Hacettepe Journal of Mathematics and Statistics
Volume 43 (2) (2014), 309 – 322

Parameter estimation by anfis where dependent variable has outlier

Türkan Erbay Dalkılıç [a] *, Kamile Şanlı Kula [b] †, and Ayşen Apaydın [c] ‡

Abstract

Regression analysis is investigation the relation between dependent and independent variables. And, the degree and functional shape of this relation is determinate by regression analysis. In case that dependent variable has outlier, the robust regression methods are proposed to make smaller the effect of the outlier on the parameter estimates. In this study, an algorithm has been suggested to define the unknown parameters of regression model, which is based on ANFIS (Adaptive Network based Fuzzy Inference System). The proposed algorithm, expressed the relation between the dependent and independent variables by more than one model and the estimated values are obtained by connected this model via ANFIS. In the solving process, the proposed method is not to be affected the outliers which are to exist in dependent variable. So, to test the activity of the proposed algorithm, estimated values obtained from this algorithm and some robust methods are compared.

Keywords: Adaptive network, fuzzy inference, robust regression.

2000 AMS Classification: 62J05, 62K25

[a]Department of Statistics and Computer Science, Faculty of Science, Karadeniz Technical University, Trabzon, Turkey.
*Email: tedalkilic@gmail.com
[b]Department of Mathematics, Faculty of Sciences and Arts, Ahi Evran University, Kirsehir, Turkey
†Email: sanli2004@hotmail.com
[c]Department of Statistics, Faculty of Science, Ankara University, 06100, Tandogan, Ankara, Turkey
‡Email: apaydin@science.ankara.edu.tr

1. Introduction

In a regression analysis, it is assumed that the observations come from a single class in a data cluster and the simple functional relationship between the dependent and independent variables can be expressed using the general model; $Y = f(X) + \varepsilon$. However; a data set may consist of a combination of observations that have different distributions that are derived from different clusters. When faced with issues of estimating a regression model for fuzzy inputs that have been derived from different distributions, this regression model has been termed the 'switching regression model' and it is expressed with $Y^L = f^L(X) + \varepsilon^L$ $(L = \prod_{i=1}^{p} l_i)$. Here l_i indicates the class number of each independent variable and p is indicative of the number of independent variables [18, 19, 21]. In case that, the class numbers of the data and the number of the independent variables are more than two, simultaneously the numbers of sub-models are increased. At this stage, the method attempts to utilize the neural networks, which are intended to solve complex problems and systems. When faced with issues in which the data belong to an indefinite or fuzzy class, the neural network, termed the adaptive network, is used for establishing the regression model. In this study, adaptive networks have been used to construct a model that has been formed by gathering obtained models. There are methods that suggest the class numbers of independent variables heuristically. Alternatively, in defining the optimal class number of independent variables, the use of suggested validity criterion for fuzzy clustering has been aimed. There are many studies on the use of the adaptive network for parameter estimation. In a study by Chi-Bin, C. and Lee, E. S. a fuzzy adaptive network approach was established for fuzzy regression analysis [4] and it was studied on both fuzzy adaptive networks and the switching regression model [5]. Jang, J. R. studied the adaptive networks based on a fuzzy inference system [16]. In a study of Takagi, T. and Sugeno, M., the method for identifying a system using it's input-output data was presented [23]. James, P. D. and Donalt, W., were studied fuzzy regression using neural networks [15]. In a study by Cichocki, A. and Unbehauen, R., the different neural networks for optimization were explained [2]. There are different studies about fuzzy clustering and the validity criterion. In the study of Mu-Song, C. and Wang, S.W. 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 [20]. Bezdek, J.C. has conducted important studies on the fuzzy clustering topic [1]. One such study is by Hathaway R.J. and Bezdek J.C. were studied on switching regression and fuzzy clustering [7]. In 1991, Xie, X.L. and Beni, G. suggested a validity criterion for fuzzy clustering [24]. In this study we used the Xie-Beni validity criterion for determining optimal class numbers. Over the years, the least squares method(LSM) has commonly been used for the estimation of regression parameters. If a data set conforms to LSM assumptions, LSM estimates are known to be the best. However, if outliers exist in the data set, the LSM can yield bad results. In the conventional approach, outliers are removed from the data set, after which the classical method can be applied. However, in some research, these observations are not removed from the data set. In such cases, robust methods are preferred to the LSM [17]. The remainder of the paper is organized as follows. Section 2 explores the fuzzy if-then rules and the use of these rules will be introduced using adaptive networks for analysis. In Section 3 an algorithm for parameter estimation based ANFIS is given.In Section 4, we provide definitions of M methods of Huber, Hampel, Andrews and Tukey, which are commonly used in the literature. In Section 5, a numerical application examining the work and validity of the suggested algorithm as well as a comparison of the algorithm with these robust methods and LSM is provided.In the last part, a discussion and conclusion are provided.

