International Journal of Studies in Business Management, Economics and Strategies
Volume 03, Issue 01, January, 2024
ISSN (E): 2949-883X
Scholarsdigest.org

USING THE NSGA2 ALGORITHM TO HYBRIDIZE FUZZY REGRESSION WITH MULTI-OBJECTIVE PROGRAMMING APPROACHES WITH APPLICATION

Fatima Othman Eatiah Al-Abadi
Master's Student at College of Administration and Economics, Basrah University, Iraq
Email: fothman996@gmail.com

Dr. Sahera Hussein Zain Al-Thalabi
Administration and Economics College, Basrah University, Iraq
Email: sahera.zain@uobasrah.edu.iq

Abstract

Fuzzy regression analysis was used to model the relationship between the response variable and explanatory variables in an ambiguous environment. A multi-objective optimization approach was adopted using NSGA-II algorithms with the aim of minimizing two objectives: the prediction error is the mean squared error between the predicted and actual concentrations, which indicates the accuracy of the model and the other objective is fuzziness. In the model that represents the uncertainty in the model predictions. The lower these values are, the better the model’s performance in terms of accuracy and reliability.

Keywords: Fuzzy regression, multi-objective programming, Tanaka model, fuzzy least squares method, fuzzy moments method, fuzzy Bayesian method, NSGA2 hybridization algorithm.

Introduction

The researcher faces several problems when estimating statistical regression functions, for example when the model is indeterminate, or the relationship between the model parameters is ambiguous, or the data is hierarchical, or our problem does not meet the assumptions of probit regression (i.e., the coefficients of the regression relationship must be constant), or when the data Available data are small, so fuzzy regression is used instead of traditional regression to more robustly deal with uncertainty and imprecision that may be present in environmental data.[2]

As a result of the presence of high levels of salinity in the waters of the Shatt al-Arab, several solutions must be found to reach the optimum to reduce the levels of salinity as much as possible at the lowest costs. Given that the levels of salinity are not fixed, but rather are fuzzy or random and fuzzy together, it would have been better to study this problem using fuzzy regression with programming. Multiple targets to increase efficiency and improve forecast accuracy.

We will discuss some of what has been studied in previous years on the topic of fuzzy regression and its hybridization with multi-objective programming. Among these studies are:

1. The study (Sakawa.M, 1989) dealt with fuzzy linear regression models, where the input and output data were fuzzy numbers (ambiguous), and the study also used three indicators to equate the fuzzy numbers, but due to the conflict of ambiguous factors, three types of multi-objective programming problems were formulated to obtain fuzzy linear regression models corresponding to the three indicators, and then an interactive algorithm was developed to derive the optimal Pareto solution.[13]
2. The study (Sahed and Makidish, 2014) aimed to predict oil prices in the State of Algeria, the study compared the method of artificial neural networks and the method of using programming by objectives in the analysis of ambiguous regression (foggy), and to compare between the two methods, the standard of average absolute values of the error rate was relied upon, and the results reached that the method of using programming by objectives in the analysis of vague regression is superior to the method of artificial neural networks.[12]
3. The study (Jianga. H, et al,2019) Fuzzy regression analysis to model the relationship between explanatory variables and response variable in a fuzzy environment and by suggesting different methods for performing regression analysis and using multi-objective programming approaches in order to improve the prediction accuracy of fuzzy regression models, and the aim of the study was to reduce the blurring of fuzzy outputs, reduce the impact of outliers and reduce the average absolute percentage error of modeling. Multi-objective based on multi-objective programming, ambiguous least squares regression and fuzzy regression methods probability in terms of training errors and prediction accuracy.[5]

Fuzzy regression [6]

The analysis of regression models is one of the statistical methods that are frequently used in science, and probability theory is used to address the ambiguity (uncertainty) present in these models, which is attributed to randomness. However, on the applied side, there are regression models in which the ambiguity (uncertainty) accompanying the model is caused by blurring. Not randomness or both (randomness and blurring).

