Insurance: Mathematics and Economics 45 (2009) 236–241

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Neural networks approach for determining total claim amounts in insurance

Turkan Erbay Dalkilic [a], Fatih Tank [b,*], Kamile Sanli Kula [c]

[a] Karadeniz Technical University, Faculty of Arts and Sciences, Department of Statistics and Computer Sciences, 61080, Trabzon, Turkey
[b] Kirikkale University, Faculty of Arts and Sciences, Department of Statistics, 71100 Yahsihan-Kirikkale, Turkey
[c] Ahi Evran University, Faculty of Arts and Sciences, Department of Mathematics, 40200 Kirsehir, Turkey

ARTICLE INFO

Article history:
Received February 2009
Received in revised form
June 2009
Accepted 24 June 2009

Keywords:
Neural networks
Least squares method
Total claim amount
Claim amount payments
Fuzzy if-then rules

ABSTRACT

In this study, we present an approach based on neural networks, as an alternative to the ordinary least squares method, to describe the relation between the dependent and independent variables. It has been suggested to construct a model to describe the relation between dependent and independent variables as an alternative to the ordinary least squares method. A new model, which contains the month and number of payments, is proposed based on real data to determine total claim amounts in insurance as an alternative to the model suggested by Rousseeuw et al. (1984) [Rousseeuw, P., Daniels, B., Leroy, A., 1984. Applying robust regression to insurance. Insurance: Math. Econom. 3, 67–72] in view of an insurer.

1. Introduction

An insurance company is supposed to keep itself ready for all uncertain events such as claim amount demands from the insured. Consequently, in order to estimate future payments of claims, the insurance company sets up various models and checks their validity. Multiple linear regression is one of the most widely used statistical tools in practice. In actuarial statistics, situations occur that do not fit comfortably in such settings and they may generate some critical problems due to strong assumptions. Many problems can be prevented by using Generalized Linear Models, investigated by Kaas et al. (2008), instead of ordinary multiple linear regression. Many studies can be found as an alternative to multiple linear regression in the literature.

There are many studies on the use of the neural networks for parameter estimation. A fuzzy adaptive network approach was established for fuzzy regression analysis by Cheng and Lee (1999) and it was studied on both fuzzy adaptive networks and the switching regression model (Cheng and Lee, 2001). Jang (1993) studied adaptive networks based on a fuzzy inference system. In a study of Takagi and Sugeno (1985), the method for identifying a system using its input–output data was presented. James and Donalt (1999) studied fuzzy regression using neural networks.

There are different studies of fuzzy clustering and the validity criterion. In the study of Mu-Song and Wang (1999), the analysis of fuzzy clustering was done for determining fuzzy memberships, and in this study a method was suggested for indicating the optimal cluster numbers that belong to the variables. Xie and Beni (1991) suggested a validity criterion for fuzzy clustering. In this study we used the Xie–Beni validity criterion for determining optimal cluster numbers.

Various studies have used fuzzy clustering in insurance, such as Verrall and Yakoubov (2008), who specified a data-based procedure for grouping by age, using a fuzzy c-means algorithm. Ebanks et al. (1992) presented how to use the measures of fuzziness to risk classification for life insurance. The article by Horgby (1998) describes how to classify by using a fuzzy inference methodology instead of a risk classification according to the numerical rating system. In the study of Shapiro (2004), fuzzy clustering and the other fuzzy logic topics are discussed.

Detailed information about the historical development of neural networks, fuzzy logic and genetic algorithms and their useful application areas in insurance can be found in Shapiro (2002).

In this paper, we intend to highlight the importance of the neural networks approach to estimating total claim amount payments. The remainder of the paper is organized as follows. Section 2 introduces the parameter estimation in multiple linear regression. In Section 3, the fuzzy if-then rules and the use of these rules are introduced, using adaptive networks for analysis. In Section 4, an algorithm for parameter estimation using a neural network is given,

* Corresponding author. Tel.: +90 5322674950.
E-mail address: fatihtank@gmail.com (F. Tank).

0167-6687/$ – see front matter
doi:10.1016/j.insmatheco.2009.06.004

and in Section 5 a numerical study, based on real data, is investigated by using the algorithm suggested in Section 4. Finally, in Section 6, a discussion and conclusion are provided.

