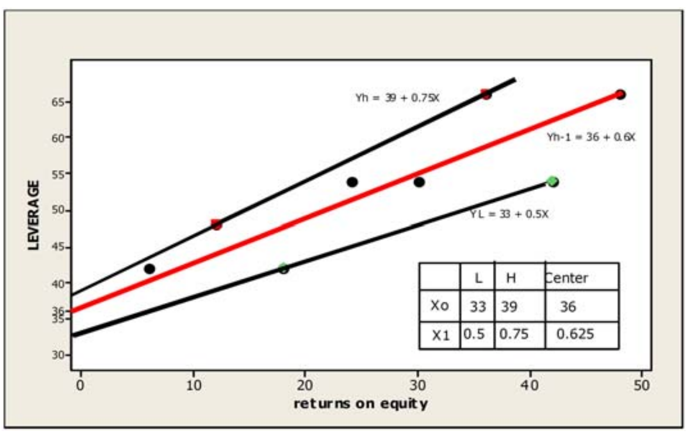

Figure 4.4. Construction of fuzzy regression mode.

The fuzzy regression intervals

For any given data pair, (X_i, Y_i), the foregoing conceptualizations can be evaluated and checked by the fuzzy regression interval (Y_i^L, Y_i^U) shown in Figure 4.5.

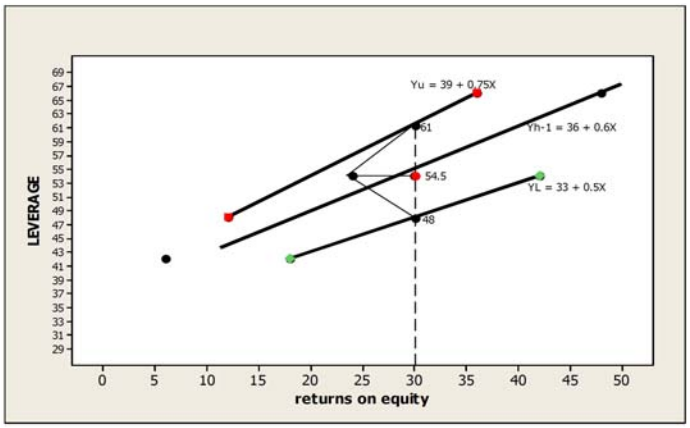

Figure 4.5. Fuzzy regression interval.

$Y_i^{h=1}$ is the mode of the MF and if a SFTN is assumed, $Y_i^{h=1} = (Y_i^U + Y_i^L)/2$. Given the parameters, $(Y^U, Y^L, Y^{h=1})$ which characterize the fuzzy regression model, the *i*th data pair (x_i, y_i), is associated with the model parameters $(Y_i^U, Y_i^L, Y_i^{h=1})$.

In possibilistic regression based on STFN, the data points involved in determining the upper and lower bounds determine the structure of the model, as depicted in Figure 4.3.

Verification if the assumption of STFN's holds

It can be shown that the fuzzy coefficients are triangular fuzzy numbers (TFNs), since the left spread here is equal to the right spread. Their membership functions (MFs), $\mu_A(a)$, can be represented as shown in Figure 4.6 below.

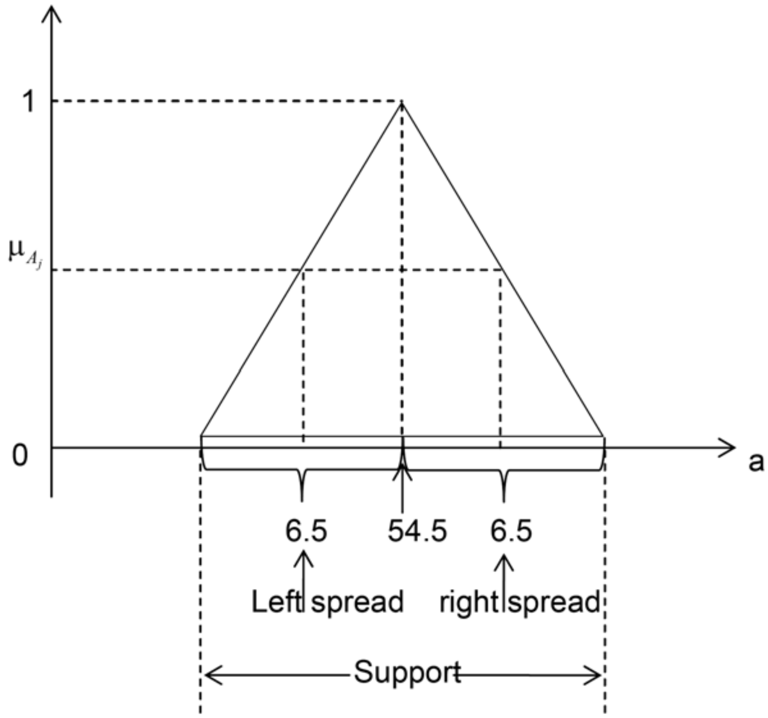

Figure 4.6. Fuzzy coefficient.

$$\mu_{Aj}(a) = \max\left\{1 - \frac{|57.75 - 54.5|}{6.5},\ 0\right\}$$

$$= \max\{0.5,\ 0\},$$

where 54.5 is the mode and 6.5 is the spread, and represented as shown in Figure 4.7 and Figure 4.8.