2. ANFIS: Adaptive Network based Fuzzy Inference System

The most popular application of fuzzy methodology is known as fuzzy inference systems. This system forms a useful computing framework based on the concepts of fuzzy set theory, fuzzy reasoning and fuzzy if-then rules. Fuzzy inference systems usually perform on input-output relation, as in control applications where the inputs correspond to system state variables, and the outputs are control signals [3, 5, 16]. The fuzzy inference system is a powerful function approximater. The basic structure of a fuzzy inference system consist of five conceptual components;a rule base which contains a selection of fuzzy rules, a database which defines the membership functions of the fuzzy sets used in the fuzzy rules, a decision-making unit which performs inference operations on the rules, a fuzzification interface which transforms the crisp inputs into degrees of match with linguistic values, and a defuzzification interface which transform the fuzzy results of the inference into a crisp output [3, 15, 16]. The adaptive network used to estimate the unknown parameters of regression model is based on fuzzy if-then rules and fuzzy inference system. When issues of estimating a regression model to fuzzy inputs from different distributions arose, the Sugeno Fuzzy Inference System is appropriate and the proposed fuzzy rule in this case is indicated as

$$R^L = If; \left(x_1 = F_1^L \text{ and } x_2 = F_2^L \text{ and } \quad \dots \quad x_p = F_p^L\right).$$

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

Here, F_i^L stands for fuzzy cluster and Y^L stands for system output according to the R^L rule [16, 23].

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

$$\widehat{Y} = \frac{\sum\limits_{L=1}^{m} w^L Y^L}{\sum\limits_{L=1}^{m} w^L}.$$

Here; w^L weight is indicated as,

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

$\mu_{F_i^L}(x_i)$ is a membership function defined on the fuzzy set F_i^L, and m is fuzzy rule number [13, 14].

Neural networks that enable the use of fuzzy inference systems for fuzzy regression analysis is known as adaptive network and called ANFIS. An adaptive network is a multilayer feed forward network in which each node performs a particular function on incoming signals as well as a set of parameters pertaining to this node. The formulas for the node functions may wary from node to node and the choice of each node function depends on the overall input-output function of the network. Neural networks are used to obtain a good approach to regression functions and were formed via neural and adaptive network connections consistingof five layers $[4, 12-14, 15]$.

Fuzzy rule number of the system depends on numbers of independent variables and fuzzy class number forming independent variables. When independent variable number is indicated with p and if the fuzzy class number associated with each variable is indicated by l_i $(i = 1, ..., p)$, the fuzzy rule number indicated by

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

To illustrate how a fuzzy inference system can be represented by ANFIS, let us consider the following example. Suppose a data set has two-dimensional input $X = (x_1, x_2)$. For input x_1, there are two fuzzy sets "tall" and "short" and for input x_2, three fuzzy set "thin", "normal" and "fat". In this case a fuzzy inference system contains the following six rules:

$$\begin{aligned}
R^1 &: If(x_1 \textit{ is tall and } x_2 \textit{ is thin}), then; (Y^1 = c_0^1 + c_1^1 x_1 + c_2^1 x_2),\\
R^2 &: If(x_1 \textit{ is tall and } x_2 \textit{ is normal}), then; (Y^2 = c_0^2 + c_1^2 x_1 + c_2^2 x_2),\\
R^3 &: If(x_1 \textit{ is tall and } x_2 \textit{ is fat}), then; (Y^3 = c_0^3 + c_1^3 x_1 + c_2^2 x_2),\\
R^4 &: If(x_1 \textit{ is short and } x_2 \textit{ is thin}), then; (Y^4 = c_0^4 + c_1^4 x_1 + c_2^4 x_2),\\
R^5 &: If(x_1 \textit{ is short and } x_2 \textit{ is normal}), then; (Y^5 = c_0^5 + c_1^5 x_1 + c_2^5 x_2),\\
R^6 &: If(x_1 \textit{ is short and } x_2 \textit{ is fat}), then; (Y^6 = c_0^6 + c_1^6 x_1 + c_2^6 x_2).
\end{aligned}$$

This fuzzy system is represented by the ANFIS as shown in Figure 1. The functions of each node in Figure 1 defined as follows.

Layer 1: The output of node h in this layer is defined by the membership function on F_h

$$f_{1,h} = \mu_{F_h}(x_1) \quad for \quad h = 1, 2$$
$$f_{1,h} = \mu_{F_h}(x_2) \quad for \quad h = 3, 4, 5$$

where fuzzy cluster related to fuzzy rules are indicated with $F_1, F_2, ..., F_h$ and μ_{F_h} is the membership function relates to F_h. Different membership functions are can be define for F_h. In this example, the Gaussian membership function will be used whose parameters can be represented by $\{v_h, \sigma_h\}$.