2. Multiple linear regression analysis

The main aim of regression analysis is to explain a dependent variable through its relationship with p independent variables. Consider a multiple linear regression model with p independent variables:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon. \quad (1)$$

In this sense the insurer wants to find linear relations between the dependent and independent variables in insurance. The observations, $y_1, y_2, \ldots, y_n$, recorded for each of these p levels can be expressed in the following way:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \cdots + \beta_p x_{p1} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \cdots + \beta_p x_{p2} + \varepsilon_2 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \cdots + \beta_p x_{pn} + \varepsilon_n. \end{aligned} \quad (2)$$

The system of n equations shown previously can be represented in matrix notation as follows:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (3)$$

The matrix $\mathbf{X}$ in Eq. (3) is referred to as the design matrix and it contains information about the levels of the predictor variables at which the observations are obtained. The vector $\boldsymbol{\beta}$ contains all the regression coefficients. To obtain the regression model, $\boldsymbol{\beta}$ should be known and can be estimated using least squares (LS) estimates. The following equation can be obtained:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{Y}. \quad (4)$$

Knowing the estimates, $\widehat{\boldsymbol{\beta}}$, the multiple linear regression model can be estimated, and the sum of squares error can be calculated as

$$\widehat{\mathbf{Y}}_{\mathrm{LS}} = \mathbf{X}\widehat{\boldsymbol{\beta}} \quad (5)$$

$$\widehat{\varepsilon}_{\mathrm{LS}} = \sum_{k=1}^{n} (y_k - \widehat{y}_k)^2 \quad (6)$$

respectively.

3. Fuzzy if-then rules and adaptive networks

The adaptive network used in estimating the unknown parameters of the regression model is based on fuzzy if-then rules and a 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 indicated as

$$\begin{aligned} R^K &: \text{IF } \left(x_1 = F_1^K, x_2 = F_2^K, \ldots, x_p = F_p^K\right) \\ &\quad \text{THEN } \left(Y = Y^K = c_0^K + c_1^K x_1 + \cdots + c_p^K x_p\right). \end{aligned} \quad (7)$$

Here F_i^K stands for fuzzy clusters or fuzzy terms associated with the input x_i in the Kth rule, and Y^K is the system output due to rule R^K (Cheng and Lee, 1999; Takagi and Sugeno, 1985). For instance, suppose a data set has two-dimensional input $X = [x_1, x_2]$. For input x_1, there are two fuzzy clusters, "small" and "large", and for input x_2, two fuzzy clusters, "light" and "heavy". In this case the fuzzy inference system contains the following four rules:

$$\begin{aligned} R^1 &: \text{IF } (x_1 \text{ is small and } x_2 \text{ is light}) \\ &\quad \text{THEN } (Y^1 = c_0^1 + c_1^1 x_1 + c_2^1 x_2) \end{aligned} \quad (8)$$

$$\begin{aligned} R^2 &: \text{IF } (x_1 \text{ is small and } x_2 \text{ is heavy}) \\ &\quad \text{THEN } (Y^2 = c_0^2 + c_1^2 x_1 + c_2^2 x_2) \end{aligned} \quad (9)$$

$$\begin{aligned} R^3 &: \text{IF } (x_1 \text{ is large and } x_2 \text{ is light}) \\ &\quad \text{THEN } (Y^3 = c_0^3 + c_1^3 x_1 + c_2^3 x_2) \end{aligned} \quad (10)$$

$$\begin{aligned} R^4 &: \text{IF } (x_1 \text{ is large and } x_2 \text{ is heavy}) \\ &\quad \text{THEN } (Y^4 = c_0^4 + c_1^4 x_1 + c_2^4 x_2). \end{aligned} \quad (11)$$

Here the multivariate input vector $X = \left[x_1, x_2, \ldots, x_p\right]$, univariate output vector Y and posterior parameter set c_i^K are crisp numbers, but the degrees of belonging to determined rules of the input vector X are fuzzy. The fuzzy system which is suitable for the fuzzy rules in Eqs. (8)–(11) is illustrated in Fig. 1, and the corresponding equivalent adaptive network architecture is shown in Fig. 2 (Mu-Song and Wang, 1999; Cheng and Lee, 1999; Shapiro, 2002).

