Gazi University Journal of Science
GU J Sci
28(1):103-113 (2015)

Parameter Estimation by Fuzzy Adaptive Networks and Comparison with Robust Regression Methods

Türkan ERBAY DALKILIÇ[1], Kamile ŞANLI KULA[2,♠]

[1]*Karadeniz Technical University, Statistics and Computer Sciences Department, Trabzon*

[2]*Ahi Evran University, Mathematics Department, Kırşehir*

Received: 27/02/2014 Accepted: 27/12/2014

ABSTRACT

Fuzzy adaptive networks used for estimating the unknown parameters of a regression model are based on fuzzy if-then rules and a fuzzy inference system. In regression analysis, data analysis is very important, because, every observation may have a large influence on the parameters estimates in the regression model. When a data set has outliers, robust methods such as the M method (Huber, Hampel, Andrews and Tukey), Least Median of Squares (LMS) and Reweighed Least Squares Based on the LMS (RLS) are used for estimating parameters. In this study, a method and an algorithm have been suggested to define the parameters of a switching regression model. Adaptive networks have been used in constructing one model that has been formed by gathering obtained models. There are methods that suggest the class numbers of independent variables heuristically. Alternatively, to define the optimal class number of independent variables, we aimed to use the suggested validity criterion. The proposed method has the properties of a robust method, because the process does not give permission to the intuitional and is not affected by the outliers, which exist in the independent variable. Consequently, another aim of this study is, to compare the proposed method with the robust methods that are mentioned above. For the comparison the cross-validation method is used.

Keywords: Fuzzy adaptive network, robust regression, switching regression.

1. INTRODUCTION

In regression analysis, a data set can be formed by collecting observations that have been obtained from more than one class. The regression model, that will be formed when each class is defined by a function, f_i and the number of classes is given by c, can be called a switching regression model, and it is expressed by $Y_i = f_i(X) + \varepsilon_i \ \left(1 \le i \le c\right)$[1, 2, 3].

There are two significant steps in the process of prediction for switching regression models. Thr first step is to determine an a priori parameter set characterizing the classes that the data come from and to update these parameters within the process. The other step is to determine a posteriori parameters belonging to the regression models that are formed. To obtain prediction values for data coming from different classes, the adaptive network based on a fuzzy inference system will be used. However, it is necessary to determine fuzzy class numbers related to independent variables first. Since the independent variables are fuzzy, we consider the proposed validity criterion for determining the optimal class number. The validity criterion has been designed to find optimal numbers of class and to determine compacted and separated clusters. There are many validity criteria for clusters in the literature. In this study the *S* function also called the Xie – Beni index will be used since it is easy to count and is comprehensible [4].

There are many validity criteria for clusters in the literature. In this study the *S* function also called the Xie

♠Corresponding author, e-mail: kskula@ahievran.edu.tr

– Beni index will be used since it is easy to count and is comprehensible [4].

There are different studies about fuzzy clustering and the validity criterion. In the study of Mu-Song, C. the analysis of fuzzy clustering was done for determining fuzzy memberships and in this study a method was suggested for indicating the optimal class numbers that belong to the variables [5]. Bezdek, J.C. has conducted important studies on the fuzzy clustering topic. One such study is by Hathaway R.J. and Bezdek J.C. (1993), and is on switching regression and fuzzy clustering [5, 6]. In 1991, Xie, X.L. and Beni, G. suggested a validity criterion for fuzzy clustering. In this study we used the Xie-Beni validity criterion for determining optimal class numbers [4].

There have also been various studies on the usage of the adaptive network for parameter prediction. Cheng, C.B. and Lee, E.S. used the fuzzy adaptive network approach for fuzzy regression analysis [7], and they studied both fuzzy adaptive networks and the switching regression model [8]. Jhy-Shing, R.J. studied on the adaptive networks based on the fuzzy inference system [9]. Nasrabadi, E. and Hashemi, S.M., propose a training algorithm for fuzzy neural networks with general fuzzy number weights, biases, inputs and outputs for computation of nonlinear fuzzy regression models [10].

There are many robust regression methods in the literature. In the second part of the study, we will explain several robust regression methods that are commonly used, along with the Fuzzy c-Means Algorithm. The ways in which adaptive networks provide efficient solutions when data class membership is unclear, as well as the fuzzy inference system, will be dealt with in the third part of the study. In the fourth part, we describe practices, and in the fifth part we present our conclusions.

2. ROBUST REGRESSION METHODS

In the determining of test statistics and coefficients, the role of each observation must be taken into consideration. The data details must also be tested, because the results of the parameter estimation may be related to an observation, and removal of this observation from the data may change the result of the analysis. This kind of observation, which has a bigger residual value than the others, is called an outlier. In the event of an outlier value, robust methods are used that are less affected than the LSM method during the estimation of the regression model [11]. In this section, we provide definitions of robust methods, M, LMS and RLS, which are commonly used in the literature. In addition we provide definitions of the Fuzzy *c*-Means (FCM) algorithm to the comparison.

2.1. M methods

M method is minimizing the residual function. Regression coefficients $\hat{\beta}_j$ are obtained by minimizing the sum

$$\sum_{i=1}^{n} \rho\left[\left(y_i - \sum_{j=1}^{p} x_{ij}\hat{\beta}_j\right) / d\right] \tag{1}$$

where, $x_{ij} = (x_{i1}, x_{i2}, \dots, x_{ip})$ are independent variables, y_i is depend variable and $d = median\left|r_i - median(r_i)\right| / 0.6745$, r_i is the i[th] observed error.

