IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 07, Issue 10 (October. 2017), ||V2|| PP 70-78

Switching Regression Analyze Based on Fuzzy Clustering in Case that Input Variables Come From Normal Distribution

Türkan Erbay Dalkılıç[1], Ayşen Apaydın[2]
[1]*Department of Statistics and Computer Sciences, Karadeniz Technical University, Trabzon, Turkey*
tedalkilic@gmail.com
[2]*Department Department of Insurance and Actuarial Sciences, Ankara University, Ankara, Turkey*
aapaydin@ankara.edu.tr

Abstract : In regression analysis, data set can be formed by collecting observations that have been obtained from more than one class. The regression model, which will be formed when each class is defined by an $\boldsymbol{f_i}$ function and c shows the number of classes, can be called switching regression model. To constitute this model, the first step is to determine fuzzy sets numbers depending on class numbers related to independent variables. Then, fuzzy rules are formed depending on specified fuzzy sets. In the studies carried out, class numbers of data have been proposed intuitively at first and in forming fuzzy sets numbers and rules, these class numbers have been taken as basis. The aim of this study is to use validity criterion in determining optimal class number of independent variables and to use adaptive network in obtaining parameter estimations related to the models to be formed.

Keywords: - Adaptive network, Fuzzy rule, Switching regression, Validity criterion

---

Date of Submission: 28-10-2017 Date of acceptance: 16-11-2017

I. INTRODUCTION

When data set is formed by bringing observations obtained from more than one class together and *c* stands for class number, regression model to be formed when each class is indicated with different functions is called switching regression model [1-3]. And it is indicated as

$$Y_i = f_i(X) + \varepsilon_i \quad (1 \le i \le c) \tag{1}$$

To obtain prediction values for data coming from different classes, the adaptive network based on fuzzy inference system will be used. However, it is necessary to determine fuzzy class numbers related to independent variables at first. Since independent variables are fuzzy, the proposed validity criterion for determining optimal class number is given with

$$S = \frac{\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^2\left\|v_i - x_j\right\|}{n\min_{i,j}\left\|v_i - v_j\right\|^2} \tag{2}$$

There have been various studies on fuzzy clustering and validity criterion. Xuanlı Xie and Gerardo Beni proposed a validity criterion for fuzzy clustering in their study in 1991 [4]. Fuzzy clustering analyze for determining fuzzy memberships was used in the study carried out by Mu-song Chen in 1999 and in this study a method for determining optimal class numbers related to variables was proposed. N. Zahid, M. Limoury and A. Essaid deal with a new fuzzy validity criterion for fuzzy clustering in their study in 1999 [5].

Adaptive networks giving efficient solutions when it is unclear that which class data belong to and fuzzy inference system will be dealt with in the second part of the study. There have been also various studies on the usage of adaptive network in parameters estimations. C. B. Cheng and E. S. Lee dealt with the approach of fuzzy adaptive network for fuzzy regression analysis in their study in 1999 [6] and they studied on switching regression and fuzz adaptive networks in 2001 [7]. Jhy-Shing Roger Jang used adaptive network based on fuzzy inference system in his study [8]. In the second part of the study, fuzzy if-then rules and adaptive networks are will be explained. In the third part of the study, validity criterion for fuzzy clustering will be dealt with. In the fourth part, the algorithm which proposed for prediction of parameters related to switching regression model will be dealt with. In the fifth part, practices and in the sixth part conclusion and discussion take place.

II. FUZZY IF-THEN RULES AND ADAPTIVE NETWORK

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

$$\begin{gathered} R^L = If \quad ;\left(x_1 \text{ is } F_1^L \text{ and } x_2 \text{ is } F_2^L, \dots, x_p \text{ is } F_p^L\right) \\ Then\ ; Y = Y^L = c_0^L + c_1^L x_1 + \cdots + c_p^L x_p \end{gathered} \tag{3}$$

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

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

Neural networks enabling the use of fuzzy inference system for fuzzy regression analysis is known as adaptive network. Used for obtaining a good approach to regression functions and formed via neurals and connections adaptive network consists of five layers [9-12].