]

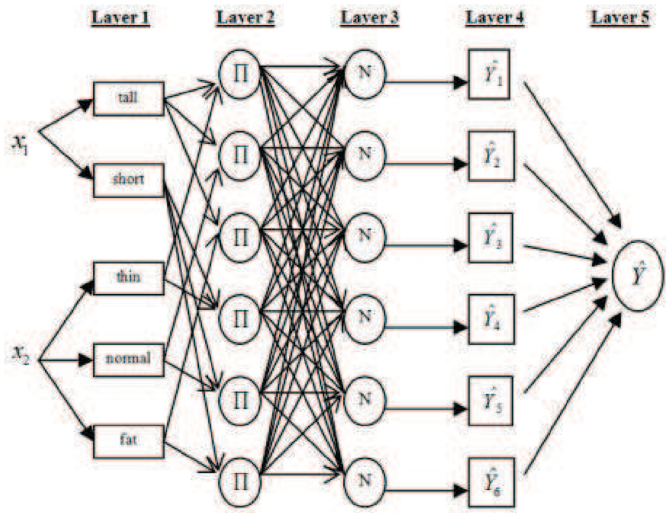

FIGURE 1. The ANFIS architecture

$$\mu_{F_h}(x_1) = \exp\left[-\left(\frac{x_1 - v_h}{\sigma_h}\right)^2\right] \quad for \quad h = 1, 2$$

$$\mu_{F_h}(x_2) = \exp\left[-\left(\frac{x_2 - v_h}{\sigma_h}\right)^2\right] \quad for \quad h = 3, 4, 5.$$

The parameter set $\{v_h, \sigma_h\}$ in this layer is referred to as the premise parameters.

Layer 2: Each nerve in the second layer has input signals coming from the first layer and they are defined as multiplication of these input signals. This multiplied output forms the firing strength w^l for rule l:

$$f_{2,1} = w^1 = \mu_{F_1}(x_1) \times \mu_{F_3}(x_2),$$
$$f_{2,2} = w^2 = \mu_{F_1}(x_1) \times \mu_{F_4}(x_2),$$
$$f_{2,3} = w^3 = \mu_{F_1}(x_1) \times \mu_{F_5}(x_2),$$
$$f_{2,4} = w^4 = \mu_{F_2}(x_1) \times \mu_{F_3}(x_2),$$
$$f_{2,5} = w^5 = \mu_{F_2}(x_1) \times \mu_{F_4}(x_2),$$
$$f_{2,6} = w^6 = \mu_{F_2}(x_1) \times \mu_{F_5}(x_2).$$

Layer 3: The output of this layer is a normalization of the outputs of the second layer and nerve function is defined as

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

Layer 4: The output signals of the fourth layer are also connected to a function and this function is indicated with

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

where, Y^L stands for conclusion part of fuzzy if-then rule and it is indicated with

$$Y^L = c_0^L + c_1^L x_1 + c_2^L x_2,$$

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

Layer 5: There is only one node which computes the overall output as the summation of all the incoming signals

$$f_{5,1} = \widehat{Y} = \sum_{L=1}^{6} \overline{w}^L Y^L.$$

3. An Algorithm for Parameter Estimation Based ANFIS

The estimation 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 estimation. 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 [6]. 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.

Step 1: Optimal class numbers related to the data set associated with the independent variables are determined. Optimal value of class number l_i, ($l_i = 2$, $l_i = 3$ . . . $l_i = \max$) can be obtained by minimizing the fuzzy clustering validity function S_i. This function is expressed by

$$S_i = \frac{\frac{1}{n} \sum_{i=1}^{l_i} \sum_{j=1}^{n} (\ u_{ij})^m \left\| v_i - x_j \right\|^2}{\min_{i \neq j} \left\| v_i - v_j \right\|^2}.$$

As it can be seen in this statement, cluster centers, which are well-separated produce a high value of separation such that a smaller S_i value is obtained. When the lowest S_i value is observed, class number (l_i) with the lowest value is defined as an optimal class number.