The formula for the general linear model of fuzzy linear regression is as follows:[8]

$$Y_i = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{x}_1 + \cdots + \tilde{\beta}_n \tilde{x}_n + \varepsilon_i$$

Since:

- ❖ $\tilde{y}_i$: It represents the fuzzy response variable
- ❖ $\tilde{\beta}_0 , \ldots , \tilde{\beta}_n$: Represents the fuzzy regression parameters.
- ❖ $\tilde{x}_1, \ldots , \tilde{x}_n$: Represents fuzzy explanatory variables.
- ❖ ε_i : It represents random error, which is normally distributed.

Fuzzy linear regression aims to model an inaccurate or ambiguous phenomenon using the parameters of the fuzzy model. If the assumptions of the ordinary least squares method are met, the fuzzy regression is more effective and more flexible for different problems as an alternative to classical regression.[7]

Fuzzy regression model estimation methods:

Fuzzy regression is based on combining model elements with fuzzy numbers. Fuzzy linear regression models are used to find out the ideal linear relationship between explanatory variables on the response variable in a fuzzy environment.

There are several ways to estimate the fuzzy regression model, as follows

1. Tanaka model [3]

It is the first fuzzy regression method presented by the scientist (Tanaka et al., 1982).

In this method, a fuzzy linear function is applied to determine the regression of an ambiguous phenomenon. In traditional regression, the deviations in general between the observed values and the estimated values arise from errors in the observations. However, in this method the deviations depend on not specifying the structure of the system, so these deviations are considered fuzzy for the system parameters. The relationship The fuzzy line is as follows:

$$Y = A_0 + A_1X_1 + \cdots + A_nX_n \quad \ldots \quad (1)$$
$$= \acute{A}X$$

$X = [1, X_1, X_2, \ldots, X_n]$ A vector of explanatory variables

$A = [0, A_1, A_2, \ldots, A_n]$ Vector of fuzzy model parameters

Fuzzy regression proposed by Tanaka uses fuzzy numbers. Using fuzzy numbers improves the modeling of problems where the output variable (both scalar and continuous) is affected by imprecision.

2. Fuzzy Least Squares Linear Regression (FLSLR) [14]

A statistical technique presented by the scientist (Savic and Pedrycz, 1991) as another version of the fuzzy linear regression method, and its use was confirmed (Woodall & Redden, 1994) in a later study for ease of calculation. It is implemented to reach the solution of the unknown (fuzzy) parameters in the regression model and is considered the most common method. . The fuzzy least squares method treats the regression parameters as fuzzy numbers. An estimate of the parameters of the fuzzy regression model can be obtained by the fuzzy least squares method as follows:

$$(B_i^k)^T = (X^TX)^{-1}\, X^TAY$$
$$(a_i^k)^T = (X^TX)^{-1}X^TaY$$

3. Fuzzy Moments Method [10]

This method depends on finding k parameters, then equating the population moments with the corresponding sample moments, where we obtain k equations with the number of parameters, and then the required estimates are obtained by solving the resulting equations.

4. Fuzzy Bayes Method [11]

This method is of great importance in inferential statistics and in recent decades Bayes' theorem has been a valid and powerful alternative to traditional statistical perspectives.

In the Bayesian method, the parameters are assumed to be random variables and have a prior distribution that is determined from previous information and experience of similar studies.

After that, the prior distributions are combined with the likelihood function to obtain the joint posterior density function, which becomes the basic solution in the inference for Bayesian estimation.

Multi-objective programming [1]

Statistical methods seek to find the best ways to achieve goals through research and studies that complement each other. These methods always lead to achieving the optimal (best) goal, and with the continuous development of departments, they have achieved many goals at the same time, and these goals often have multiple dimensions. And diverse, as it made linear programming lead to weak solutions that do not meet the management's ambitions.

The multi-objective programming method aims to find the best and closest solutions through a number of previously defined goals to reduce the total deviations from the specified goals to the minimum possible. The pre-determined goals are used as constraints that must be achieved as closely as possible, so through them the best solutions are obtained due to their ability to contain more than one goal so that they are written in a linear form by improving the solution based on one of the variables at each stage of the solution.