Neurons which are forming network is characterized with parameter functions. The processing of adaptive network consisting five layers is as follows. Functional relation 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 numbers of independent variables and class or fuzzy sets number forming independent variables. When independent variable number is indicated with *p*, if level number belonging to each variable is indicated with $l_i\ (i = 1, \dots, p)$ fuzzy rule number is indicated with $L = \prod_{i=1}^{p} l_i$.

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

$$f_{1,h} = \mu_{F_h}(x_i) \quad h = 1, \dots, L \quad i = 1, \dots, p \tag{4}$$

When fuzzy clusters related to fuzzy rules are indicated with $F_1, F_2, \dots, F_h$. Here μ_{F_h} is the membership function relates to F_h. Different membership functions are can be define for F_h. Here, membership functions are defined as

$$\mu_{F_h}(x_i) = exp\left[-\left(\frac{x_i - v_h}{\sigma_h}\right)^2\right] \tag{5}$$

because it is thought that data come from Normal Distribution and parameter set of Normal Distrubution are $\{v_h, \sigma_h\}$. Parameter set $\{v_h, \sigma_h\}$ in this layer indicates priori parameters.

Each nerve in the second layer is fixed layer. They have input signals coming from the first layer and they are defined as multiplication 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 and nerve function is defined as

$$f_{3,L} = \overline{w}^L = \frac{w^L}{\sum_{L=1}^{m} w^L} \tag{6}$$

The output signals of the fourth layer are also connected to a function and this function is indicated with

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

here, Y^L stands for conclusion part of fuzzy if-then rule and it is indicated with $Y^L = c_0^L + c_1^L x_1 + \cdots + c_p^L x_p$, where 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 and it is counted by

$$f_{5,1} = \hat{Y} = \sum_{L=1}^{m} \overline{w}^L Y^L \tag{7}$$

as total of all signals [6,7].

III. FUZZY CLUSTERING AND DETERMINING OPTIMAL CLASS NUMBER

The aim of clustering analysis is to divide data set into sub groups having fewer data. In the suggested methods for classic clustering analysis, each datum belongs to only one cluster. Such methods are known as definite clustering methods. On the other hand, the formulation of fuzzy clustering assumes that one datum or observation might belong to different clusters at the same time. Widely used algorithms for definite clustering and fuzzy clustering are respectively definite c-means and fuzzy c-means algorithm, because of counting activity. It is necessary to define explicitly *c*-class number in the usage of these algorithms.

Cluster validity used for determining the optimal cluster or class number of data set and various functions called validity criterion take place in literature.

Clustering algorithm based on fuzzy *c*-means (FCM) is equivalent with definite algorithm and it is based on minimizing of objective function given with

$$J_m = \sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij})^m d^2(x_j, v_i) \tag{8}$$

according to v_i indicated as. Cluster center and to μ_{ij} defined as fuzzy membership. Here;

d : distance between each observation and cluster centers

c : class numbers

n : observation numbers

m : fuzziness index

Clustering algorithm based on fuzzy *c*-means consists of the following steps:

Step 1: Initial membership of belonging *i*. class of x_j, $\mu_{ij}$are defined as

$$\sum_{i=1}^{c}\mu_{ij} = 1$$

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

$$v_i = \frac{\sum_{j=1}^{n}(\mu_{ij})^m x_j}{\sum_{j=1}^{n}(\mu_{ij})^m} \quad i = 1, \dots, c \tag{9}$$

Step 3: Fuzzy memberships defined in step l are updated by using Eq. (10)

$$\mu_{ij} = \frac{\left(\frac{1}{d^2(x_j, v_i)}\right)^{\frac{1}{(m-1)}}}{\sum_{i=1}^{c}\left(\frac{1}{d^2(x_j, v_i)}\right)^{\frac{1}{(m-1)}}} \tag{10}$$

Step 4: Updating in step 2 and 3 are carried on till the amount of decrease in the value of J_m defined as objective function will become lower than a small stable specified in advance [4,5].

So, memberships and cluster centers giving optimal value to objective function are determined. However, in forming a switching regression model one of the important points is that it is necessary to determine how many clusters and therefore how many fuzzy sets that data set should be divided into related to independent variables.