Step 2: 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

$$v_i = (\min X_i) + \frac{\max(X_i) - \min(X_i)}{l_i - 1}(i-1), \qquad i = 1, 2, ..., p.$$

Step 3: $\overline{w}^L$ weights are counted which are used to form matrix B to be used in counting the a posteriori parameter set. L is the fuzzy rule number. 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 by

$$f_{1,h} = \mu_{F_h}(x_i) \qquad h = 1, 2, ..., \sum_{i=1}^{p} l_i.$$

$\mu_{F_h}(x_i)$ 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

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

From the defined membership functions, membership degrees related to each class forming independent variables are determined. The w^L weights are indicated as

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

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

$$\overline{w}^L = \frac{w^L}{\sum\limits_{L=1}^{m} w^L}.$$

Step 4: On the condition that the independent variables are fuzzy and the dependent variables are crisp, a posteriori parameter set $c_i^L = (a_i^L, b_i^L)$ is obtained as crisp numbers in the shape of, $c_i^L = a_i^L$ $(i = 1, ..., p)$. In that condition, $Z = \left(B^T B\right)^{-1} B^T Y$ equation is used to determine the a posteriori parameter set. Here B is the data matrix which is weighted by membership degree and its dimension is $[(p+1) \times m \times n]$, Y dependent variable vector and Z is posterior parameter vector which is defined by

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

Step 5: By using a posteriori parameter set $c_i^L = a_i^L$ obtained in Step 4, the regression model indicated by

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

are constituted. Setting out from the models and weights specified in Step 1, the estimation values are obtained using

$$\widehat{Y} = \sum_{L=1}^{m} \overline{w}^L Y^L.$$

Step 6: The error related to model is counted as

$$\varepsilon_k = \sum_{k=1}^{n} \left(y_k - \widehat{y}_k\right)^2 .$$

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 $\varepsilon < \phi$, then, Step 6 begins. Here ϕ, is a law stable value determinated by the decision maker.
Step 7: Central a priori parameters specified in Step 2 are updated with

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

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;

$$t = \frac{\max\left(x_{ji}\right) - \min\left(x_{ji}\right)}{a} \qquad j = 1, 2, ..., n; \quad i = 1, 2, ..., p$$

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

Step 8: Estimations for each a priori parameter obtained by change and the error criteria related to these estimations are counted. The lowest of the error criterion is defined. A priori parameters giving the lowest error specified, and the estimation obtained via the models related to these parameters is taken as output.

In the proposed algorithm, the estimated values which are obtained from the fuzzy adaptive network are not to be affected by the outliers that may exist in the dependent 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.

4. M methods

The classical LSM is widely used in regression analysis because computing its estimate is easy and traditional. However, least square estimators are very sensitive to outliers and to deviations from basic assumptions of normal theory [11, 25]. The importance of eachobservation should therefore be recognized, and the data should be tested in detail whenit is analyzed. This is important because sometimes even a single observation can changethe value of the parameter estimates, and omitting this observation from the data maylead to totally different estimates. If there exist outliers in the data set, robust methods are preferred to estimate parameter values [22]. Now, we discuss the widely used methods of the Huber, Hampel, Andrews and Tukey M estimators. The M estimator utilizes minimizing of residual functions much more thanminimizing the sum of the squared residuals. Regression coefficients are obtained by theminimizing sum:

$$\sum_{i=1}^{n}\rho\left[\left(y_i-\sum_{j=1}^{p}x_{ij}\widehat{\beta}_j\right)\diagup d\right]. \tag{4.1}$$

By taking the first partial derivative of the sum in Equation (4.1) with respect to each $\widehat{\beta}_j$ and setting it to zero, it may be found regression coefficient that p equations:

$$\sum_{i=1}^{n}x_{ij}\Psi\left[\left(y_i-\sum_{j=1}^{p}x_{ij}\widehat{\beta}_j\right)\diagup d\right]=0 \quad j=1,2,...,p$$

where $\Psi(z)=\rho^{'}(z)$. When the data contains outliers, standard deviations are not good measures of variability, and other robust measures of variability are therefore required. One robust measure of variability is d. In the case where r_i is the residual of i^{th} observation, $d=median\left|r_i-median\left(r_i\right)\right|\diagup 0.6745,\ i=1,2,...,n$. Therefore, the standardized residuals may be defined as $z=r_i\diagup d$. Inaddition $r_i=y_i-\sum\limits_{j=1}^{p}x_{ij}\widehat{\beta}_j$.

Huber's Ψ function is defined as:

$$\Psi\left(z\right)=\begin{cases}-k & z<-k\\ z & |z|\leq k\\ k & z>k\end{cases}$$

with k=1.5.

The Hampel Ψ function is defined as:

$$\Psi(z)=\begin{cases}|z| & 0<|z|\leq a\\ asgn(z) & a<|z|\leq b\\ a\left(\frac{c-|z|}{c-b}\right)sgn(z) & b<|z|\leq c\\ 0 & c<|z|\end{cases} \qquad sgn(z)=\begin{cases}+1 & z>0\\ 0 & z=0\\ -1 & z<0\end{cases}$$

Reasonably good values of the constants are $a=1.7$, $b=3.4$ and $c=8.5$.

Andrews (sine estimate) Ψ function is defined as

$$\Psi\left(z\right)=\begin{cases}\sin\left(z\diagup k\right) & |z|\leq k\pi\\ 0 & |z|>k\pi\end{cases}$$

with k =1.5 or $k=2.1$.