The basic idea of programming goals is to set a priority for each goal, then assign a specific weight to each goal within the priority level, then search for a solution that reduces the sum of the deviations of the goal functions from their respective goals.

The best hypothesis when using multi-objective programming is to completely achieve all objectives, but the percentage of achieving objectives varies, some of which are achieved completely (completely), and conflicting objectives are achieved at different levels depending on the priority and weight of the objective that has been allocated by the decision maker or management.

Multi-objective programming seeks to reduce the deviations between the achieved and targeted goals to a minimum that may reach zero, instead of directly reducing or maximizing a single objective function as in single-objective linear programming.

The mathematical model is as follows:[4]

$$\text{Min Z} = \sum_{v=1}^{k} \left| \sum_{j=1}^{n} C_{vj} X_j - g_v \right|$$

$$\begin{Bmatrix} \sum_{j=1}^{n} C_{vj} X_j - \delta_v^+ + \delta_v^- = g_v & , v = 1,2,\dots,k \\ \sum_{j=1}^{n} a_{ij} X_j \le = \ge b_i & , i = 1,2,\dots,m \\ X_j, \delta_v^+, \delta_v^- \ge 0 \qquad (\ j = 1,2,\dots,n) & , (v = 1,2,\dots,k) \end{Bmatrix}$$

whereas:

❖ X_j: represents the decision variable (j).

- C_{vj}: The coefficient represents the decision variable X_j In the objective function v where$(v = 1,2, \dots, k)$
- a_{ij} : Represents the resource parameter i Available for the resolution variant j where $(j = 1,2, \dots, n)$.
- g_v : Represents the ambition level of the objective function v.
- b_i :Type available resources i where $i = 1,2, \dots, m$.
- $\delta_v^+ \& \delta_v^-$ Represents positive and negative deviations from the ambition level.

Fuzzy regression and its hybridization with multi-objective programming approaches:
The fuzzy approach is used to model uncertain parameters. Moreover, it has been repeatedly emphasized in the literature that the fuzzy approach has had a significant impact in modeling preference and the multi-objective problem and has helped in approximating preference techniques.[9]

Formulation of a multi-objective programming model in case of fuzzy data:
The distinctive circumstances that the researcher faces are cases of uncertainty and inaccuracy regarding some information and data related to the issues, where the researcher is unable to accurately determine the level of ambition for a specific goal.
In 1978, the scientist Zimmerman gave the first formulation of multi-objective linear programming under conditions characterized by certainty and ambiguity, as follows:

$$\text{Fuzzy GP} \begin{cases} Min\ Z = \sum_{v=1}^{k} (\delta_v^+ + \delta_v^-) \\ sc \begin{cases} \sum_{j=1}^{n} C_{vj} X_j - \delta_v^+ + \delta_v^- \approx g_v & ,(v = 1,2, \dots k) \\ \sum_{j=1}^{n} a_{ij} X_j \approx b_i & ,(i = 1,2, \dots k) \\ X_j, \delta_v^+, \delta_v^- \geq 0 \quad (j = 1,2, \dots n), (v = 1,2, \dots k) \end{cases} \end{cases}$$

Applied side:
The fuzzy regression model was developed using multi-objective programming techniques. The goal of this model is to reduce both prediction error and ambiguity at the same time. Ambiguity here expresses the degree of uncertainty in the effect of the independent variables on the response variable. After the optimization process, several possible solutions were obtained within what is known as the “Pareto front”. One solution from this front was selected based on a specific criterion and used in prediction on the test set.

How to use the NSGA2 Hybridization Algorithm:[3]
It is an evolutionary genetic algorithm to perform dominance-based multi-objective optimization, and it directly applies the Pareto dominance relationship and elitism strategy to maintain the best results along the optimization process. Optimization refers to finding the best possible solution to a problem under a set of constraints, where two solutions are chosen from the group and then the best solution is chosen, as both solutions may be possible or not possible, or one of them is possible and the other is not.