Membership function for triangular fuzzy number

As indicated, the salient features of the TFN are its mode, its left and right spreads, and its support. Since the two spreads are equal, the TFN is known as a symmetrical TFN (STFN).

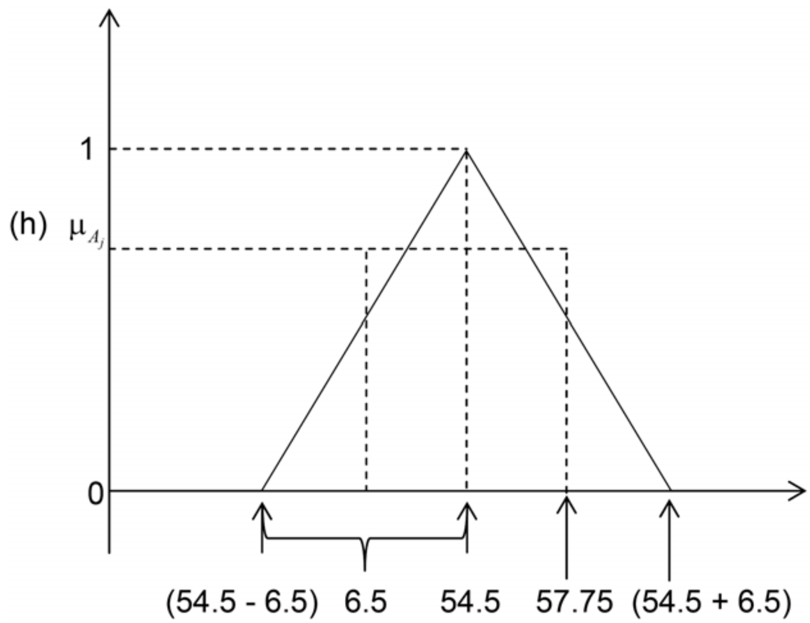

Figure 4.7. Symmetrical fuzzy parameters.

Defining the bounded intervals

$$\tilde{A}_j = \{54.5,\ 6.5\}_L = \{\tilde{A}_j : 54.5 - 6.5 \le \tilde{A}_j \le 54.5 + 6.5\}_L,$$

$$j = 0, 1, 2, \ldots 8.$$

The "*h*-certain" factor

If, as in Figure 4.5, the supports are just sufficient to include all the data points of the sample, there would be only limited confidence in out-of-sample projection using the estimated fuzzy regression model. This is resolved for fuzzy regression model, just as it is with statistical regression, by extending the supports. Consider the membership function associated with the *j*th fuzzy coefficient, the presentation of which is shown in Figure 4.7.

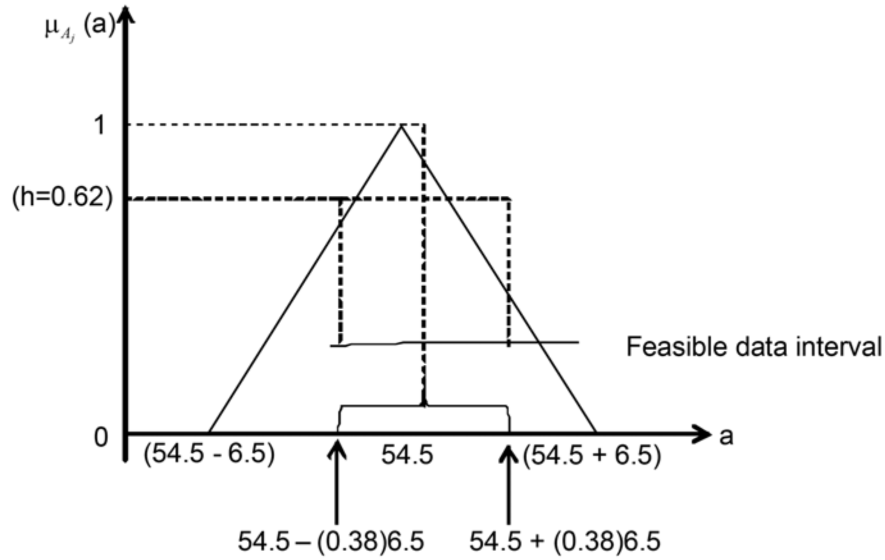

Figure 4.8. Estimating A_j using the "*h*-certain" factor.