The Tukey (biweight estimate) Ψ function is defined as:

$$\Psi\left(z\right)=\begin{cases}\left(z\left(1-\left(z\diagup k\right)^2\right)^2\right) & |z|\leq k\\ 0 & |z|>k\end{cases}$$

with k $=5.0$ or 6.0 $[8-11]$.

5. Numerical Example

The values related to the data set having three independent variables and one dependent variable is shown in Table 1. The values in the data set have been generated from normal distribution such that $X_1 \sim (\mu = 20; \sigma = 3)$, $X_2 \sim (\mu = 50; \sigma = 12)$, $X_3 \sim (\mu = 32; \sigma = 13)$, and dependent variable Y is depend on independent variables value. 5^{th} observation of the dependent variable is changed with $(y_{15} + 50)$ to work up this observation into outlier. The regression models and estimations for this model are obtained via the proposed algorithm for this data set. Moreover, estimations have been obtained using the robust regression methods are used for comparison. The proposed algorithm was executed with a program written in MATLAB. From the initial step of the proposed algorithm, fuzzy class numbers for each variable are defined as two. Number of fuzzy inference rules to be formed depending on these class numbers is obtained as

$$L = \prod_{i=1}^{p=3} l_i = l_1 \times l_2 \times l_3 = 8.$$

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

No	X_1	X_2	X_3	Y	No	X_1	X_2	X_3	Y
1	21.8101	50.5397	49.8319	125.4057	**16**	25.2815	50.2143	54.2714	128.9194
2	19.8248	78.9993	35.1925	137.4526	**17**	20.2663	30.6749	40.9755	60.9541
3	16.6740	46.2813	33.5450	97.0719	**18**	27.7867	64.8650	33.4679	130.0444
4	26.4327	52.2510	37.0010	116.3851	**19**	17.9736	58.2030	17.8756	87.9184
5	15.9415	61.3724	31.0880	107.0015	**20**	28.3604	40.6314	11.7420	78.4568
6	21.3711	43.6916	24.4820	92.0244	**21**	19.9495	56.3718	40.2863	116.6304
7	21.1735	36.6127	38.1010	100.6000	**22**	20.8150	75.6140	26.7405	118.1328
8	26.2190	30.8922	48.8959	90.8950	**23**	17.2577	54.2523	26.7568	103.9698
9	19.0300	64.0981	53.2524	136.5460	**24**	14.1459	52.7804	33.0930	101.7193
10	24.4044	55.8217	22.8635	104.7410	**25**	19.0477	65.4558	26.3405	113.7832
11	18.4928	69.7458	42.4943	133.6250	**26**	21.7650	49.8381	24.6859	93.9008
12	20.6288	44.5492	18.6431	83.1755	**27**	22.4870	33.9999	43.4148	101.4928
13	22.2644	62.1052	48.8284	134.8870	**28**	14.9754	43.3239	21.4096	85.7995
14	17.1554	74.5928	32.1941	126.0083	**29**	14.2331	59.0672	28.6413	105.1197
15	21.8395	57.2242	34.8432	**166.4707**	**30**	18.6900	39.0578	38.4129	99.5382

Models obtained via eight fuzzy inference rules are;

$$\begin{aligned}
\widehat{y}_1 &= 1308 + 346x_1 - 84x_2 - 314x_3 \\
\widehat{y}_2 &= 10896 - 145x_1 + 175x_2 - 230x_3 \\
\widehat{y}_3 &= 9022 - 211x_1 - 126x_2 + 263x_3 \\
\widehat{y}_4 &= -27061 - 24x_1 + 202x_2 + 207x_3 \\
\widehat{y}_5 &= -20670 + 701x_1 - 51x_2 + 436x_3 \\
\widehat{y}_6 &= -6201 - 405x_1 - 155x_2 + 341x_3 \\
\widehat{y}_7 &= 18219 - 610x_1 + 19x_2 - 316x_3 \\
\widehat{y}_8 &= 25742 + 283x_1 - 204x_2 - 283x_3 \qquad (5.1)
\end{aligned}$$

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

Table 2. The estimation of regression parameters

	Constant	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$
LMS	−10.4360	1.0404	1.2420	0.9412
Huber	3.0366	0.8125	1.0329	1.0085
Hampel	5.5338	0.7794	0.9778	1.0412
Tukey	5.3224	0.8127	0.9625	1.0563
Andrews	5.2896	0.7775	0.9809	1.0430

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 Equation (5.1). For this reason, eight different weights, which are called membership degrees of observation, are located in Table 3.