There are two levels of nodes in Layer 1. The first level includes nodes "*small*" and "*large*" and the second level includes nodes "*light*" and "*heavy*". The output of the layer is the membership function based on the linguistic value of the input. Nodes in Layer 2 output the products w^L. The function of a node in this layer is to synthesize the information in the premise section of the fuzzy if-then rule; for example, the first node in Layer 2 includes IF (x_1 is small and x_2 is light), and each node output represents the firing strength of a rule. Layer 3 performs a normalization of the output signals from Layer 2. Each node in Layer 4 corresponds to the consequence of each fuzzy if-then rule; for example, the first node in Layer 3 includes THEN ($\widehat{Y}^1 = c_0^1 + c_1^1 x_1 + \cdots + c_p^1 x_p$). Finally, the single node in Layer 5 computes the overall output as the summation of all incoming signals, which is equivalent to performing an aggregation of all the fuzzy if-then rules.

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 clusters is indicated with this weighted mean.

A neural network enabling the use of a fuzzy inference system for fuzzy regression analysis is known as an adaptive network. Used for obtaining a good approach to regression functions and formed via neurals and connections, such an adaptive network consists of five layers (Hisao and Tanaka, 1992; Hisao and Manabu, 2001; Horia and Costel, 1996).

The nodes which form a network can be separated into two main groups: adaptive nodes and fixed nodes. The nodes which are forming Layer 1 and Layer 4 are called "adaptive nodes", since the nodes located in Layer 1 generate the membership degrees by Eq. (14) which have various values depending on the parameters $\{v_h, \sigma_h\}$. Similarly, the nodes located in Layer 4 generate the outputs, which have various values depending on the parameters c_i^K. In contrast, the nodes which form Layer 2, Layer 3 and Layer 5 do not include any parameter; thus they are depend on the numerical value from the previous layer, and are called "*fixed nodes*".

The fuzzy rule number of the system depends on the number of independent variables and the cluster or fuzzy sets number forming independent variables. There are significant numbers of validity criteria for fuzzy clusters in the literature. In this study, the fuzzy clustering validity function S, also called the Xie–Beni index, which is proposed by Xie and Beni (1991), will be used. This function will be explained in Section 4. When the independent variable number is indicated with p, if the cluster number

