



Type 1 fuzzy function approach based on ridge regression for forecasting

Eren Bas¹ · Erol Egrioglu¹ · Ufuk Yolcu² · Crina Grosan^{3,4}

Received: 20 February 2018 / Accepted: 10 July 2018 / Published online: 19 July 2018
© Springer Nature Switzerland AG 2018

Abstract

Fuzzy function approach is a kind of fuzzy inference system that can produce successful results for the analysis of forecasting problems. In a fuzzy function approach, a fuzzy function corresponding to each fuzzy set is generated using multiple regression analysis. The number of explanatory variables in multiple regression analysis is increased via the non-linear transformations of the membership functions to improve the prediction performance of the model. In a fuzzy function approach, it can be found a high correlation between the non-linear transformations of membership functions, and therefore, the multiple linear regression method used to define fuzzy functions which has multicollinearity problem. The contribution of this paper is to propose a new fuzzy forecasting method to overcome this problem. In this paper, a new fuzzy function approach using ridge regression instead of multiple linear regression in Type 1 fuzzy function approach is proposed. The proposed new Type 1 approach is applied to various real world time series data and the results are compared to the ones obtained from other techniques. Thus, it is concluded that the results present superior forecasts performance.

Keywords Type 1 fuzzy function approach · Ridge regression · Fuzzy inference system · Forecasting · Fuzzy c-means

1 Introduction

Future planning has become very important in the world we live in. It is vital to obtain accurate and reliable forecasts in making plans for the future. Fuzzy inference systems have been also added to the methods used to obtain predictions in recent years. Fuzzy inference systems work based on fuzzy sets created for input and output dataset on hand. In the literature, fuzzy inference system proposed by Takagi and Sugeno (1985) and adaptive network fuzzy inference

system (ANFIS) proposed by Jang (1993) have been commonly used for forecasting problem. Both of these fuzzy inference systems work based on a rule base. The determination of these rules requires expert opinion and this situation prevents the objective working of the inference system.

Therefore, fuzzy function approach was proposed by Turksen (2008). Fuzzy function approach does not need rule base using fuzzy functions instead of rules. The different applications of fuzzy function approach were developed by Celikyilmaz and Turksen (2008a, b) and Turksen (2009). The detailed information about fuzzy functions can be found in Celikyilmaz and Turksen (2009). And also, a hybrid fuzzy function approach was proposed by Zarandi et al. (2013). The fuzzy function approach for forecasting problem was also reconsidered in the study of Beyhan and Alci (2010) and Aladag et al. (2014, 2016). In recent years, Granular computing has been an emerging computing paradigm of information processing. Granular computing is more a theoretical perspective than a set of methods. Granular computing approaches are data analysis techniques like fuzzy methods and these methods try to recognize various levels of scales. The recent contributions of granular computing are given by Pedrycz and Chen (2011, 2015a, b) and Maciel et al. (2016).

✉ Eren Bas
eren.bas@giresun.edu.tr

¹ Forecast Research Laboratory, Department of Statistics, Faculty of Arts and Science, Giresun University, 28200 Giresun, Turkey

² Forecast Research Laboratory, Department of Econometrics, Faculty of Economics and Administrative Sciences, Giresun University, 28200 Giresun, Turkey

³ Department of Computer Science, College of Engineering Design and Physical Sciences, Brunel University London, London UB83PH, UK

⁴ Department of Computer Science, Babes-Bolyai University, Cluj-Napoca, Romania

Fuzzy function approach can produce successful results for prediction problems. In a fuzzy function approach, while obtaining fuzzy functions, the input matrix is extended with the use of various nonlinear transformations of memberships. It should not be a significant linear relationship between the explanatory variables in multiple regression method used in the obtaining of fuzzy functions. Linear relationship between the explanatory variables in multiple regression leads to multicollinearity problem and this situation causes to increase the variance of estimators and to obtain inconsistent prediction results. It is clearly seen that the nonlinear functions of the memberships used to obtain fuzzy functions are related between each other.