By taking the first partial derivative of the sum in Eq. (1) with respect to each $\hat{\beta}_j$ and setting it to zero, the regression coefficients are obtained from Eq. (2).

$$\sum_{i=1}^{n} x_{ij}\psi\left[(y_i - \sum_{j=1}^{p} x_{ij}\hat{\beta}_j) / d\right] = 0 \tag{2}$$

Huber's ρ function is defined as

$$\rho(z) = \begin{cases} \dfrac{z^2}{2} & |z| \le k \\ k|z| - \dfrac{k^2}{2} & |z| > k \end{cases} \tag{3}$$

$$z = r_i / d$$
$$d = median\left|r_i - median(r_i)\right| / 0.6745$$

where *k* is called the tuning constant and *k* is set at 1.5 and r_i is the i^{th} observed error. Sometimes the numerator of *d* is called the median of the absolute deviations (MAD). The following function is obtained by taking the derivative of Eq. (3)

$$\psi(z) = \begin{cases} -k & z < -k \\ z & |z| \le k \\ k & z > k \end{cases} \tag{4}$$

The function ψ is the derivative of ρ. They are typically set up such that large residuals will be given only marginal or zero ψ weights in Eq. (3). So ψ is often labeled as "redescending to zero".

The Hampel ψ function is defined as

$$\psi(z) = (sign z)\begin{cases} |z| & 0 \le |z| \le a \\ a & a \le |z| \le b \\ a\left(\dfrac{c - |z|}{c - b}\right) & b \le |z| \le c \\ 0 & c \le |z| \end{cases} \tag{5}$$

The constant values are selected as a=1.7, b=3.4 and c=8.5 in general.

Andrews (sin estimate) ψ function is defined as

$$\psi(z) = \begin{cases} \sin(z/k) & |z| \le k\pi \\ 0 & |z| > k\pi \end{cases} \quad (6)$$

where k is taken to be 1.5 or k=2.1.

In Tukey's biweight estimate, the ψ function is defined as

$$\psi(z) = \begin{cases} z(1-(z/k)^2)^2 & |z| \le k \\ 0 & |z| > k \end{cases} \quad (7)$$

where k is selected as 5 or 6 [12, 13, 14, 15].

2.2. Least median of squares (LMS)

Minimizing the median of the squared residual is the goal of this method, and the objective function is defined as

$$\min_{\hat{\theta}} \ \underset{i}{med} \ r_i^2 \quad (8)$$

where r_i is the i^{th} observed error and $\hat{\theta}$ is defined as estimated. In the least median of squares method, a weight for the i^{th} observation is defined as

$$w_i = \begin{cases} 1 & |r_i / s_0| \le 2.5 \\ 0 & |r_i / s_0| > 2.5 \end{cases} \quad (9)$$

where n is the total number of observations and p is the number of variables.

Then, we have $s_0 = 1,4826\left(1+\frac{5}{n-p}\sqrt{\underset{i}{med}\ r_i^2}\right)$.

The LMS method is calculated by Rousseuw and Leroy (1987) was taken with the resampling algorithm [16].

2.3. Reweighed least squares based on the LMS (RLS)

The RLS aims to minimize the weighted median of the squared residual, and the objective function is defined as

$$\min_{\hat{\theta}} \sum_i w_i r_i^2 \quad (10)$$

In the RLS method, weights are defined as

$$w_i = \begin{cases} 1 & |r_i / \hat{\sigma}| \le 2.5 \\ 0 & |r_i / \hat{\sigma}| > 2.5 \end{cases} \quad (11)$$

Where $\hat{\sigma}$ is defined as

$$\hat{\sigma} = \sqrt{\frac{\sum w_i r_i^2}{\sum w_i - p}}$$

[16].

2.4. Fuzzy *c*-Means (FCM) Algorithm for Fuzzy *c*-Regression Model

Several clustering criteria have been proposed for identifying optimal fuzzy c-partitions. Of these, the most popular and well-studied method is associated with the generalized least-squared errors functional

$$J_m = \sum_{i=1}^{c} \sum_{j=1}^{n} \left(\mu_{ij}\right)^m d^2\left(x_j, v_i\right) \quad (12)$$

with v_i indicating the cluster center and μ_{ij} defined as fuzzy membership. Here,

d : distance between each observation and cluster centers
c : class numbers
n : observation numbers
m : fuzziness index
v : fuzzy cluster center.

The clustering algorithm based on fuzzy c-means consists of the following steps:

Step 1: Initial membership of belonging i. class of x_j, μ_{ij} are defined as

$$\sum_{i=1}^{c} \mu_{ij} = 1 \quad (13)$$

Step 2: Fuzzy cluster centers, v_i's are counted as

$$v_i = \frac{\sum_{j=1}^{n} \left(\mu_{ij}\right)^m x_j}{\sum_{j=1}^{n} \left(\mu_{ij}\right)^m} \qquad i=1,2,...,c \quad (14)$$

Step 3: Fuzzy memberships defined in step l are updated by using

$$\mu_{ij} = \frac{\left(\frac{1}{d^2\left(x_j, v_i\right)}\right)^{\frac{1}{(m-1)}}}{\sum_{i=1}^{c}\left(\frac{1}{d^2\left(x_j, v_i\right)}\right)^{\frac{1}{(m-1)}}} \tag{15}$$

equality.

Step 4: Updating in step 2 and 3 are carried on until the amount of decrease in the value of J_m, defined as the objective function, will become lower than a small stable specified in advance [6, 12, 17].

3. PARAMETER ESTIMATION BY FUZZY ADAPTIVE NETWORKS

The fuzzy adaptive network used in predicting the unknown parameters of a regression model is based on fuzzy if-then rules and the fuzzy inference system. When the problem is to estimate a regression line to fuzzy inputs coming from different distributions, the Sugeno Fuzzy Inference System is appropriate and the proposed fuzzy rule in this case is

R^L = If; ($x_1 = F_1^L$ and $x_2 = F_2^L$ and $\ldots x_p = F_p^L$)

Then; $Y = Y^L = c_0^L + c_1^L x_1 + \ldots + c_p^L x_p$

Here, F_i^L stands for fuzzy cluster and Y^L stands for system output according to the R^L rule.