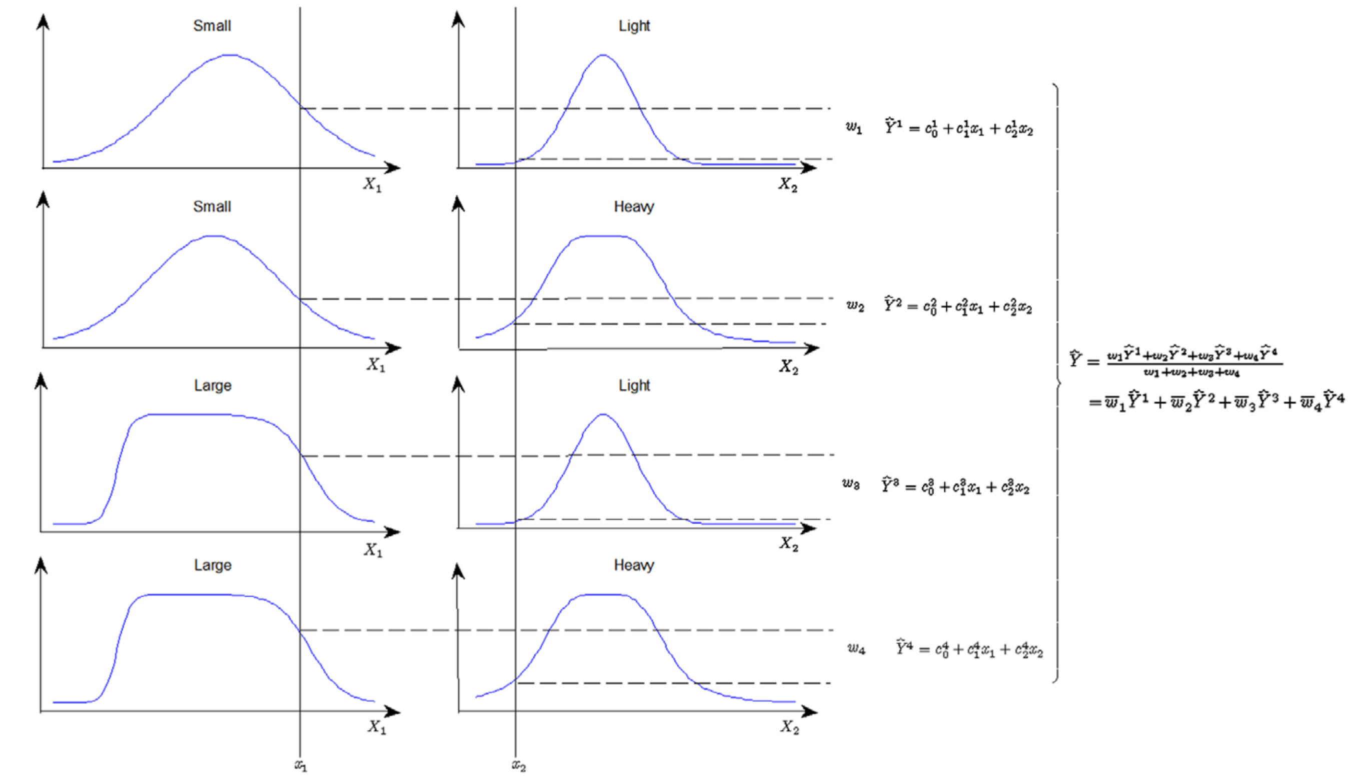

Fig. 1. Fuzzy inference system.

belonging to each variable is indicated with l_i $(i = 1, 2, \ldots, p)$, the fuzzy rule number is indicated with

$$L = \prod_{i=1}^{p} l_i. \tag{12}$$

The method of the determining the fuzzy rule number in Eq. (12) is conceptually based on the product rule in probability theory. For instance, if there are three input vectors in any system, and these inputs consist of two, three and two fuzzy clusters, respectively, then we can generate a $2 \times 3 \times 2 = 12$ fuzzy rule for this system.

The hth node in the first layer is adaptive, and is defined as

$$f_{1,h} = \mu_{F_h}(x_i), \quad h = 1, 2, \ldots, \quad \sum_{i=1}^{p} l_i, i = 1, 2, \ldots, p \tag{13}$$

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

$$\mu_{F_h}(x_i) = \exp\left[-\left(\frac{x_i - v_h}{\sigma_h}\right)^2\right] \tag{14}$$

because it is thought that the data which are used in numerical examples come from $N(v_h, \sigma_h)$, which is proved by using the Kolmogorov–Smirnov and Shapiro–Wilk tests with p-value $= 0.405 > 0.05$ and p-value $= 0.07 > 0.05$, respectively. The parameter set $\{v_h, \sigma_h\}$ indicates a priori parameters in this layer.

Each node in the second layer is a fixed node, and multiplication of the membership degrees, which are the outputs of the first layer, is labelled by Π in this layer. They have input signals coming from the first layer, and they are defined as multiplications of these input signals.

$$f_{2,K} = w^K = \prod_{i=1}^{p} \mu_{F_h}(x_i), \quad K = 1, 2, \ldots, L. \tag{15}$$

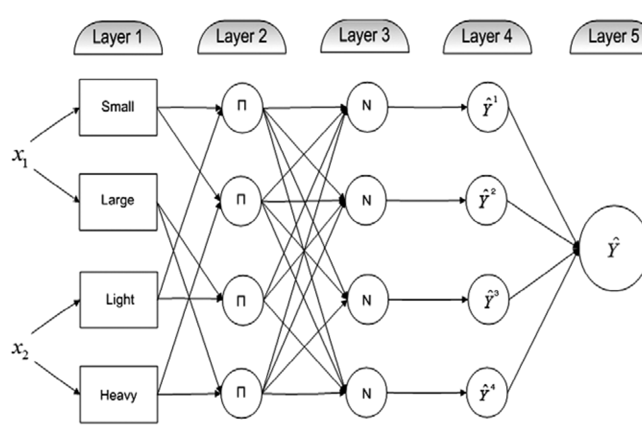

Fig. 2. The adaptive network architecture.