As a result, Type 1 fuzzy function (T1FF) approach work based on regression models with multicollinearity problem. Ridge regression technique can be used in the presence of multicollinearity problem instead of classic regression analysis. It can be obtained biased but small variance estimators using shrinkage parameter in ridge regression. Ridge regression is a remedy used in the presence of multicollinearity problem and it was first proposed by Hoerl and Kennard (1970). Ridge regression method has two important advantages according to the ordinary least squares (OLS) method. One of them is to solve the multicollinearity problem and the other one is to decrease the mean square error (MSE) of predictors. A simple formula to obtain shrinkage parameter in ridge regression was proposed by Hoerl et al. (1975). The motivation of this paper is to try ridge regression method to eliminate multicollinearity problem in T1FF approach.

In this paper, the multicollinearity problem in T1FF approach is solved using ridge regression technique and this is the contribution of the paper. The proposed new T1FF approach is called as “Type 1 fuzzy function approach based on ridge regression (T1FFRR)”. The proposed new T1FFRR approach was applied to many real world time series datasets and the results are compared to the ones obtained from other techniques.

The rest of the paper can be outlined as below: The second section of the paper is about T1FF approach. Ridge regression technique is briefly summarized in Sect. 3. In

$$X^{(i)} = \begin{bmatrix} \mu_{i1} & \mu_{i1}^2 & \exp(\mu_{i1}) & \ln((1 - \mu_{i1})/\mu_{i1}) & x_{11} & \cdots & x_{p1} \\ \mu_{i2} & \mu_{i2}^2 & \exp(\mu_{i2}) & \ln((1 - \mu_{i2})/\mu_{i2}) & x_{12} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mu_{in} & \mu_{in}^2 & \exp(\mu_{in}) & \ln((1 - \mu_{in})/\mu_{in}) & x_{1n} & \cdots & x_{pn} \end{bmatrix}, \quad i = 1, 2, \dots, c \tag{3}$$

the fourth section of the paper, the proposed new T1FFRR approach was introduced with details. Section 5 presents the results from the application of the proposed method to real life datasets and finally Sect. 6 presents conclusions and discussions.

2 Type 1 fuzzy function approach

Fuzzy inference system proposed by Takagi and Sugeno (1985) and ANFIS are fuzzy inference systems that require the creation of rule base. Fuzzy function approach proposed by Turksen (2008) is a fuzzy inference system working without rules. In a fuzzy function approach, a linear function is composed via linear regression method for each fuzzy set obtained from fuzzy c-means and the output of the system is obtained from the output of weighted fuzzy function predictions with memberships as weights. The fuzzy function approach for forecasting problem was also reconsidered in the study of Beyhan and Alci (2010), Aladag et al. (2014, 2016). Four different fuzzy function approaches were given in Celikyilmaz and Turksen (2009). The algorithm of T1FF method is given in Algorithm 1.

Algorithm 1. T1FF Approach.

Step 1 Generate a matrix of lagged variables of time series.

Step 2 Determine the degrees of belonging to the fuzzy sets.

The matrix consisting of lagged variables is clustered using FCM technique, and thus the membership values ($\mu_{ik} \ i = 1, 2, \dots, c; k = 1, 2, \dots, n$). Where c and n represent # fuzzy sets and crisp observations.

Step 3 Constitute the fuzzy regression functions.

The fuzzy regression functions for each fuzzy set are constituted and they can be expressed as given in Eq. (1).

$$Y^{(i)} = X^{(i)}\beta^{(i)} + \epsilon^{(i)}, \quad i = 1, 2, \dots, c \tag{1}$$

Celikyilmaz and Turksen (2009) stated that various logarithmic and exponential transformations of membership values may increase the performance of systems. In Eqs. (2) and (3) represents the targeted outputs and the inputs of the system.

$$Y^{(i)} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad i = 1, 2, \dots, c \tag{2}$$

where p represents # crisp inputs.