The weighted mean of the models obtained according to fuzzy rules is the output of the Sugeno Fuzzy Inference System and the common regression model for data coming from different classes is indicated with this weighted mean.

Neural networks enabling the use of a fuzzy inference system for fuzzy regression analysis are known as adaptive networks. They are formed via neural connections, consist of five layers and are used for obtaining a good approach to regression functions [18, 19, 20].

Neurons which are forming network are characterized with parameter functions. The processing of an adaptive network consisting of five layers is as follows. Functional relationships between dependent and independent variables in the processing of adaptive network are modeled and estimates based on these models are obtained.

Fuzzy rule number of the system depends on the number of independent variables and the class or fuzzy sets numbers forming the independent variables. Letting p be the independent variable number, and if the level number belonging to each variable is denoted by $l_i \ (i = 1,\ldots, p)$ then the fuzzy rule number is indicated with

$$L = \prod_{i=1}^{p} l_i \tag{16}$$

The neuron h in the *first layer* is defined as

$$f_{1,h} = \mu_{F_h}(x_i) \qquad h=1,2,..,L \quad i=1,2,...,p \tag{17}$$

With the fuzzy clusters related to fuzzy rules are denoted by $F_1, F_2, \ldots, F_h$. Here μ_{F_h} is the membership function that relates to F_h. Different membership functions can be defined. Here, functions of membership are defined as

$$\mu_{F_h}(x_i) = \exp\left[-\left(\frac{x_i - v_h}{\sigma_h}\right)^2\right] \tag{18}$$

since it is thought that the data comes from a normal distribution, where the parameter set of normal distribution is $\{v_h, \sigma_h\}$. The parameter set $\{v_h, \sigma_h\}$ in this layer indicates priori parameters.

Each nerve in the *second layer* has input signals coming from the first layer and they are defined as the product of these input signals.

The nerves in the *third layer* are the fixed nerves as well as the nerves in the second layer. The output of this layer is a normalization of the outputs of the second layer, the nerve function is defined as

$$f_{3,L} = \overline{w}^L = \frac{w^L}{\sum_{L=1}^{m} w^L} \tag{19}$$

where,

$$w^L = \prod_{i=1}^{p} \mu_{F_i^L}(x_i).$$

The output signals of the *fourth layer* are also connected to a function, and this function is given by

$$f_{4,L} = \overline{w}^L Y^L \tag{20}$$

where, Y^L stands for conclusion part of the fuzzy if-then rule. It is denoted by

$$Y^L = c_0^L + c_1^L x_1 + c_2^L x_2 + \ldots + c_p^L x_p \tag{21}$$

where the c_i^L are fuzzy numbers and stands for posteriori parameters.

In the *fifth layer*, there is only one nerve and it is a fixed nerve. It is counted

$$f_{5,1} = \hat{Y} = \sum_{L=1}^{m} \overline{w}^L Y^L \tag{22}$$

as the total of all signals [7, 8].

The prediction of parameters with an adaptive network is based on the principle of the minimizing of error criterion. . There are two significant steps in the process of prediction. First, we must determine the a priori parameter set characterizing the class from which the data comes and then update these parameters within the process. The second step is to determine a posteriori parameters belonging to the regression models to be formed. The process of determining parameters for the switching regression model begins with determining class numbers of independent variables and a priori parameters [21].

The algorithm related to the proposed method for determining the switching regression model in the case of independent variables coming from a normal distribution is defined as follows.

3.1. An algorithm for parameter estimation

Step 0: Optimal class numbers (c) related to the data set belonging to independent variables are determined.

Different values of the S function are obtained with

$$S_k = \frac{\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n}\left(\mu_{ij}\right)^m \left\|v_i - x_j\right\|^2}{\min_{i\neq j}\left\|v_i - v_j\right\|^2} \qquad k = 1,\ldots,c \tag{23}$$

where v_i is the fuzzy cluster center of the i^{th} class, m is the fuzziness index and μ_{ij} is fuzzy membership. For all the values of c denoting for class number (c=2, c=3, ... ,c=max) and the c used in counting the lowest of S_k values is defined as optimal class number.

Step 1: A priori parameters are determined.

Spreading is determined intuitively according to the space in which input variables gain value and to the fuzzy class numbers of the variables. Center parameters are based on the space in which variables gain value and fuzzy class numbers and it is defined by

$$v_i = \min(X_i) + \frac{\max(X_i) - \min(X_i)}{(c_i - 1)}(i-1)$$
$$i = 1,\ldots,p \tag{24}$$

Here c (c>1) stands for the optimal class number related to the variables specified in step 1, and p indicates number of independent variables.

Step 2: $\overline{w}^L$ weights are counted which are then used to form matrix B to be used in counting the a posteriori parameter set. L is the fuzzy rule number. As they are defined in the third part, the $\overline{w}^L$ weights are outputs of the nerves in the third layer of the adaptive network, and they are counted based on a membership function related to the distribution family to which independent variable belongs. Nerve functions in the first layer of the adaptive network are defined in Eq. (17). $\mu_{F_h}(x_i)$ is called the membership function. Here, when the normal distribution function which has the parameter set of $\{v_h, \sigma_h\}$ is considered, membership functions are defined as in Eq. (18). From the defined membership functions, membership degrees related to each class forming independent variables are determined. The w^L weights are indicated as

$$w^L = \mu_{F_L}(x_i).\mu_{F_L}(x_j) \tag{25}$$

They are obtained via mutual multiplication of membership degrees at an amount depending on the number of independent variables and the fuzzy class numbers of these variables. $\overline{w}^L$ weight is a normalization of the weight defined as $\overline{w}^L$ and they are counted with Eq. (19).

Step 3: In the case where independent variables are composed of fuzzy numbers, and dependent variables are composed of definite numbers, the a posteriori parameter set is obtained as definite numbers, in the form, $c_i^L = a_i^L$. In this case, to determine the a posteriori parameter set,

$$Z = (B^T B)^{-1} B^T Y \tag{26}$$

and equality is used [8].