The nodes in the third layer are fixed nodes as well, and since the "Normalizations" of the outputs from the second layer are carried out in this layer, the nodes here labelled with N. The output of this layer is a normalization of the outputs of the second layer, and the node function is defined as

$$f_{3,K} = \overline{w}^K = \frac{w^K}{\sum_{K=1}^{L} w^K}. \tag{16}$$

The nodes in the fourth layer are adaptive nodes. The output signals of the fourth layer are also connected to a function, and this function is indicated with

$$f_{4,K} = \overline{w}^K Y^K, \quad K = 1, 2, \ldots, L \tag{17}$$

where $\widehat{Y}^K$ stands for the conclusion part of the fuzzy if-then rule as in Eq. (7). In the fifth layer, there is only one node; it is a fixed node, and it is calculated,

$$f_{5,1} = \widehat{Y} = \sum_{K=1}^{L} \overline{w}^K Y^K, \tag{18}$$

as the total of all signals (Cheng and Lee, 1999).

4. An algorithm for parameter estimation using a neural network

The process of determining the parameters for the regression model begins with determining the cluster numbers of the independent variables and a priori parameters. In this study, a validity criterion based on fuzzy clustering is used to determine the cluster numbers. A structure is formed which includes determining the a priori parameters as dependent to change the interval of the data set. In updating the a priori parameters, instead of the method in which spreading back the error brings about a series of transaction, a process has been formed which enables one to review all the values so that parameter might have the lowest error. The algorithm related to the proposed method for determining the regression model in the case of independent variables coming from a normal distribution is defined as follows.

Step 0. Optimal cluster numbers related to the data set associated with the independent variables are determined. The optimal value of the cluster number l_i, ($l_i = 2, 3, \ldots, \max$) can be obtained by minimizing the fuzzy clustering validity function S_k. This function is expressed by (Xie and Beni, 1991)

$$S_k = \frac{\frac{1}{n}\sum_{k=1}^{l_i}\sum_{j=1}^{n}\left(\mu_{kj}\right)^m \left\|v_k - x_j\right\|^2}{\min_{k\neq j}\left\|v_k - v_j\right\|^2}. \tag{19}$$

Note that $\|.\|$ is the usual Euclidean norm. Here, μ_{kj} is the fuzzy membership degree of the jth observation belonging to the kth cluster, and $m > 1$ is the fuzziness index (or weighting exponent). At the time of the first suggested use of validity criteria, there did not exist any theoretical method for obtaining the m parameter, which is used for determination of the fuzzy c-means algorithm and the optimal class number for fuzzy clustering. Bezdek et al. (1984) and James and Donalt (1999) suggested the $(1.5 \leq m \leq 3.0)$ interval for the optimal value of m and showed that this interval gave good results. Furthermore, Zahid et al. (1999) suggested that $m = 2$ is often the most useful value for many users of the fuzzy c-means algorithm. Recently, many articles can be found with recommendations for the optimal selection of m (Okeke and Karnieli, 2006; Gao et al., 2000). In new studies it is seen that the optimal value of m is $[1.5 \leq m \leq 2.5]$. In this study we preferred to use $m = 2$, which can be found in the literature, and gives good results, according to the suggestion of Xie and Beni (1991).

As can be seen, cluster centers ($k = 2, 3, \ldots, l_i$) which are well separated (the norm between the centers is high) produce a high value of separation such that a smaller S_k value is obtained because this value is located in the denominator of Eq. (19). Thus, our goal is to find the optimal fuzzy l_i-partition with the smallest S. The cluster number (l_i) value is defined as the optimal cluster number with the smallest S_k when the smallest S_k value is observed.

Step 1. An a priori parameter set is determined. The a priori parameters that display the cluster centers, which belong to the independent variables, depend on the difference between the maximum and minimum values and fuzzy cluster numbers of the independent variables (Dalkilic and Apaydin, 2008). This is indicated by

$$v_i = \min(X_i) + \frac{\max(X_i) - \min(X_i)}{(l_i - 1)}(i-1), \quad i = 1, 2, \ldots, p. \tag{20}$$

Step 2. $\overline{w}^K$ weights are calculated which are used to form matrix **B** to be used in counting the a posteriori parameter set by Eq. (16). They depend on the membership function related to the distribution family which the independent variable belongs to. The node functions in the first layer of adaptive networks are defined by Eq. (13). For F_h there may be several functions that are appropriate membership functions. The observations come from the normal distribution and each cluster of observations has an a posteriori parameter set $\{v_h, \sigma_h\}$, so here, the fuzzy membership function, which is defined in Eq. (14), is used to determine the membership degrees of belonging to fuzzy clusters of each observations.