At this stage, there is problem of determining optimal numbers of clusters. In parameter prediction studies carried out via adaptive networks, class numbers of data sets related to independent variables are determined intuitively at first. In this study it is aimed to use validity criterion for determining optimal class number.

Validity criterion has been designed to find optimal numbers of class and determine compacted and separated clusters. There are lots of validity criterions for cluster in literature. In this study *S* function also called as Xie – Beni index for it is easy to count and it is comprehensible, intuitively will be used [4].

When total validity is defined as

$$\sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij})^m \|v_i - x_j\|^2 \tag{11}$$

compactness of a fuzzy division is defined as

$$comp = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij})^m \|v_i - x_j\|^2 \tag{12}$$

and measurement of difference between the cluster centers defined also as separation of a fuzzy division is given as

$$sep = \min_{i \neq j} \|v_i - v_j\|^2 \tag{13}$$

Thus, validity function of fuzzy clustering is

$$S = \frac{comp}{sep} \tag{14}$$

As it can be seen in this statement, cluster centers which are separated well produce a high value for separation so a smaller S value is obtained. Optimal value of class number (c), can be obtained by minimizing S for $c = 2, c = 3, \dots, c = max$. When the lowest S value is found, class number (c) giving this lowest S value is defined as optimal class number [13].

IV. AN ALGORITHM FOR PARAMETERS ESTIMATION OF SWITCHING REGRESSION MODEL

Prediction of parameter with adaptive network is based on the principle of minimizing of error criterion. An algorithm has been purposed forming regression models related to data coming from different classes by Chi-Bin C. (1999) for the process of obtaining a common prediction set based on these regression models. There are two significant steps in the process of prediction. First of them is to determine priori parameter set characterizing class from which data come and to update these parameters within the process and the other one is to determine posteriori parameters belonging to regression models to be formed. In algorithm which is suggested by Chi-Bin, posteriori parameter set $c_i^L = (a_i^L, b_i^L)$ are obtained via solving the linear programming problem which is succested by Tanaka [6, 7], and the problem is indicated as

$$\begin{gathered} min \sum_{k=1}^{N} \sum_{L=1}^{m} \sum_{i=1}^{p} \overline{w}^L b_i^L x_{ik} \qquad \left(= min \sum_{k=1}^{N} \hat{e}_k \right) \\ b_i^L \geq 0 \quad i = 1, \dots, p \quad L = 1, \dots, m \\ \sum_{L=1}^{m} \sum_{i=0}^{p} \overline{w}^L a_i^L x_{ik} + (1-\alpha) \sum_{L=1}^{m} \sum_{i=0}^{p} \overline{w}^L b_i^L x_{ik} \geq y_k + (1-\alpha) e_k \\ -\sum_{L=1}^{m} \sum_{i=0}^{p} \overline{w}^L a_i^L x_{ik} + (1-\alpha) \sum_{L=1}^{m} \sum_{i=0}^{p} \overline{w}^L b_i^L x_{ik} \geq -y_k + (1-\alpha) e_k \end{gathered} \tag{15}$$

Updating of priori parameters determined intuitively at first within the process is based on back probagation errors defined as

$$\epsilon_{5,1} = \frac{\partial \varepsilon_k^2}{\partial y_k} \qquad \qquad \epsilon_{r,l} = \sum_{h=1}^{M_{r+1}} \epsilon_{r+1,h} \frac{\partial F_{r+1,h}}{\partial f_{r,l}}$$

and it is obtained via

$$\Delta\rho = -\eta \frac{\partial (y_k - \hat{y}_k)^2}{\partial \rho} \tag{16}$$

equality. Here ρ stands for priori parameter set and η stands for learning rate. Learning rate having value on the interval (0 1] is determined by decision maker.

The process of determining parameters for switching regression model begins with determining class numbers of independent variables and priori parameters. In this study, it is aimed to use validity criterion based on fuzzy clustering as an alternative to intuitive methods in determining class numbers. Moreover, a structure is formed which includes determining priori parameters as dependent to change interval of data set. In updating priori parameters, instead of the method in which spread back error bringing a series of transaction, a process has been formed which spread back error bringing a series of transaction, a process has been formed which enables to review all the values that parameter might have the lowest error. The algorithm related to the proposed method for determining switching regression model in case of independent variables coming from normal distribution is defined as follows

Step 0: Optimal class numbers related to data set belonging to independent variables are determined. Different values of S function is obtained with

$$S_k = \frac{\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij})^m \|v_i - x_j\|^2}{\min_{i\neq j}\|v_i - v_j\|^2} \quad k = 1, \dots, c$$

for all the values that *c* standing for class number $c = 2,\ c = 3, \dots, c = max$ and c used in counting the lowest of S_k values is defined as optimal class number.