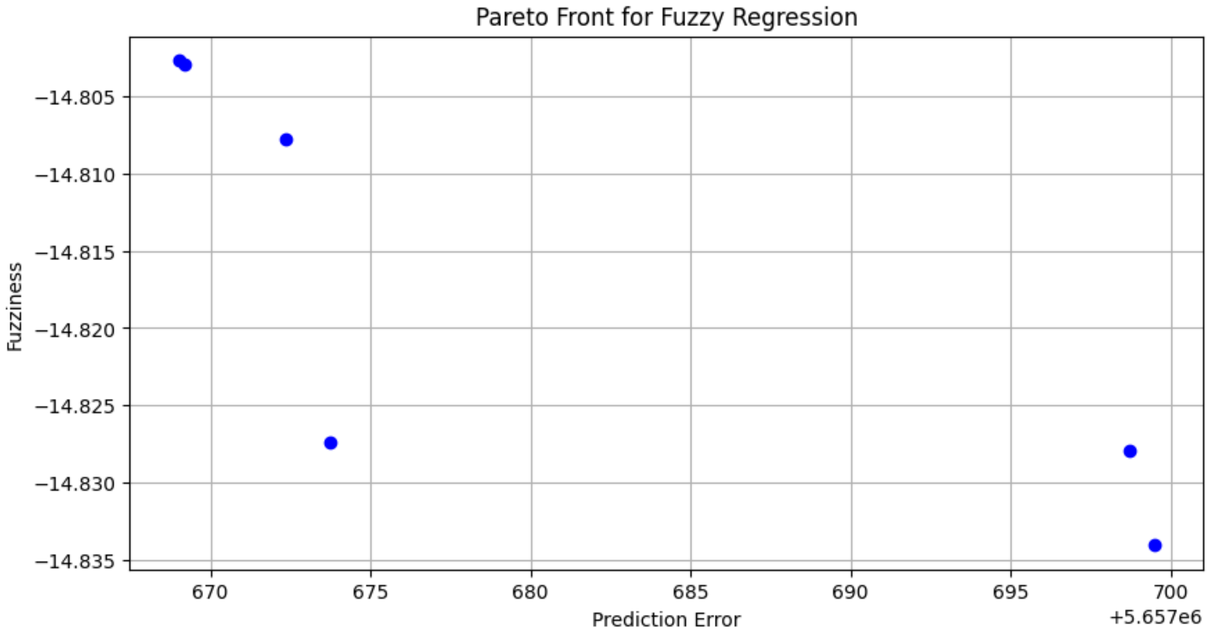
Adopting a multi-objective optimization approach using NSGA-II algorithms with the aim of reducing two objectives: prediction error and blurring in the model. Prediction error is the mean squared error between predicted and actual concentrations, indicating model accuracy, while blurring represents uncertainty in model predictions. The lower these values are, the better the model performs in terms of accuracy and reliability.

In this experiment, a fuzzy regression model was developed to predict the concentration of a specific substance, using features related to weather conditions such as precipitation (PRCP), minimum temperature (TMIN), average temperature (TAVG), and maximum temperature (TMAX). Fuzzy regression differs from traditional regression by using fuzzy set theory to more strongly deal with uncertainty and imprecision that may exist in environmental data.

After dividing the data into training and test groups, the problem is created with the initialization of the chosen algorithm NSGA2 and the sampling is equal to 100 samples and repeating the process 100 times also to extract the best approximation measures for the Pareto front solutions, where both solutions can be possible or one is possible and the other is not or both are not possible

Table (1) Algorithm outputs type NSGA2

n_gen	n_eval	n_nds	eps	indicator
1	100	4	-	-
2	200	5	18804051240.	ideal



The Pareto front is represented by the NSGA2 method

The goal of using the NSGA2 type genetic screening algorithm is to minimize the result (i.e. improve prediction error and ambiguity)

Table (2) Pareto Optimal Pareto Solutions by NSGA2

5.657	-1.480
5.657	-1.482

Table (3) Criteria and coefficient of determination for the estimator model using the NSGA2 method

MSE	R-squared	MAE	MEDAE
665,613	0.47	617	506

After evaluating the optimization results, a set of parameters was selected for the fuzzy regression model based on the lowest prediction error. This model was then subjected to a local search optimization using a stepwise descent technique to refine the parameters further, with the aim of achieving the lowest possible prediction error while maintaining an acceptable level of ambiguity.