Step 4 Estimate the fuzzy regression functions.

The fuzzy regression functions for each fuzzy set are estimated by OLS method.

$$\hat{Y}^{(i)} = X^{(i)}\hat{\beta}^{(i)}, \quad i = 1, 2, \dots, c \tag{4}$$

Step 5 Obtain the outputs.

The outputs obtained by each estimated fuzzy regression function are weighed in proportion to corresponding membership values and the final outputs are calculated by Eq. (5).

$$\hat{y}_k = \frac{\sum_{i=1}^c \hat{y}_{ik} \mu_{ik}}{\sum_{i=1}^c \mu_{ik}}, \quad i = 1, 2, \dots, c; \quad k = 1, 2, \dots, n \quad (5)$$

In a fuzzy function approach, fuzzy function is obtained as the number of clusters and requires fewer parameters than other fuzzy inference systems. The non-linear relationship between input and output can be modeled on the grounds that fuzzy function approach is an approach based on data. The most important issue in fuzzy function approach is that it has multicollinearity problem and also there is no other solution method in the literature on this subject. In this study, T1FF approach has been modified by solving multicollinearity problem.

3 Ridge regression

Ridge regression is a remedy used in the presence of multicollinearity problem and it was first proposed by Hoerl and Kennard (1970). Ridge regression method has two important advantages according to the OLS method. One of them is to solve the multicollinearity problem and the other one is to decrease the mean square error (MSE). The solution technique of ridge regression is similar with OLS. Besides, the difference between ridge regression and OLS is the k value. This k value is also called as biased parameter or shrinkage parameter and it takes values between 0 and 1. This k value is added to the diagonal elements of the correlation matrix and thus biased regression coefficients are obtained. That means ridge regression is a biased regression method. There are several methods of determining multicollinearity problem. Some graphs and statistics can be used for this aim. Variance inflation factors (VIF), eigenvalues of correlation matrix of the independent variables and signs of the regression coefficients can demonstrate signals of multicollinearity problem.

The OLS estimates of regression coefficients and ridge estimates of regression coefficients are shown in Eqs. (6) and (7) respectively.

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (6)$$

$$\hat{\beta}_R = (X'X + kI)^{-1} X'Y \quad (7)$$

As noted above, ridge regression is a biased regression method. The proof of this situation is shown in Eq. (8).

$$\begin{aligned} \hat{\beta}_R &= (X'X + kI)^{-1} X'Y \\ &= (X'X + kI)^{-1} X'X \hat{\beta} = Z \hat{\beta} \end{aligned} \quad (8)$$

$$E(\hat{\beta}_R) = E(Z \hat{\beta}) = Z \beta$$

It is clearly seen that ridge estimates of regression coefficients ($\hat{\beta}_R$) are biased estimates. One of the most important points to be considered in the ridge regression is the k value. There are many methods proposed in the literature to find the optimal k value. Ridge trace is one of these methods. Ridge trace is a plot of the elements of the ridge estimator versus k usually in the interval (0, 1). The other method in the literature used to find the optimal k value is given in Eq. (9) and it was proposed in Hoerl et al. (1975).

$$k = \frac{p \hat{\sigma}^2}{\hat{\beta}' \hat{\beta}} \quad (9)$$

The other method used to determine the effects of multicollinearity problem is VIF. The diagonal elements of $\text{Var}(\hat{\beta})$ is called as **VIF** and is given by Eq. (10).

$$\text{VIF}_j = \frac{1}{(1 - R_j^2)} \quad j = 1, \dots, p \quad (10)$$

In Eq. (10), R_j^2 is the determination coefficient obtained from the multiple regression of X_j on the remaining $(p - 1)$ regressor variables in the model. It can be said that there is a multicollinearity problem among the independent variables if these VIF values increase (VIF values ≥ 10).