Step 1: 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 variables Center parameters are based on the space in which variables gain value and fuzzy class number and it is defined with

$$v_i = min(x_i) + \frac{max(x_i) - min(x_i)}{(c-1)}(i-1) \quad i = 1, \dots, p \tag{17}$$

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

Step 2: $\bar{w}^L$ weights are counted which are used to form matrix B to be used in counting posteriori parameter set. As they are defined in the second part, $\bar{w}^L$weights are outputs of the nerves in the third layer of adoptive network and they are counted depending on membership function related to distribution family which independent variable belongs to. Nerve functions in the first layer of adaptive network are defined as

$$f_{1,h} = \mu_{F_h}(x_i)$$

with the said membership functions. For F_h there may be several functions appropriate membership function. Here, when Normal Distribution function parameter set of which is $\{v_h, \sigma_h\}$ is thought membership functions are defined as

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

Here Parameter set $\{v_h, \sigma_h\}$ indicates priori parameters. From defined membership functions, membership degrees related to each class forming independent variables are determined. w^L weights are indicated as

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

They are obtained via mutual multiplication of membership degrees at an amount depending on number of independent variable and fuzzy class numbers of these variables. $\bar{w}^L$ weights are normalization of the weights defined as w^L and they are counted with

$$\bar{w}^L = \frac{w^L}{\sum_{L=1}^{m} w^L}$$

Step 3: In case that independent variables are composed of fuzzy numbers, and depended variables are composed of definite numbers, posteriori parameter set is obtained as definite numbers, in such form $c_i^L = (a_i^L, b_i^L),\ c_i^L = a_i^L$. In this case, to determine posteriori parameter set,

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

equality is used [1].

Here Y, Z and B defined as

$Y = [y_1, y_2, \dots, y_n]^T$

$Z = \left[a_0^1, \dots, a_0^m, a_1^1, \dots, a_1^m, a_p^1, \dots, a_p^m\right]^T$

and

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

Step 4: By using posteriori parameter set $c_i^L = (a_i^L, b_i^L)$ obtained in Step 3, switching regression model indicated with

$$Y^L = c_0^L + c_1^L x_1 + c_2^L x_2 + \cdots + c_p^L x_p$$

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

$$\hat{Y} = \sum_{L=1}^{m} \bar{w}^L Y^L$$

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

$$\varepsilon_k = Y_k\{-\}\hat{Y}_k \qquad k = 1, \dots, n \tag{18}$$

error related to model is counted as

$$\varepsilon_k = \frac{\sum_{k=1}^{n}(Y_k - \hat{Y}_k)^2}{n} \tag{19}$$

If $\varepsilon < \emptyset$, then posteriori parameter has been obtained as parameters of regression models to be formed, the process is concluded. If $\varepsilon \geq \emptyset$, then , step 6 begins, Here $\emptyset$, is a law stable value determined by decision maker, $\{-\}$ is subtraction operator in case of the fuzziness of dependent variable as well.

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

$$v_i' = v_i \pm t \tag{20}$$

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

$$t = \frac{max(x_{ij}) - min(x_{ji})}{a} \quad j = 1, \dots, n \quad i = 1, \dots, p$$

and a is stable value which is determinant of size of step and therefore iteration number.

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

V. NUMERICAL EXAMPLE

The proposed algorithm was operated with a programme written in MATLAB. In the stage of step operating, data sets having two independent variables for this study were dealt with and they were compared to the model structures which had given result. It was seen in the comparisons that models obtained via adaptive network reached prediction values giving the lowest error.