Here B, Y and Z are defined as

$$B = \begin{bmatrix} \overline{w}_1^1, & \cdots, & \overline{w}_1^m, & \overline{w}_1^1 x_{11}, & \cdots, & \overline{w}_1^m x_{11}, & \cdots, & \overline{w}_1^1 x_{p1}, & \cdots, & \overline{w}_1^m x_{p1} \\ \vdots & & & & & & \overline{w}_k^l x_{jk} & & & \vdots \\ \overline{w}_n^1 & \cdots, & \overline{w}_n^m, & \overline{w}_n^1 x_{1n}, & \cdots, & \overline{w}_n^m x_{1n}, & \cdots, & \overline{w}_n^1 x_{pn}, & \cdots, & \overline{w}_n^m x_{pn} \end{bmatrix}$$

$$Y = \left[y_1, y_2, ..., y_n\right]^T \text{ and } Z = \left[a_0^1, ..., a_0^m, a_1^1, ..., a_1^m, a_p^1, ..., a_p^m\right]^T.$$

Step 4: By using the a posteriori parameter set $c_i^L = a_i^L$ obtained in Step 3, the switching regression models given by

$$Y^L = c_0^L + c_1^L x_1 + c_2^L x_2 + ... + c_p^L x_p \tag{27}$$

are constituted. Setting out from the models and weights specified in Step 2, prediction values are obtained with

$$\hat{Y} = \sum_{L=1}^{m} \overline{w}^L Y^L . \tag{28}$$

Step 5: When the error related to each observation is given by

$$\varepsilon_k = Y_k \{-\} \hat{Y}_k \qquad k=1,\ldots,n \tag{29}$$

the error related to model is counted as

$$\varepsilon = \sum_{k=1}^{n} \left(Y_k - \hat{Y}_k\right)^2 \tag{30}$$

If $\varepsilon < \phi$, then the a posteriori parameters have been obtained as parameters of regression models to be formed, and the process is determinate.

If $\varepsilon \geq \phi$, then, step 6 begins.

Here ϕ, is a law stable value determinated by the decision maker, {-}, is the differential operation in the case of the fuzziness for the dependent variable as well.

Step 6: Central a priori parameters specified in Step l are updated with

$$v_i^{'} = v_i \pm t \tag{31}$$

in a way that it increases from the lowest value to the highest and it decreases from the highest value to the lowest. Here, t is the size of the step;

$$t = \frac{\max(x_{ji}) - \min(x_{ji})}{a} \quad j = 1,...,n \;\; i = 1,...,p \tag{32}$$

and a is a stable value, which is determinant by the size of the step, and is therefore an iteration number.

Step 7: Predictions for each a priori parameter obtained by change and the error criteria related to these predictions are counted. The lowest of the error criterion is defined. A priori parameters giving the lowest error specified, and the prediction obtained via the models related to these parameters is taken as output. This method can also be used when the dependent variable is fuzzy.

In this algorithm, the Xie-Beni index, will be used, since it is easy to count and is comprehensible. There are methods that suggest the class numbers of independent variables heuristically, but in the proposed algorithm alternatively, the validity criterion is used to define the optimal class number of independent variables. Because of the proceeding complexity, there is not used the back propagation error methods to update the priori parameters. These parameters are update by the method which is presentation in the Step 6. All of the values for the priori parameter were used to found the optimal prediction by this method.

In the proposed algorithm, the predicted values which are obtained from the fuzzy adaptive network are not to be affected by the outliers that may exist in the independent variable. This is because in this algorithm, all of the independent variables are weighted. Consequently, the proposed method has a robust method's properties, and, it is comparable to robust methods that are commonly used in literature. In addition the prediction values are compared to other switching regression model that is used by C.B. Cheng and E.S. Lee based on Fuzzy c-means (FCM). This method is defined in Section (2.4).

4. NUMERICAL EXAMPLES

The values related to the data set having two independent variables and one dependent variable is shown in Table 1. The values in the data set have been taken from the study carried out by Cheng and Lee (1998) [8]. The switching regression model and predictions for this model are obtained via the proposed algorithm for this data set. Moreover, predictions have been obtained using the robust regression methods. Also, the predictions from the study of Cheng and Lee, which is based on FCM are used for comparison.

Table 1. Data set having two independent variables and one dependent variable.

Observed Number	X_1	X_2	Y	Observed Number	X_1	X_2	Y
1	6.859	9.688	4.992	16	6.134	2.824	4.273
2	7.215	0.617	4.849	17	8.976	1.165	5.160
3	3.199	2.898	5.256	18	2.316	7.121	5.310
4	0.260	9.151	4.994	19	0.080	2.127	5.036
5	5.528	7.114	3.275	20	2.937	1.300	4.755
6	3.539	2.766	4.837	21	5.408	1.343	6.047
7	9.476	8.443	5.042	22	6.512	6.041	5.163
8	5.968	9.446	5.276	23	3.504	5.534	5.409
9	7.562	2.678	5.384	24	9.319	1.516	5.075
10	5.103	3.324	4.379	25	6.879	2.906	5.318
11	2.802	6.101	4.208	26	6.793	0.142	5.035
12	2.831	6.492	4.887	27	8.325	2.432	4.904
13	8.287	6.081	5.167	28	0.539	8.211	5.012
14	5.479	2.999	3.382	29	1.544	6.900	4.863
15	0.423	0.885	5.033	30	9.298	2.566	4.826