Table 3. The weight related to observation for all methods

No	**LMS**	**Huber**	**Hampel**	**Tukey**	**Andrews**	The membership degrees of the observation to belong to the models in Equation (5.1)							
						w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8
1	1	1	1	0.9892	0.4710	0.2558	0.8399	0.1927	0.6327	0.2563	0.8714	0.1931	0.6341
2	1	1	1	0.9274	0.4632	0.1331	0.1553	0.6541	0.7634	0.1323	0.1544	0.6504	0.7590
3	1	1	1	0.9704	0.4713	0.4513	0.4688	0.2568	0.2667	0.4431	0.4603	0.2521	0.2619
4	1	1	1	0.9996	0.4752	0.2955	0.3919	0.2492	0.3305	0.3017	0.4001	0.2544	0.3374
5	1	1	1	0.9068	0.4545	0.2619	0.2287	0.4029	0.3518	0.2564	0.2239	0.3944	0.3444
6	1	1	1	0.9812	0.4702	0.9279	0.5080	0.4451	0.2437	0.9283	0.5082	0.4453	0.2438
7	1	0.9901	1	0.9383	0.4559	0.6285	0.9008	0.1891	0.2710	0.6283	0.9005	0.1890	0.2709
8	1	0.2780	0.1984	0	0.1023	0.1777	0.5464	0.0367	0.1128	0.1813	0.5573	0.0374	0.1150
9	1	1	1	0.9555	0.4685	0.1053	0.4404	0.1939	0.8109	0.1044	0.4365	0.1922	0.8037
10	1	1	1	0.9729	0.4692	0.5702	0.2784	0.6084	0.2971	0.5774	0.2820	0.6161	0.3009
11	1	1	1	0.9853	0.4732	0.1575	0.3079	0.4207	0.8225	0.1557	0.3045	0.4160	0.8134
12	1	1	1	0.9794	0.4728	0.9492	0.3440	0.4818	0.1746	0.9468	0.3431	0.4806	0.1742
13	1	1	1	0.9999	0.4758	0.1794	0.5488	0.2896	0.8859	0.1801	0.5510	0.2908	0.8895
14	1	1	1	0.9917	0.4754	0.1503	0.1419	0.5525	0.5216	0.1478	0.1396	0.5435	0.5132
15	1	0.0793	0	0	0	0.4992	0.5684	0.5843	0.6652	0.5004	0.5697	0.5856	0.6668
16	1	1	1	0.9371	0.4691	0.1247	0.5603	0.0919	0.4131	0.1267	0.5694	0.0934	0.4198
17	1	0.1292	0	0	0	0.5070	0.8905	0.1032	0.1812	0.5050	0.8869	0.1027	0.1804
18	1	1	1	0.8298	0.4345	0.1445	0.1493	0.2798	0.2891	0.1483	0.1532	0.2872	0.2967
19	1	0.5182	0.7229	0.6013	0.3816	0.5293	0.1817	0.6608	0.2268	0.5224	0.1793	0.6521	0.2238
20	1	1	1	0.9810	0.4742	0.3005	0.0669	0.1178	0.0262	0.3091	0.0688	0.1212	0.0270
21	1	1	1	0.9724	0.4718	0.3892	0.6510	0.4306	0.7203	0.3872	0.6476	0.4284	0.7165
22	1	0.5921	0.9567	0.7679	0.4197	0.2313	0.1485	0.9096	0.5841	0.2309	0.1483	0.9080	0.5831
23	1	1	1	0.8455	0.4425	0.5032	0.3235	0.4841	0.3113	0.4952	0.3184	0.4764	0.3063
24	1	1	1	0.9931	0.4747	0.2068	0.2081	0.1806	0.1817	0.2010	0.2022	0.1755	0.1766
25	1	1	1	0.9569	0.4682	0.4008	0.2502	0.8070	0.5038	0.3973	0.2480	0.7999	0.4994
26	1	1	1	0.9086	0.4589	0.8261	0.4589	0.5943	0.3301	0.8278	0.4598	0.5955	0.3307
27	1	1	1	0.9955	0.4762	0.4479	0.9437	0.1135	0.2367	0.4501	0.9393	0.1140	0.2379
28	1	0.9584	1	0.8559	0.4441	0.4073	0.1794	0.1907	0.0840	0.3971	0.1750	0.1860	0.0819
29	1	1	1	0.9882	0.4744	0.1948	0.1431	0.2575	0.1891	0.1894	0.1392	0.2504	0.1839
30	1	1	1	0.9931	0.4726	0.5378	0.7880	0.1901	0.2785	0.5323	0.7799	0.1882	0.2757

The residuals, which belong to estimates fromregression models in Equation (5.1) and belong to estimates for models from robust regression methods, are located in Table 4. The proposed algorithm was executed with a program written in MATLAB. In the stage of step operating, data sets have one dependent variables and this variable has an outlier observation.