5.1. Prediction of parameters of switching regression models in case of two independent variables

The values related to data set having two independent variables and one dependent variable are shown in Table 1. The values in data set have been taken from the study carried out by C.B. Cheng (1999) and switching regression model and predictions for this model are obtained via algorism proposed for this data set in the 4th part. Moreover, predictions have been obtained via using the least squares method (LSM) to data set.

Prediction values $(\hat{y}_{Network})$ obtained via adaptive network related to this data set and errors related to these predictions $(e_{(Network)i}\ (i = 1, \dots, n))$, predictions obtained via the least squares method $(\hat{y}_{LSM})$ and errors related to these predictions $(e_{(LSM)i}\ (i = 1, \dots, n))$ and the results drawn by Chi-Bin $(\hat{y}_{Chi-Bin})$ and errors related to these predictions $(e_{(Chi-Bin)i}\ (i = 1, \dots, n))$ are shown in Table 1.

Table 1: Predictions and Error Values for data set related to two independent variables

X_1	X_2	Y	$\hat{y}_{Network}$	$e_{(Network)i}$	$\hat{y}_{LSM}$	$e_{(LSM)i}$	$\hat{y}_{Chi-Bin}$	$e_{(Chi-Bin)i}$
6.859	9.688	4.992	4.9906	0.0014	4.8539	0.1381	5.030	-0.038
7.215	0.617	4.849	4.8967	-0.0477	4.9538	-0.1048	5.126	-0.277
3.199	2.898	5.256	5.2841	-0.0281	4.9133	0.3427	4.776	0.480
0.260	9.151	4.994	4.9917	0.0023	4.8339	0.1601	4.994	0.000
5.528	7.114	3.275	3.2753	-0.0003	4.8766	-1.6016	3.428	-0.153
3.539	2.766	4.837	4.7801	0.0569	4.9161	-0.0791	4.572	0.265
9.476	8.443	5.042	5.0428	-0.0008	4.8776	0.1644	5.042	0.000
5.968	9.446	5.276	5.2772	-0.0012	4.8530	0.4230	5.276	0.000
7.562	2.678	5.384	5.4656	-0.0816	4.9328	0.4512	5.353	0.031
5.103	3.324	4.379	4.2823	0.0967	4.9162	-0.5372	3.830	0.549
2.802	6.101	4.208	4.2941	-0.0861	4.8770	-0.6690	5.054	-0.846
2.831	6.492	4.887	4.8481	0.0389	4.8729	0.0141	5.094	-0.207
8.287	6.081	5.167	5.1632	0.0038	4.8986	0.2684	5.203	-0.036
5.479	2.999	3.382	3.7447	-0.3627	4.9212	-1.5392	4.140	-0.758
0.423	0.885	5.033	5.0319	0.0011	4.9244	0.1086	5.033	0.000
6.134	2.824	4.273	3.9250	0.3480	4.9256	-0.6526	4.679	-0.406
8.976	1.165	5.160	5.1136	0.0464	4.9548	0.2052	4.743	0.417
2.316	7.121	5.310	5.3173	-0.0073	4.8640	0.4460	5.851	-0.541

0.080	2.127	5.036	5.0369	-0.0009	4.9095	0.1265	5.223	-0.187
2.937	1.300	4.755	4.7615	-0.0065	4.9297	-0.1747	5.074	-0.319
5.408	1.343	6.047	6.0407	0.0063	4.9389	1.1081	6.047	0.000
6.512	6.041	5.163	5.1682	-0.0052	4.8921	0.2709	4.930	0.233
3.504	5.534	5.409	5.3647	0.0443	4.8859	0.5231	4.839	0.570
9.319	1.516	5.075	5.1179	-0.0429	4.9523	0.1227	5.531	-0.456
6.879	2.906	5.318	5.3812	-0.0632	4.9276	0.3904	5.104	0.214
6.793	0.142	5.035	5.0138	0.0212	4.9573	0.0777	4.858	0.177
8.325	2.432	4.904	4.8575	0.0465	4.9384	-0.0344	5.218	-0.314
0.539	8.211	5.012	5.0196	-0.0076	4.8452	0.1668	5.221	-0.209
1.544	6.900	4.863	4.8402	0.0228	4.8634	-0.0004	4.301	0.562
9.298	2.566	4.826	4.8206	0.0054	4.9408	-0.1148	4.437	0.389
Error			$\varepsilon_{Network} = 0.0099$		$\varepsilon_{LSM} = 0.2649$		$\varepsilon_{Chi-Bin} = 0.1360$	