From the result of the residual analysis, the 5^{th}, 14^{th}, and 21^{st}, observations are outliers. The proposed algorithm was executed with a programme written in MATLAB. From the initial step of the proposed algorithm, fuzzy class numbers for each variable are defined as three. Number of fuzzy inference rules to be formed depending on these class numbers is obtained as

$$L = \prod_{i=1}^{p=2} l_i = l_1 \times l_2 = 3 \times 3 = 9$$

Models obtained via nine fuzzy inference rules are

$$\begin{aligned}
&\hat{y}_1 = 556 + 133x_1 - 1x_2 \\
&\hat{y}_2 = -1143 - 273x_1 + 100x_2 \\
&\hat{y}_3 = 624 + 563x_1 132x_2 \\
&\hat{y}_4 = -2662 + 376x_1 - 304x_2 \\
&\hat{y}_5 = 6875 - 788x_1 - 911x_2 \\
&\hat{y}_6 = 781 + 1665x_1 - 617x_2 \\
&\hat{y}_7 = -263 + 213x_1 + 614x_2 \\
&\hat{y}_8 = -2056 - 460x_1 + 1947x_2 \\
&\hat{y}_9 = -2813.1 + 1026x_1 + 1282x_2
\end{aligned} \tag{33}$$

The error related to predictions obtained via the models given with Eq. (30) is found as

$$\varepsilon_{Network} = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 = 0.2972 .$$

Regression model estimates, which are obtained from robust regression methods and the least square method, are located in Table 2.

Table 2. The predicted value of regression parameters.

Method	Constant	Regression Parameters $\hat{\beta}_1$	Regression Parameters $\hat{\beta}_2$
LSM	4.9323	0.0039	-0.0109
Huber	4.9313	0.0086	0.0064
Hampel	4.9315	0.0100	0.0104
Tukey	4.9198	0.0095	0.0131
Andrews	4.9238	0.0109	0.0087
LMS	5.04227	0.02375	-0.00659
RLS	4.97338	0.00817	0.01139

The weights related to the observations that are used in estimation methods for regression models, are located in Table 3.

The weights for robust methods are expression of that observation's effect on one model for each of the outlier observations of the robust method. On the other hand, weight obtained from the network is an expression of that observation's effect on more than one model, which are expressed in Eq. (33). For this reason, nine different weights, which are called membership degrees of observation, are located in Table 3. Some weights of outlier observations (5th, 14th, and 21st) to switching regression are high. This is because each outlier observation is belongs to at least one fuzzy rule of the switching regression model.

The residuals, which belong to estimates for switching regression models in Eq. (33) and belong to estimates for models from robust regression methods, are located in Table 4. In addition, the residuals from the C.B Cheng and E.S. Lee model, which is based on FCM, is located in same table. The proposed algorithm was execudet with a program written in MATLAB. In the stage of step operating, data sets have two independent variables and the variables have outlier observations. The defined methods M (Huber, Hampel, Tukey, Andrews) were executed with programs written in MATLAB, and the LMS and LRS methods were executed with a Progress packet program, which was written by Rousseeuw and Leroy.

The prediction of the other switching regression model, which is based on FCM, is taken from the study of Cheng and Lee [7], and, the obtained results from all methods were compared.

The residuals of the outlier observation obtained from the proposed algorithm and from the methods which is based on FCM are similar. The residuals of from the robust methods and LSM are large, but the residuals from the proposed algorithm based network and the method based on FCM are small. This is because the last two methods are bought depent on fuzzy clustering.

As it can be seen in a numerical example, error related to predictions obtained via the network according to error criterion is lower than errors obtained via all the other methods.

Table 3. The weight related to observation for all methods.

NO	LSM	Huber	Hampel	Tukey	Andrews	RLS	The membership degrees of the observation to belong to the Models in Eq. (33)								
							w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9
1	1	1	1	0,9783	0,4733	1	0.0008	0.0257	0.0907	0.0070	0.2200	0.7753	0.0066	0.2083	0.7338
2	1	1	1	0,9673	0,4674	1	0.0691	0.0196	0.0006	0.7070	0.2007	0.0064	0.8002	0.2271	0.0072
3	1	1	1	0,8943	0,4507	1	0.5028	0.4663	0.0484	0.6837	0.6340	0.0658	0.1029	0.0954	0.0099
4	1	1	1	0,9965	0,4761	1	0.0157	0.3735	0.9960	0.0049	0.1160	0.3093	0.0002	0.0040	0.0106
5	1	0,2006	0,005	0	0	0	0.0211	0.1744	0.1615	0.0924	0.7642	0.7078	0.0448	0.3706	0.3432
6	1	1	1	0,9650	0,4684	1	0.4618	0.3999	0.0388	0.7449	0.6450	0.0625	0.1330	0.1152	0.0112
7	1	1	1	0,9906	0,4750	1	0.0003	0.0045	0.0084	0.0088	0.1449	0.2675	0.0310	0.5107	0.9430
8	1	1	1	0,9538	0,4617	1	0.0020	0.0543	0.1686	0.0107	0.2966	0.9217	0.0065	0.1794	0.5576
9	1	0,9481	1	0,8161	0,4334	1	0.0437	0.0361	0.0033	0.5318	0.4399	0.0407	0.7165	0.5928	0.0549
10	1	0,5683	0,5680	0,4851	0,3492	0	0.1976	0.2286	0.0296	0.6995	0.8092	0.1048	0.2740	0.3170	0.0411
11	1	0,4462	0,4080	0,2427	0,2820	0	0.1389	0.6791	0.3718	0.1547	0.7564	0.4141	0.0191	0.0932	0.0510
12	1	1	1	0,9685	0,4708	1	0.1064	0.6376	0.4276	0.1203	0.7205	0.4832	0.0150	0.0901	0.0604
13	1	1	1	0,9881	0,4727	1	0.0055	0.0267	0.0145	0.0967	0.4681	0.2536	0.1876	0.9079	0.4919
14	1	0,2172	0,0361	0	0	0	0.1750	0.1710	0.0187	0.7479	0.7309	0.0800	0.3539	0.3458	0.0378
15	1	1	1	0,9856	0,4729	1	0.9998	0.3261	0.0119	0.3370	0.1099	0.0040	0.0126	0.0041	0.0001
16	1	0,4813	0,4729	0,3395	0,3056	0	0.1216	0.1085	0.0108	0.7224	0.6447	0.0644	0.4750	0.4239	0.0424
17	1	1	1	0,9708	0,4704	1	0.0146	0.0055	0.0002	0.3628	0.1369	0.0058	0.9948	0.3753	0.0158
18	1	1	1	0,8888	0,4456	1	0.0768	0.6379	0.5930	0.0670	0.5566	0.5174	0.0065	0.0537	0.0500
19	1	1	1	0,9884	0,4732	1	0.9109	0.5661	0.0394	0.2584	0.1606	0.0112	0.0081	0.0050	0.0004
20	1	1	1	0,9344	0,4606	1	0.6851	0.2772	0.0126	0.8166	0.3304	0.0150	0.1077	0.0436	0.0020
21	1	0,3310	0,2435	0,0231	0,1729	0	0.2345	0.0970	0.0045	0.9675	0.4003	0.0185	0.4417	0.1828	0.0085
22	1	1	1	0,9843	0,4715	1	0.0235	0.1116	0.0592	0.1691	0.8017	0.4254	0.1345	0.6374	0.3383
23	1	0,8524	0,9419	0,7897	0,4225	1	0.1564	0.5698	0.2324	0.2479	0.9030	0.3683	0.0435	0.1584	0.0646
24	1	1	1	0,9967	0,4757	1	0.0102	0.0046	0.0002	0.3002	0.1359	0.0069	0.9779	0.4426	0.0224
25	1	1	1	0,8728	0,4465	1	0.0708	0.0659	0.0069	0.6112	0.5691	0.0593	0.5843	0.5441	0.0567
26	1	1	1	0,9964	0,4757	1	0.0921	0.0204	0.0005	0.7623	0.1691	0.0042	0.6980	0.1548	0.0038
27	1	1	1	0,9757	0,4702	1	0.0237	0.0173	0.0014	0.4234	0.3083	0.0251	0.8370	0.6094	0.0497
28	1	1	1	0,9994	0,4762	1	0.0384	0.5618	0.9195	0.0137	0.2007	0.3285	0.0005	0.0079	0.0130
29	1	1	1	0,9607	0,4696	1	0.1038	0.7685	0.6370	0.0615	0.4549	0.3771	0.0040	0.0298	0.0247
30	1	1	1	0,9306	0,4592	1	0.0090	0.0071	0.0006	0.2631	0.2053	0.0179	0.8479	0.6619	0.0578