Let w^K indicate the weights. The weights are obtained via mutual multiplication of membership degrees at an amount depending on the number of independent variables, and the fuzzy cluster members of these variables are determined by Eq. (15). For example, let us assume that variable x_1 has two levels, which are called the first level and the second level, and similarly variable x_2 has two levels, which are called the third level and the fourth level, in a fuzzy inference system. In this situation, if any observation belongs to the first level for variable x_1 with $\mu_{F_1}(x_1)$ membership and belongs to the third level for variable x_2 with $\mu_{F_3}(x_2)$ membership then the weight value of the first rule is calculated by

$$w_1 = \mu_{F_1}(x_1)\,\mu_{F_3}(x_2). \tag{21}$$

The weights of other rules are calculated in the same way. The $\overline{w}^K$ weights are normalization of the weights defined as w^K and they are calculated by Eq. (16).

Step 3. In the condition that the independent variables are fuzzy and the dependent variables are crisp, the a posteriori parameter set $c_i^K = \left(a_i^K, b_i^K\right)$ is obtained as crisp numbers in the shape of $c_i^K = a_i^K$. When c_i^K is a triangular fuzzy number, a_i^K is the center and b_i^K is the spread of c_i^K. In the condition that $c_i^K = a_i^K$, the equality

$$\mathbf{Z} = \left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{Y} \tag{22}$$

is used by Cheng and Lee (2001). Here, **B** is the data matrix which is weighted by membership degree and its dimension is $[(p+1)\times L\times n]$, defined as

$$\mathbf{B} = \begin{bmatrix} \overline{w}_1^1 & \cdots & \overline{w}_1^L & \overline{w}_1^1 x_{11} & \cdots & \overline{w}_1^L x_{11} & \cdots & \overline{w}_n^1 x_{p1} & \cdots & \overline{w}_n^L x_{p1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \overline{w}_k^1 x_{jk} & \vdots & \ddots & \vdots \\ \overline{w}_n^1 & \cdots & \overline{w}_n^L & \overline{w}_n^1 x_{1n} & \cdots & \overline{w}_n^L x_{1n} & \cdots & \overline{w}_n^1 x_{pn} & \cdots & \overline{w}_n^L x_{pn} \end{bmatrix} \tag{23}$$

where **Y** is a dependent variable vector and **Z** is an a posteriori parameter vector which is defined as

$$\mathbf{Z} = \left[a_0^1, \ldots, a_0^L, a_1^1, \ldots, a_1^L, a_p^1, \ldots, a_p^L\right]^T. \tag{24}$$

Step 4. By using the a posteriori parameter set $c_i^K = \left(a_i^K\right)$ obtained in Step 3, the regression model indicated by

$$\widehat{Y}^K = c_0^K + c_1^K x_1 + c_2^K x_2 + \cdots + c_p^K x_p, \quad K = 1, 2, \ldots, L \tag{25}$$

is constituted. Setting out from the models and weights specified in Step 2, the prediction values are obtained using a neural network (NN) defined as

$$\widehat{Y}_{\mathrm{NN}} = \sum_{K=1}^{L} \overline{w}^K Y^K. \tag{26}$$

Step 5. The error related to the model which is obtained from the proposed algorithm depends on NN, and is calculated as

$$\widehat{\varepsilon}_{NN} = \sum_{k=1}^{n} (y_k - \widehat{y}_k)^2 . \quad (27)$$

If $\widehat{\varepsilon}_{NN} < \phi$, then the a posteriori parameters have been obtained as parameters of the regression models to be formed, and the process is determined; else Step 6 begins. Here ϕ is a low stable value determined by the decision maker, y_k is kth observed outcome and $\widehat{y}_k$ is the kth predicted network output of the input vector.

Step 6. The central a priori parameters, specified in Step 1, are updated with

$$v_i' = v_i \pm t \quad (28)$$

in a manner that they increase from the lowest value to the highest and decrease from the highest value to the lowest. Here t is the size of the steps, and it can be calculated as

$$t = \frac{\max(x_{ki}) - \min(x_{ki})}{a}, \quad k = 1, 2, \ldots, n, i = 1, 2, \ldots, p; \quad (29)$$

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

Step 7. The estimations for each a priori parameter obtained by a change and error criterion related to these predictions are calculated. The lowest of error criterion is defined. The a priori parameters, which give the lowest error, are specified, and the prediction obtained via the models related to these parameters is given as output.