Table 4. The residuals belong to observations for all methods

No	LMS Residual	Huber Residual	Hampel Residual	Tukey Residual	Andrews Residual	ANFIS Residual
1	3.4767	2.1904	1.5721	1.0752	1.6086	-17.6759
2	-3.9789	1.2178	2.5822	2.8061	2.5519	-7.8254
3	1.1051	-1.1466	-1.6369	-1.7818	-1.5673	-1.5879
4	-0.4020	0.5861	0.6348	0.2043	0.6978	-11.5606
5	-4.6338	-3.7318	-3.3337	-3.1872	-3.3084	-2.26604
6	2.9175	1.8044	1.6230	1.4196	1.7260	1.1031
7	7.6721	4.1178	3.0940	2.5835	3.1943	-7.6847
8	-10.3378	-14.6644	-16.1899	-17.1185	-16.0816	-28.9371
9	-2.5497	-1.8646	-1.9390	-2.1887	-1.9565	2.7615
10	-1.0641	1.1593	1.8002	1.7052	1.8737	-6.1937
11	-1.8004	0.6667	1.2379	1.2547	1.2210	-1.0513
12	-0.7286	-1.4388	-1.4064	-1.4838	-1.2970	1.2672
13	-0.9346	0.3685	0.4359	0.1152	0.4387	-13.3710
14	-4.3509	-0.4822	0.6489	0.9398	0.6329	-4.3802
15	50.3167	51.4430	51.6845	51.5151	51.7272	-10.2161
16	-0.3957	-1.2575	-1.9241	-2.6085	-1.8880	-29.6824
17	-26.3607	-31.5567	-33.0319	-33.6463	-32.9199	-9.3339
18	-0.4926	3.6793	4.5840	4.3540	4.6162	-11.3668
19	-9.4589	-7.8677	-7.1450	-6.9146	-7.0821	-5.1477
20	-2.1298	-1.4330	-1.1352	-1.4250	-0.9864	10.9992
21	-1.6218	-1.4706	-1.5166	-1.7185	-1.4846	-7.6727
22	-12.1691	-6.8861	-5.3992	-5.1320	-5.4017	-11.5501
23	3.8849	3.8895	4.0801	4.1399	4.1380	-3.2610
24	0.7363	-0.7023	-0.9030	-0.8580	-0.8578	0.0676
25	-1.6870	1.0962	1.9775	2.1547	2.0043	-7.9256
26	-3.4418	-3.1937	-3.0297	-3.1557	-2.9457	-2.9142
27	5.4420	1.2833	-0.0150	-0.6893	0.0864	-16.3709
28	6.6952	4.2542	3.9416	3.9915	4.0392	0.8268
29	0.4278	0.6231	0.9175	1.1227	0.9512	-1.3372
30	5.8638	2.2337	1.2525	0.8570	1.3398	-5.0809
Sum of Square Residual	3867.3	4087.8	4228.2	4278.2	4222.3	3580.9
Mean	128.9112	136.2609	140.9394	142.6050	140.7425	119.3650

The defined methods M (Huber, Hampel, Tukey, Andrews) were executed with programs written in MATLAB.The residuals of from the robust methods and LSM are large, but the residuals from the proposed algorithm based network are small. This is because, this method depend on fuzzy clustering.

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

6. Conclusion

In the study, we have proposed a method for obtaining optimal estimation values and compared various methods. Estimation values, which are obtained from the

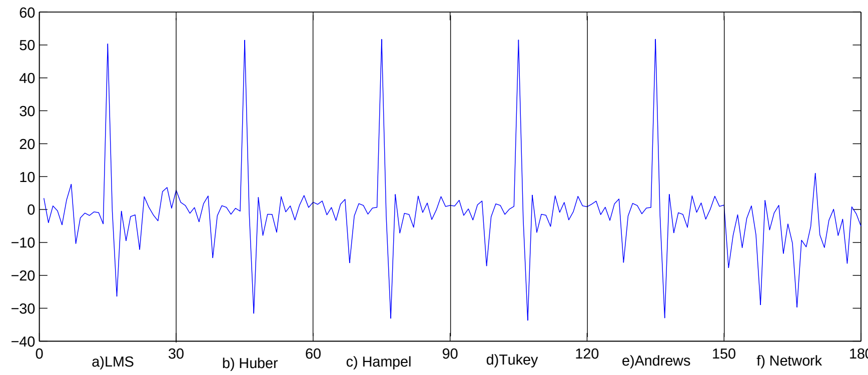

Figure 2. Graphs for errors related to data set in Table 1

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 estimations related to data are being used more frequently. 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. The difference between the obtained estimation 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 estimations 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 dependent variables. Finally, the estimation values obtained from the networks that are formed through the proposed algorithm are compared with the estimation values obtained from the robust regression methods. According to the indicated error criterion, the errors related to the estimations that are obtained from the network are lower than the errors that are obtained from the robust regression methods and LSM.The figures of errors obtained from the six methods are given in Figure 2. Figure 2(a) shows the errors related to the estimations that are obtained from the LSM, (b,c,d,e) are show the errors related to the estimations that are obtained from M Methods, and (f) shows the errors related to the estimations that are obtained from the proposed algorithm based ANFIS.