Table 4. The residuals belong to observations for all methods.

NO	LSM Residual	Huber Residual	Hampel Residual	Tukey Residual	Andrews Residual	LMS Residual	RLS Residual	Network Residual	Cheng-Lee Residual (depend on FCM)
1	0.1381	-0.0609	-0.1090	-0.1200	-0.0910	-0.1494	-0.1477	-0.0014	-0.038
2	-0.1048	-0.1485	-0.1612	-0.1476	-0.1586	-0.3606	-0.1904	0.0477	-0.277
3	0.3427	0.2784	0.2623	0.2678	0.2721	0.1568	0.2235	0.0281	0.480
4	0.1601	0.0015	-0.0352	-0.0481	-0.0126	0.0058	-0.0857	-0.0023	0.000
5	-1.6016	-1.7498	-1.7859	-1.7906	-1.7710	-1.8517	-1.8246	0.0003	-0.153
6	-0.0791	-0.1426	-0.1587	-0.1527	-0.1494	-0.2711	-0.1968	-0.0569	0.265
7	0.1644	-0.0254	-0.0723	-0.0786	-0.0586	-0.1697	-0.1050	0.0008	0.000
8	0.4230	0.2324	0.1865	0.1757	0.2048	0.1542	0.1463	0.0012	0.000
9	0.4512	0.3703	0.3489	0.3571	0.3546	0.1798	0.3183	0.0816	0.031
10	-0.5372	-0.6177	-0.6382	-0.6329	-0.6293	-0.7626	-0.6739	-0.0967	0.549
11	-0.6690	-0.7868	-0.8150	-0.8184	-0.7996	-0.8606	-0.8578	0.0861	-0.846
12	0.0141	-0.1106	-0.1404	-0.1448	-0.1243	-0.1798	-0.1834	-0.0389	-0.207
13	0.2684	0.1251	0.0892	0.0886	0.1000	-0.0321	0.0567	-0.0038	-0.036
14	-1.5392	-1.6159	-1.6356	-1.6293	-1.6275	-1.7707	-1.6703	0.3627	-0.758
15	0.1086	0.0923	0.0881	0.0976	0.0969	-0.0135	0.0461	-0.0011	0.000
16	-0.6526	-0.7294	-0.7493	-0.7422	-0.7421	-0.8964	-0.7827	-0.3480	-0.406
17	0.2052	0.1438	0.1264	0.1394	0.1285	-0.0878	0.0999	-0.0464	0.417
18	0.4460	0.3128	0.2813	0.2749	0.2988	0.2596	0.2366	0.0073	-0.541
19	0.1265	0.0903	0.0816	0.0876	0.0928	0.0058	0.0377	0.0009	-0.187
20	-0.1747	-0.2100	-0.2194	-0.2098	-0.2121	-0.3485	-0.2572	0.0065	-0.319
21	1.1081	1.0604	1.0474	1.0581	1.0527	0.8851	1.0141	-0.0063	0.000
22	0.2709	0.1366	0.1034	0.1021	0.1156	0.0058	0.0676	0.0052	0.233
23	0.5231	0.4118	0.3849	0.3834	0.3988	0.3200	0.3440	-0.0443	0.570
24	0.1227	0.0536	0.0344	0.0465	0.0367	-0.1787	0.0082	0.0429	-0.456
25	0.3904	0.3087	0.2874	0.2946	0.2941	0.1315	0.2553	0.0632	0.214
26	0.0777	0.0442	0.0340	0.0486	0.0362	-0.1677	0.0045	-0.0212	0.177
27	-0.0344	-0.1147	-0.1362	-0.1270	-0.1315	-0.3200	-0.1651	-0.0465	-0.314
28	0.1668	0.0231	-0.0103	-0.0204	0.0106	0.0110	-0.0593	0.0076	-0.209
29	-0.0004	-0.1261	-0.1557	-0.1618	-0.1379	-0.1705	-0.2016	-0.0228	0.562
30	-0.1148	-0.2020	-0.2254	-0.2160	-0.2212	-0.4202	-0.2526	-0.0054	0.389
Sum of Square Error	*8,8473*	*9.2113*	*9.4119*	*9.4228*	*9.3219*	*10.5267*	*9.7334*	*0.2972*	*4.0800*