In the proposed algorithm, the Sugeno fuzzy inference system is used to form the fuzzy rule. The use of the membership function, which is given in Eq. (14), in the formation of the fuzzy rule minimizes the errors associated with the predictions. Using the program, which was coded in the MATLAB for the proposed algorithm for NN, the value of the prime parameters can change minimally, and the prediction errors can be calculated with the use of these values. Consequently, the optimal value of prime parameters can be selected by this program (Dalkilic and Apaydin, 2008).

5. Numerical study

For our example, data were collected from a well-known insurance company in Turkey. In the model, X_1 represents the number of the calendar month, X_2 represents the claim numbers and Y represents the total claim payments (million Turkish Lira) in the related calendar month.

Applying least squares, the estimated regression model yields

$$\widehat{Y}_{LS} = -1.1836 + 0.1204X_1 + 0.00191X_2 \quad (30)$$

and according to the results of the residual analysis, the eighth observation is an outlier. We preferred the multiple linear regression model due to its reasonable coefficient of determination R^2.

In the prediction process by the neural network, the fuzzy cluster numbers for each variable are defined as 2 in the initial step of the proposed algorithm. The number of fuzzy inference rules to be formed depending on these cluster numbers is obtained as

$$L = \prod_{i=1}^{p=2} l_i = l_1 \times l_2 = 4. \quad (31)$$

The models obtained via the four fuzzy inference rules are

$$\widehat{Y}_1 = 7768 + 428x_1 - x_2 \quad (32)$$

$$\widehat{Y}_2 = 372 + 24x_1 \quad (33)$$

$$\widehat{Y}_3 = -15\,057 + 397x_1 + x_2 \quad (34)$$

$$\widehat{Y}_4 = -1995 + 45x_1. \quad (35)$$

Table 1
Results of models.

X_1	X_2	Y	$\widehat{Y}_{LS}$	$\widehat{Y}_{NN}$	$\widehat{\varepsilon}_{LS}$	$\widehat{\varepsilon}_{NN}$
1	1270	1.25	1.3847	1.2504	−0.1347	0.0004
2	2630	3.87	4.1264	3.8593	−0.2564	−0.0107
3	3653	5.89	6.2186	5.8826	−0.3286	−0.0074
4	3045	5.91	5.1671	5.9758	0.7429	0.0658
5	3232	6.09	5.6480	6.0383	0.4420	−0.0517
6	3681	6.54	6.6338	6.5660	−0.0938	0.0260
7	3169	6.31	5.7674	6.2465	0.5426	−0.0635
8	3448	5.45	6.4255	5.4473	−0.9755	−0.0027
9	3163	5.83	5.9966	5.9166	−0.1666	0.0866
10	3096	6.06	5.9879	6.0242	0.0721	−0.0358
11	3795	7.53	7.4556	7.5236	0.0744	−0.0064
12	4481	8.98	8.8982	8.9745	0.0818	−0.0055
Sum of squares error					2.2392	0.0207

The parameters of these models are obtained by solving Eq. (24).

The prediction values which are dependent on Eqs. (32)–(35) are calculated by using Eq. (18), and these values are shown in Table 1.

The predictions that are obtained with LS, and the errors related with these predictions, are also displayed in Table 1. As Table 1 shows, there is a significant difference between least squares methods the and neural networks approach in terms of sum of squares error, which is crucial for estimating total claim amounts payments.

6. Conclusions

This paper has given an account of and the reasons for the use of a neural networks approach instead of a least squares method. A new model, which consists of the number of months and the number of claims in each month, is suggested for estimating claim amount payments by an insurer. According to the solution, the sum of squares errors are 2.239 and 0.0207 for the least squares method and the suggested model, respectively. In conclusion, solution of the suggested model gives much better results by the neural networks approach, because each set of observations in the NN approach with the fuzzy rules for the membership degrees is included in the model. Thus, the contribution of the outlier value to the models is limited by membership degrees and possible negative effects on the model may be minimized. Despite this, the effect of the outlier value to the model with other observations is much more than NN approach. As a result, in our numerical examples the NN approach is not affected by an outlier value in the data set and smaller estimates for errors have been obtained, compared to the LS approach. The findings of this study have a number of important implications for future practice in estimating claim amount payments, especially in the case of at least one outlier existing, and they also have important implications for developing a new model that is not affected by outliers.