For illustrative purposes, a non-symmetric TFN is shown where c_j^L and c_j^R represent the left and right spread, respectively. Beyond that, what makes this MF materially different from the one shown in Figure 4.6, is that it contains a point "*h*" on the *y*-axis, called an "*h-certain factor*," which, by controlling the size of the feasible data interval (the base of the shaded area), extends the support of the MF. In particular, as the *h*-factor increases for a given dataset, so increases the spreads, c_j^L and c_j^R.

Forecasting using the fuzzy regression model

Having verified the assumptions of the model as seen above, we can proceed to predict future financial leverage values for Stratus Nig. Ltd.

Using the fuzzy regression model $Y_m = 36 + 0.625x$ to predict financial leverage values, assuming we have return in equity as 52, 58, 64, 70, 76, 82, 88 and 94 in the next eight years as seen below.

Predicted leverage values from 2011 to 2018 using fuzzy linear regression model

Year	Predicted Leverage	Return in Equity
2011	68.5	52
2012	72.3	58
2013	76.0	64
2014	79.8	70
2015	83.5	76
2016	87.3	82
2017	91.0	88
2018	94.8	94

The fuzzy regression model obtained able to predict the values of the leverage, given values of return in equity. These predicted values were very close to the original values given. This is because the data satisfies most the assumptions of the fuzzy linear regression model since the data set is too small, and the fuzzy numbers have been shown to be triangular fuzzy numbers.

5. Conclusion

This work has shown beyond reasonable doubt that forecasting using the fuzzy linear regression model gives reasonably accurate results when compared with the existing classical regression model which is problematic due to the condition stated above. However, one of the greatest problems as seen from this work is the inability to know the precise equation to use when working under classical linear regression.

Acknowledgement

The authors thank the anonymous referees for their valuable suggestions which led to the improvement of the manuscript.

References

[1] E. Abdelrafe and H. Burairah, Comparison of stepwise and fuzzy multiple regression analysis techniques for managing software project risks: analysis phase, J. Comput. Sci. 10(9) (2014), 1725-1742.

[2] T. R. Berry-Stoyzle, M. C. Koissi and A. F. Shapiro, Detecting fuzzy relationships in regression models: the case of insurer solvency surveillance in Germany, Working Paper, 2009.

[3] F. Campobasso, A. Fanizzi and M. Tarantini, Some results in a multivariate generalization of the fuzzy least square regression, Proceeding of the International Conference on Fuzzy Computation, Madeira, 2009, pp. 75-78.

[4] T. D. Dongale, S. R. Ghatage and R. R. Mudholkar, Application philosophy of fuzzy regression, Int. J. Soft Comput. Engineer. (IJSCE) 2(6) (2013), 170-172.

[5] V. Marmer, D. Feir and T. Lemieux, Supplement to “Weak identification in fuzzy regression discontinuity design”, UBC Working paper, 2014.

[6] E. Gbodratollah, S. A. Ehrahim and F. Fatemeh, Presenting a fuzzy ARIMA model for forecasting stock market, J. Basic Applied Scientific Research 3(6) (2013), 37-43.

[7] U. M. Pirzada and S. Jaita, Linear fuzzy regression model, AMO - Advanced Modeling and Optimization 17(1) (2015), 67-71.

[8] R. B. Rashmi and G. Barnali, An efficient regression based demand forecasting model including temperature with fuzzy ideology for Assam, Inter. J. Adv. Res. Electrical, Electronics and Instrumentation Engineering 4(1) (2015), 331-338.

[9] S. Rui and S. Yan, Application of fuzzy linear regression model in predicting the freight volume of city, J. Inform. Comput. Sci. 12(1) (2015), 191-200.

[10] A. Shuang, S. Hong, H. Qinghua, L. Xiaoqi and D. Jianwu, Fuzzy rough regression with application to wind speed prediction, Inform. Sci. 282 (2014), 388-400.

[11] H. Tanaka, S. Uejima and K. Asai, Linear regression analysis with fuzzy model, IEEE Trans. Systems, Man and Cybernetics 12 (1982), 903-907.

[12] L. A. Zadeh, Fuzzy sets, Information and Control 8(3) (1965), 338-353.

[13] L. A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, Inform. Sci. 8 (1975), 190-249.

[14] L. A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1(1) (1978), 3-28.

[15] H. Zareamoghaddam and Z. Zareamoghaddam, A new algorithm for fuzzy linear regression with crisp inputs and fuzzy output, Int. J. Nonlinear Sci. 17(2) (2014), 128-134.