In order to test and compare the prediction ability of the regression models, a cross validation method is used. The observation set that is located in Table 1 is separated into two parts. One of them includes 25 observations and is called training set. The other includes 5 observations and is called the test set. The test set is (observation no: 3, 9, 17, 24, 29) selected by a simple random sampling method. The outlier observations are not included in the selection process.

The residuals that are obtained from the regression models belonging to the proposed algorithm for the training set and the residuals from the regression models belonging to the robust method for the training set are located in Table 5. The sum of squares of errors for predictions on the training set and test set are located in same table. According to the error criterion, it can be seen that the predictions error from the algorithm that is proposed in this study is smaller than the other.

Table 5. The residuals belonging to observations for training and test set.

NO		X_1	X_2	Y	Network Residual	LSM Residual	Huber Residual	Hample Residual	Tukey Residual	Andrews Residual	LMS Residual	RLS Residual
1	1	6.859	9.688	4.992	-0.3922	0.1461	-0.0537	-0.0777	-0.1093	-0.0807	-0.0059	-0.0887
2	2	7.215	0.617	4.849	-0.1830	0.0059	-0.0362	-0.0405	-0.0334	-0.0499	-0.1421	0.0098
3	4	0.260	9.151	4.994	0.6161	0.1113	-0.0453	-0.0598	-0.0875	-0.0512	-0.0500	-0.0494
4	5	5.528	7.114	3.275	-0.0230	-1.5781	-1.7258	-1.7424	-1.7627	-1.7446	-1.7310	-1.7323
5	6	3.539	2.766	4.837	-0.0357	-0.0268	-0.0879	-0.0924	-0.0935	-0.0937	-0.1809	-0.0468
6	7	9.476	8.443	5.042	-0.0009	0.2108	0.0196	-0.0050	-0.0320	-0.0135	0.0631	-0.0148
7	8	5.968	9.446	5.276	-0.3340	0.4252	0.2342	0.2118	0.1815	0.2103	0.2720	0.2047
8	10	5.103	3.324	4.379	-0.0233	-0.4761	-0.5551	-0.5627	-0.5667	-0.5665	-0.6282	-0.5253
9	11	2.802	6.101	4.208	0.2504	-0.6602	-0.7762	-0.7871	-0.8024	-0.7851	-0.8166	-0.7626
10	12	2.831	6.492	4.887	0.2447	0.0189	-0.1041	-0.1159	-0.1329	-0.1137	-0.1376	-0.0942
11	13	8.287	6.081	5.167	0.1452	0.3294	0.1857	0.1679	0.1514	0.1601	0.1809	0.1777
12	14	5.479	2.999	3.382	-0.0749	-1.4710	-1.5461	-1.5535	-1.5562	-1.5583	-1.6224	-1.5149
13	15	0.423	0.885	5.033	-0.1282	0.1520	0.1398	0.1435	0.1512	0.1466	-0.0059	0.2105
14	16	6.134	2.824	4.273	-0.0516	-0.5763	-0.6517	-0.6596	-0.6616	-0.6656	-0.7267	-0.6215
15	18	2.316	7.121	5.310	0.3017	0.4390	0.3075	0.2949	0.2755	0.2986	0.2815	0.3138
16	19	0.080	2.127	5.036	0.1943	0.1530	0.1207	0.1220	0.1245	0.1266	-0.0059	0.1814
17	20	2.937	1.300	4.755	-0.2218	-0.1120	-0.1443	-0.1447	-0.1395	-0.1459	-0.2664	-0.0875
18	21	5.408	1.343	6.047	-0.2440	1.1938	1.1481	1.1445	1.1489	1.1388	1.0429	1.1947
19	22	6.512	6.041	5.163	0.0356	0.3155	0.1816	0.1661	0.1502	0.1615	0.1644	0.1810
20	23	3.504	5.534	5.409	0.1600	0.5448	0.4351	0.4246	0.4116	0.4250	0.3896	0.4511
21	25	6.879	2.906	5.318	-0.0038	0.4728	0.3922	0.3832	0.3806	0.3759	0.3235	0.4187
22	26	6.793	0.142	5.035	-0.2936	0.1896	0.1580	0.1553	0.1645	0.1464	0.0412	0.2100
23	27	8.325	2.432	4.904	0.1316	0.0669	-0.0128	-0.0225	-0.0235	-0.0327	-0.0801	0.0123
24	28	0.539	8.211	5.012	0.6100	0.1310	-0.0106	-0.0233	-0.0470	-0.0158	-0.0295	-0.0072
25	30	9.298	2.566	4.826	0.3066	-0.0056	-0.0927	-0.1039	-0.1057	-0.1158	-0.1514	-0.0727
Error Training					*1.6830*	*8.3719*	*8.6962*	*8.7708*	*8.8671*	*8.7846*	*8.9565*	*8.7065*
					Network Test Residual	LSM Test Residual	Huber Test Residual	Hample Test Residual	Tukey Test Residual	Andrews Test Residual	LMS Test Residual	RLS Test Residual
1	3	3.199	2.898	5.256	-0.1735	0.3903	0.3285	0.3242	0.3227	0.3235	0.2355	0.3698
2	9	7.562	2.678	5.384	-0.0525	0.5428	0.4626	0.4531	0.4516	0.4444	0.3942	0.4885
3	17	8.976	1.165	5.160	0.3361	0.3268	0.2661	0.2580	0.2628	0.2459	0.1808	0.3001
4	24	9.319	1.516	5.075	-0.0806	0.2437	0.1750	0.1730	0.1690	0.1532	0.0981	0.2045
5	29	1.544	6.900	4.863	0.2279	-0.0123	-0.1362	-0.1471	-0.1655	-0.1425	-0.1710	-0.1246
Error Test					*0.2043*	*0.6133*	*0.4419*	*0.4285*	*0.4331*	*0.4064*	*0.2824*	*0.5228*