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 is as

$$\begin{aligned}
\hat{y}_1 &= 556 + 133x_1 - 1x_2 \\
\hat{y}_2 &= -1143 - 273x_1 + 100x_2 \\
\hat{y}_3 &= 624 + 563x_1 + 132x_2 \\
\hat{y}_4 &= -2662 + 376x_1 - 304x_2 \\
\hat{y}_5 &= 6875 - 788x_1 - 911x_2 \\
\hat{y}_6 &= 781 + 1665x_1 - 617x_2 \\
\hat{y}_7 &= -263 + 213x_1 + 614x_2 \\
\hat{y}_8 &= 2056 - 460x_1 + 1947x_2 \\
\hat{y}_9 &= -2813.1 + 1026x_1 + 1282x_2
\end{aligned} \tag{21}$$

The error related to predictions obtained via the models given with (21) equality is found as

$$\varepsilon_{Network} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n} = 0.0099$$

The model obtained via the least squares method for data is as

$$\hat{y}_{LSM} = 4.9323 + 0.0039x_1 - 0.0109x_2$$

and error value related to this model is

$$\varepsilon_{LMS} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n} = 0.2949$$

Moreover the error related to predictions obtained by Chi-Bin is

$$\varepsilon_{Chi-Bin} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n} = 0.1360$$

From proposed algorithm, initial center values defined while forming the models related to fuzzy rules and central values giving the best predictions are shown in Table 2.
The graphs of errors obtained via, propossed algorithm, least square method and ChiBin's model are shown as compared and separated in Figure l. In Figur 1-a, estimation errors from fuzzy adaptive network which is reladet to porposed algorithim in this work, in Figur 1-b, estimation errors from LSM, and in Figur 1-c, estimation error from Chi-bin's method, are shown.
As it can be seen in Figure l d, errors related to predictions obtained via network proposed for data, are lower this conclusion is also supported by values obtained via error criterion.

Table 2. Initial and conclusion values related to the centers obtained via network.

		X_1			X_2		
		Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
Initial	v_i	0.0800	4.7780	9.4760	0.1420	4.9150	9.6880
Conclusion	v_i	1.0216	4.7780	8.5344	1.0836	4.9150	8.7464