Conclusions:

1- The combination of the fuzzy model and multi-objective programming helps improve the forecasting process by eliminating ambiguity and reducing forecast error.

2- The NSGA2 method is an effective tool for the hybridization of fuzzy regression and multi-objective programming methods, as it provides the lowest values for the statistical criteria used.

3- The mean square error criterion measures the expected and actual concentrations, which indicates the accuracy of the model. The lower these values are, the better the model's performance in terms of accuracy and reliability.

4- This means that about 47% of the variance in the response variable can be explained by the model. It also indicates that on average the model predictions are about 815 units away from the actual values because the square root standard of the MSE will give us the root mean square error, which can be interpreted on the same scale as the original values.

5- The MAE and MEDAE are also similar across this algorithm, again indicating a mean error of these sizes in the model predictions.

References

1. Abd Ali, A. T., 2019, "Using the Bayesian method and objective programming in estimating regression parameters," Master's thesis in statistics, University of Karbala.
2. Bisserier. A, Galichet.S, Boukezzoula. R, 2008, "Fuzzy Piecewise Linear Regression" , HongKong, China. pp. 2089-2094.
3. Deb, K.& Member, A., 2002, " A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II", IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, VOL. 6, NO. 2
4. Ibn al-Din, Ahmed, (2019) "Using non-replaceable multi-objective linear programming models in production decision-making, an applied study on the Ibache Foundation in Adrar," Journal of Economics and Business Administration, Volume 1, Issue 4.

5. Jianga.H, , Kwong. C.K., , Chan C.Y., Yung K.L.,2019," A multi-objective evolutionary approach for fuzzy regression analysis", Expert Systems With Applications 130 ,225–235
6. Mohammed,J.M., Abbas. M.S,2018, "Estimation Nonparametric Fuzzy Regression model using simulation", Economics & Administration of Journal The ,41,Volume115.
7. Mosleh. M, Otadi. M, Abbasbandy. S, 2011,". Fuzzy polynomial regression with fuzzy neural networks", Applied Mathematical Modelling, Volume 35, Issue 11, Pages 5400-5412.
8. Muhammad, Muhammad Jassim, (2007), "Powerful Estimations for Fuzzy Regression," a dissertation for the degree of Doctor of Philosophy in Statistics, College of Administration and Economics, University of Baghdad.
9. Najah, Pasha, 2014, "Modeling and rationalizing the administrative decision using goal programming, a case study of the Cable Industry Corporation - General Cable Branch, Biskra -", Master's thesis in Management Sciences, Mohamed Kheidar University - Biskra, Faculty of Economic and Commercial Sciences and Management Sciences, Department of Management Sciences.
10. Nasrallah, Mahdi Wahab Nimah, Ali, Bashar Khaled, 2019, "A comparison between the maximum potential method and the method of moments in estimating the fuzzy reliability of the Fregate distribution using simulation," Karbala University Scientific Journal, Volume Seventeen, Issue One.
11. Saeed, Shorouk Abdel Reda, Hussein Benin Ahmed, 2022, "Comparison between the Bayesian method and the method of moments in estimating the fuzzy reliability function for the inverse gamma distribution using simulation," Warith Scientific Journal, Volume 4, Issue 11.
12. Sahid, Abdelkader, Mukaidish, Mohamed, 2014, "A comparative study between fuzzy regression using objective programming and artificial neural networks to predict oil prices," Abi Bakr Belkaid University of Tlemcen - Algeria, Issue 14
13. Sakawa.M,1989, " Multiobjective fuzzy linear regression analysis for fuzzy input-output data", Fuzzy Sets and Systems 47 , 173-181 North-Holland
14. Yoon.J.H., Choi.S.H., 2013, "Fuzzy Least Squares Estimation with New Fuzzy Operations", Springer-Verlag Berlin Heidelberg 55, pp. 193–202.