5. CONCLUSION

Obtaining prediction values related to the data that we study is an important component of regression analysis. In the study, we have proposed a method for obtaning optimal prediction values and compared various methods. Prediction values, which are obtained from the proposed algorithm, have the lowest error values.

Recently, in our field as well as others, adaptive networks that fall under the heading of neural networks and yield efficient predictions related to data are being used more frequently, and various algorithms have been proposed. In the proposed algorithms, the fuzzy class number of the independent variable is defined intuitively at first, and within the on going process, these class numbers are taken as the basis. In this study, it has been thought to use validity criterion based on fuzzy clustering at the stage of defining level numbers of independent variables. Moreover, as it can be observed in the algorithm in Section 3, an algorithm different from other proposed algorithms has been used for updating central parameters.

A network is formed that contains; the determination of the a priori parameter set and the optimal class numbers related to the data sets that belong to the independent variables; the obtaining of the membership functions appropriate to the distributions that are used in the step of calculation of the a posteriori parameter set; the formation of the switching regression models by using an a posteriori parameter set; and the process of obtaining prediction values from these formed models. The difference between the obtained prediction values and the observed values, that is, the network that decreases the errors to the minimum level, is formed based on the adaptive network architecture that includes a fuzzy inference system based on the fuzzy rules. The process followed in the proposed method can be accepted as an ascendant from other methods since it does not allow intuitional predictions and it brings us to the smallest error. At the same time, this method is robust, since it is not affected by the contradictory observations that can occur at independent variables.

Finally, the prediction values obtained from the networks that are formed through the proposed algorithm are compared with the prediction values obtained from the robust regression methods. According to the indicated error criterion, the errors related to the predictions that are obtained from the network are lower than the errors that are obtained from the robust regression methods and Least Square Method.

CONFLICT OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

1. L. Lung-Fei and H.P. Robert, Switching Regression Models with Imperfect Sample Separation Information-With an Application on Cartel Stability, *Econometrica* **52** (1984) 391-418.

2. M. Michel, Fuzzy Clustering and Switching Regression Models Using Ambiguity and Distance Rejects, *Fuzzy Sets and Systems* **122** (2001) 363-399.

3. E.Q. Richard, A New Approach to Estimating Switching Regressions, *Journal of the American Statistical Association*, **67** (1972) 306-310.

4. X.L. Xie and G. Beni, A validity measure for fuzzy clustering, *IEEE Trans Pattern Anal. Machine Intell.* **13** (1991) 841-847.

5. C. Mu-Song. And S.W. Wang, Fuzzy Clustering Analysis for Optimizing Fuzzy Membership Functions, *Fuzzy Sets and Systems*, **103** (1999) 239-254.

6. C.J. Bezdek, R. Ehrlich and W. Full, FCM: The Fuzzy c_Means Clustering Algorithm, *Computer and Geoscience*, **10** (1984) 191-203.

7. C.B. Cheng and E.S. Lee, Applying fuzzy adaptive network to fuzzy regression analysis, *An International Journal Computers & Mathematics with Applications* **38** (1999) 123-140.

8. C.B. Cheng and E.S. Lee, Switching Regression Analysis by Fuzzy Adaptive Network, *European Journal of Operational Research*, **128** (2001) 647-663.

9. R.J. Jyh-Shing, ANFIS: Adaptive-Network-Based Fuzzy Inference System, *IEEE Transaction on Systems, Man and Cybernetics,* **23** (1993) 665-685

10. E. Nasrabadi and S.M. Hashemi, Robust Fuzzy Regression Analaysis Using Neural Networks, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **16** (2008) 579-598

11. K.S. Kula and A. Apaydın, Fuzzy Robust Regression Analaysis Based on The Rankinf of Fuzzy Sets, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **16** (2008)

12. F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel, *Robust Statistics.* (John-Willey & Sons, New-York 1986).

13. R.V. Hogg, Statistician robustness: One View of Its Use in Applications Today, *The American Statistician*, **33**(1979) 108-115.

14. P.J. Huber, *Robust statistics.(* John Willey & Son 1981).

15. H. Huynh, A Comparison of For Approaches to Robust Regression, *Psychological Bulletin*, **92**, (1982) 505-512.

16. P.J. Rousseeuw and A.M. Leroy, *Robust regression and outlier detection.* (John Willey & Son1987).

17. R.J. Hathaway and J.C. Bezdek, Switching Regression Models and Fuzzy Clustering, *IEEE Transaction on Fuzzy Systems*, **1** (1993)195-204.

18. F. Horia and S. Costel, A New Fuzzy Regression Algorithm, *Anal. Chem.* **68** (1996)771-778.

19. H. Ishibuchi and M. Nii, Fuzzy regression using asymmetric fuzzy coefficients and fuzzed neural networks, *Fuzzy Sets and Systems* **119** (2001) 273-290.

20. H. Ishibuchi and H. Tanaka, Fuzzy Regression Analysis Using Neural Networks, *Fuzzy Sets and Systems*, **50** (1992) 257-265.

21. T.E. Dalkılıç and A. Apaydın, A Fuzzy Adaptive Network Approach to Parameter Estimation in case Where Independent Variables Come From Exponential Distribution, *Journal of Computational and Applied Mathematics*, 233,(2009) 26-45.