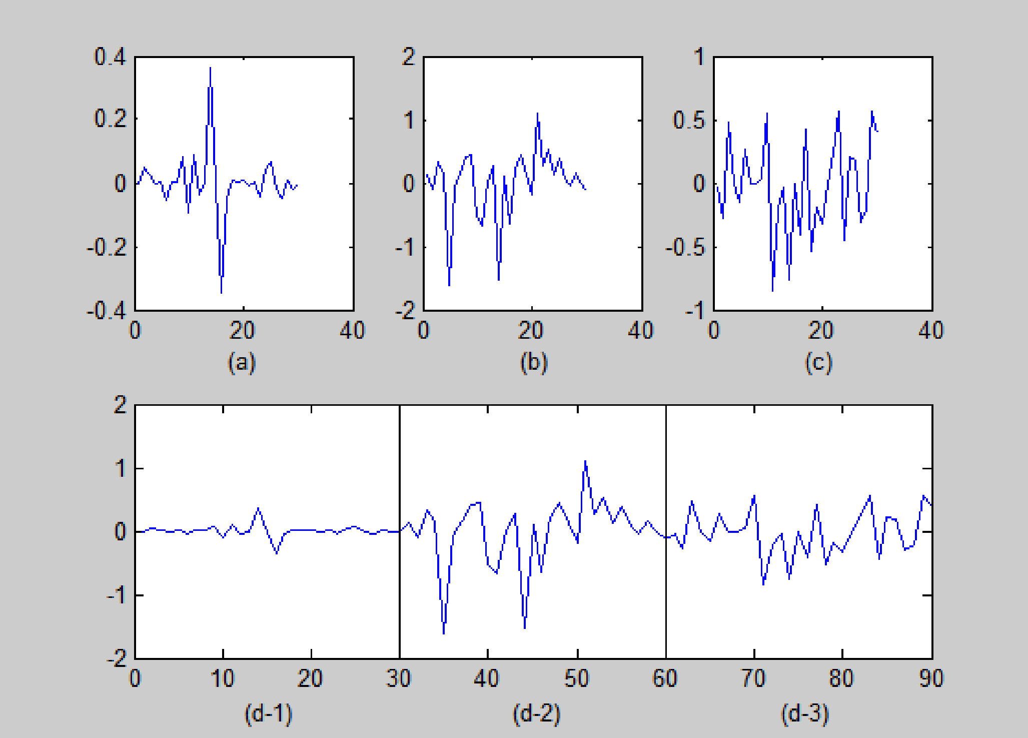

Figure l. Graphs for errors related to data set in Table l.

VI. CONCLUSION

Obtaining prediction values related to data which we studied on is an important part of regression analysis. In the studies, various methods for obtaining prediction values having the best value, in other words having the lowest error values have been proposed and compared with other methods.

Recently, as well as in the other fields, adaptive networks taking place under the heading of neural networks giving efficient results in obtaining predictions related to data is used and various, algorithm are proposed. In the proposed algorithm, fuzzy class number of independent variable is defined intuitively at first and within ongoing process, these class numbers are taken as basis. In this study, it has been thought to use validity criterion based on fuzzy clustering at the stage of defining fuzzy class numbers of independent variables, moreover, as it can be observed in algorithm in the 4th part, an algorithm different from other proposed algorithms has been used for updating central parameters. Prediction values obtained via network to be formed via bringing all these conditions together have been compared with the least squares method. Error criterion used in the comparison is based on the rate of total squares of differentials between real values and predictions. As it can be seen in numerical examples, errors related to predictions obtained via network according to error criterion is lower than errors obtained via network proposed before the errors obtained via the least squares method.

REFERENCES

[1] Lung-Fei L, Robert HP (1984) Switching regression models with imperfect sample separation information-with an application on cartel stability, Econometrica, (52):391-418
[2] Michel M (2001) Fuzzy clustering and switching regression models using ambiguity and distance rejects, Fuzzy Sets and Systems (122):363-399
[3] Richard EQ (1972) A new approach to estimating switching regressions, Journal of the American Statistical Association 67 (338):306-310
[4] Xie XL, Beni G (1991) A validity measure for fuzzy clustering, IEEE Trans Pattern Anal. Machine Intell. 13 (8): 841-847
[5] Zahid N, Limouri M, Essaid A (1999) A new cluster-validity for fuzzy clustering, The Journal of the Pattern Recognition 32:1089-1097
[6] Chi-Bin C, Lee ES (1999) Applying fuzzy adaptive network to fuzzy regression analysis, An International Journal Computers & Mathematics With Applications 38:123-140
[7] Chi-Bin C, Lee ES (2001) Switching regression analaysis by fuzzy adaptive network, Europen Journal of Operational Research 128:647-663

[8] Jyh-Shing Roger Jang (1993) ANFIS: Adaptive-network-based fuzzy inference system, IEEE Transaction on Systems, Man and Cybernetics 23 (3): 665-685
[9] Hisao I, Manabu N (2001) Fuzzy regression using asymmetric fuzzy coefficients and fuzzied neural networks, Fuzzy Sets and Systems 119: 273-290
[10] Hisao I, Tanaka H (1992) Fuzzy regression analaysis using neural networks, Fuzzy Sets and Systems 50:257-265
[11] Horia F, Costel S (1996) A new fuzzy regression algorithm, Anal. Chem. 68:771-778
[12] James PD, Donalt W (2000) Fuzzy regression by fuzzy number neural networks, Fuzzy Sets and Systems 112:371-380
[13] Mu-Song Chen, Wang SW, (1999) Fuzzy Clustering Analaysis for Optimizing Fuzzy Membership Functions, Fuzzy Sets and Systems, Vol 103, 239-254

Türkan Erbay Dalkılıç Switching Regression Analyze Based on Fuzzy Clustering in Case that Input Variables Come From Normal Distribution." IOSR Journal of Engineering (IOSRJEN) , vol. 7, no. 10, 2017, pp. 70-78.