References

[1] Bezdek, C. J, Ehrlich. R. and Full, W. *FCM: The Fuzzy c Means Clustering Algorithm*, Computer and Geoscience **10**,191–203, 1984.

[2] Cichocki, A. and Unbehauen R. *Neural Networks for Optimization and Signal processing*, (John Wiley and Sons, New York, 1993).
[3] Cherkassky, V. and Muiler, F. *LearningFrom Data Concepts, Theory and Methods*,(John Wiley and Sons, New York, 1998).
[4] Chi-Bin, C. and Lee, E. S. *Applying Fuzzy Adaptive Network to Fuzzy Regression Analysis*, An International Journal Computers and Mathematics with Applications **38**,123–140, 1998.
[5] Chi-Bin, C. and Lee, E. S. *Switching Regression Analysis by Fuzzy Adaptive Network*, Europen Journal of Operational Research **128**, 647–663, 2001.
[6] Erbay, D.T. and Apaydın, A. *A Fuzzy Adaptive Network Approach to Parameter Estimation in case Where Independent Variables Come From Exponential Distribution*, Journal of Computational and Applied Mathematics **233**, 36–45, 2009. 1, 195–204, 1993.
[7] Hathaway,R.J. and Bezdek, J.C. *Switching Regression Models and Fuzzy Clustering*, IEEE Transactions on Fuzzy Systems **1**, 195–204, 1993.
[8] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel W. A. *Robust statistics.* (John-Willey and Sons, New-York, 1986).
[9] Hogg, R. V. *Statistican Robustness: One View of its use in Applications Today*, The American Statistican **33**, 108–115, 1979.
[10] Huber, P. J. *Robust statistics*, (John Willey and Son, 1981).
[11] Huynh, H. A. *Comparison of Approaches to Robust Regression*, Psychological Bulletin **92**, 505–512, 1982.
[12] Ishibuchi, H. and Nei, M. *Fuzzy Regression using Asymmetric Fuzzy Coefficients and Fuzzied Neural Networks*, Fuzzy Sets and Systems **119**, 273–290, 2001.
[13] Ishibuchi, H. and Tanaka, H. *Fuzzy Regression Analysis using Neural Networks*, Fuzzy Sets and Systems **50**, 257–265, 1992.
[14] Ishibuchi, H., Tanaka, H. and Okada, H. *An Architecture of Neural Networks with Interval Weights and its Application to Fuzzy Regression Analysis*, Fuzzy Sets and Systems **57**, 27–39, 1993.
[15] James, D. and Donalt, W. *Fuzzy Number Neural Networks*, Fuzzy Sets and Systems **108**, 49–58, 1999.
[16] Jang, J. R. *ANFIS: Adaptive-Network-Based Fuzzy Inference System*, IEEE Transaction on Systems, Man and Cybernetics **23**, 665–685, 1993.
[17] Kula, K. S. and Apaydın, A. *Fuzzy Robust Regression Analysis Based on The Ranking of Fuzzy Sets*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **16**, 663–681, 2008.
[18] Lung-Fei L. and Robert H. P. *Switching Regression Models With Imperfect Sample Separation Information-With an Application on Cartel Stability*,Econometrica **52**, 391–418, 1984.
[19] Michel M. *Fuzzy Clustering and Switching Regression Models Using Ambiguity and Distance Rejects*, Fuzzy Sets and Systems **122**, 363–399, 2001.
[20] Mu-Song C. and Wang S. W. *Fuzzy Clustering Analysis for Optimizing Fuzzy Membership Functions*, Fuzzy Sets and Systems **103**, 239–254, 1999.
[21] Richard E. Q. *A New Approach to Estimating Switching Regressions*, Journal of the American Statistical Association **67**, 306–310, 1972.
[22] Rousseeuw, P. J. and Leroy, A. M. *Robust regression and Outlier Detection*, (John Willey and Son, 1987).
[23] Takagi, T. and Sugeno, M. *Fuzzy Identification of Systems and Its Applications to Modeling and Control*, IEEE Trans. On Systems, Man and Cybernetics **15**, 116–132, 1985.
[24] Xie, X. L. and Beni, G. *A Validity Measure for Fuzzy Clustering*, IEEE Transactions on Pattern Analysis and Machine Intelligence **13**, 841–847, 1991.
[25] Xu, R. and Li, C. *Multidimensional Least Squares Fitting with a Fuzzy Model*, Fuzzy Sets and Systems **119**, 215–223, 2001.