Acknowledgements

First and foremost the authors offer their sincerest gratitude to their Ph.D. supervisors, Prof. Dr. Aysen Apaydin (Ankara University, Turkey) and Prof. Dr. Omer L. Gebizlioglu (Ankara University, Turkey), who have continuously provided support for them throughout their studies and academic life. The authors also express their thanks to the referees for their very constructive comments and guidance.

References

Bezdek, J.C., Ehrlich, R., Full, W., 1984. FCM: The fuzzy *c*-means clustering algorithm. Computers & Geosciences 10, 191–203.

Cheng, C.-B., Lee, E.S., 1999. Applying fuzzy adaptive network to fuzzy regression analysis. An International Journal Computers & Mathematics with Applications 38, 123–140.
Cheng, C.-B., Lee, E.S., 2001. Switching regression analysis by fuzzy adaptive network. European Journal of Operational Research 128, 647–663.
Dalkilic, T.E., Apaydin, A., 2008. A fuzzy adaptive network approach to parameter estimation in cases where independent variables come from an exponential distribution. Journal of Computational and Applied Mathematics doi:10.1016/j.cam.2008.07.057.
Ebanks, B., Karwowskiz, W., Ostaszewski, K., 1992. Application of measures of fuzziness to risk classıfication in insurance. In: Proceedings of the Fourth International Conference on Computing and Information. IEEE Computer Society Press, Los Alamitos, California, pp. 290–291.
Gao, X., LI, J., Xie, W., 2000. Parameter optimization in FCM clustering algorithms. In: Proceedings of ICSP2000, pp. 1457–1461.
Hisao, I.I., Manabu, N., 2001. Fuzzy regression using asymmetric fuzzy coefficients and fuzzied neural networks. Fuzzy Sets and Systems 119, 273–290.
Hisao, I.I., Tanaka, H., 1992. Fuzzy regression analysis using neural networks. Fuzzy Sets and Systems 50, 257–265.
Horgby, P.J., 1998. Risk classification by fuzzy inference. The Geneva Papers on Risk and Insurance Theory 23, 63–82.
Horia, F., Costel, S., 1996. A new fuzzy regression algorithm. Analytical Chemistry 68, 771–778.
James, D., Donalt, W., 1999. Fuzzy number neural networks. Fuzzy Sets and Systems 108, 49–58.
Jang, J.R., 1993. ANFIS: Adaptive-network-based fuzzy inference system. IEEE Transaction on Systems, Man and Cybernetics 23, 665–685.
Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M., 2008. Modern Actuarial Risk Theory Using R. Springer-Verlag, Berlin Heidelberg.
Mu-Song, C., Wang, S.W., 1999. Fuzzy clustering analysis for optimizing fuzzy membership functions. Fuzzy Sets and Systems 103, 239–254.
Okeke, F., Karnieli, A., 2006. Linear mixture model approach for selecting fuzzy exponent value in fuzzy c-means algorithm. Ecologicalinformatics 1, 117–124.
Rousseeuw, P., Daniels, B., Leroy, A., 1984. Applying robust regression to insurance. Insurance: Mathematics and Economics 3, 67–72.
Shapiro, A.F., 2002. The merging of neural networks, fuzzy logic, and genetic algorithms. Insurance Mathematics and Economics 31, 115–131.
Shapiro, A.F., 2004. Fuzzy logic in insurance: First 20 years. Insurance: Mathematics and Economics 35, 399–424.
Takagi, T., Sugeno, M., 1985. Fuzzy identification of systems and its applications to modeling and control. IEEE Transactions on Systems, Man and Cybernetics 15, 116–132.
Verrall, R., Yakoubov, Y.H., 2008. A fuzzy approach to grouping by policyholder age in general insurance. Actuarial Research Paper no. 104, Department of Actuarial Science and Statistics, City University, London.
Xie, X.L., Beni, G., 1991. A validity measure for fuzzy clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence 13, 841–847.
Zahid, N., Limouri, M., Essaid, A., 1999. A new cluster validity for fuzzy clustering. Pattern Recognition 32, 1089–1097.