4 The proposed method

In T1FF approach, ridge regression technique can be used to eliminate the multicollinearity problem that arises from the relationship between non-linear transformations of the membership values. The forecasting performance will be increased in a fuzzy function approach which does not have multicollinearity problem. The main difference of proposed method from T1FF method is the usage of ridge regression method for obtaining fuzzy functions. In the ridge regression the k value is determined using Eq. (9). T1FFRR approach is given in Algorithm 2.

Algorithm 2: The Proposed T1FFRR Method.

Step 1. The inputs of the system are lagged variables (number of p).

The model order (p) is determined according to the structure of the time series. Time series is shown as column vectors in the form of $X_t = [x_1, x_2, \dots, x_n]'$. The matrix Z

is composed of the input and outputs of the system and its dimension is $(n - p) \times (p + 1)$.

$$Z = [X_t \ X_{t-1} \ X_{t-2} \ \dots \ X_{t-p}] \tag{11}$$

The elements of the matrix are clustered using fuzzy clustering method (FCM) technique proposed by Bezdek (1981). FCM clustering technique is applied iteratively using the following Eqs. (12) and (13). In these equations, $c, v_i (i = 1, 2, \dots, c)$ and $\mu_{ik} (i = 1, 2, \dots, c; k = 1, 2, \dots, n)$ represent the number of fuzzy sets, cluster centers and membership values, respectively.

$$v_i = \frac{\sum_{k=1}^n (\mu_{ik})^f z_k}{\sum_{k=1}^n (\mu_{ik})^f}, \quad i = 1, 2, \dots, c \tag{12}$$

$$\mu_{ik} = \left[\sum_{j=1}^c \left(\frac{d(z_k, v_i)}{d(z_k, v_j)} \right)^{\frac{2}{f-1}} \right]^{-1}, \quad i = 1, 2, \dots, c; k = 1, 2, \dots, n \tag{13}$$

where f is the degree of fuzziness, z_k is a vector whose elements are the elements that compose of k^{th} row of Z . μ_{ik} is the degree of belongingness of k^{th} observation to i^{th} cluster and $d(z, v)$ is Euclidian distance and is computed using Eq. (14).

$$d(z_k, v_i) = z_k - v_i \tag{14}$$

Step 2 The membership values are determined using the Eq. (13) for the input dataset according to the center determined from FCM technique.

$$\mu_{ik} = \left[\sum_{j=1}^n \left(\frac{d(x_k, v_i)}{d(x_k, v_j)} \right)^{\frac{2}{f-1}} \right]^{-1}, \quad i = 1, 2, \dots, c; k = 1, 2, \dots, n \tag{15}$$

where x_k is a row of the input matrix which is generated for lagged variables and it is a vector whose elements are the elements that compose of k^{th} row and this value is taken as zero when $\mu_{ik} \leq \alpha$. f_i is fuzziness index.

$$X = [X_{t-1} \ X_{t-2} \ \dots \ X_{t-p}] \tag{16}$$

Step 3 For each cluster i , the membership values of each input data sample (μ_{ik}), non-linear transformations of membership values and original inputs are used as explanatory variables and i th fuzzy function is obtained from predicting $Y^{(i)} = X^{(i)}\beta^{(i)} + \epsilon^{(i)}$ multiple regression model.

$X^{(i)}$ and $Y^{(i)}$ matrices are as follows when the number of lagged variables is p and using transformations $\mu_{i1}^2, \exp(\mu_{i1})$ and $\ln((1 - \mu_{i1})/\mu_{i1})$ for non-linear transformation of the membership values.

$$X^{(i)} = \begin{bmatrix} \mu_{i1} & \mu_{i1}^2 & \exp(\mu_{i1}) & \ln((1 - \mu_{i1})/\mu_{i1}) & x_1 & \dots & x_p \\ \mu_{i2} & \mu_{i2}^2 & \exp(\mu_{i2}) & \ln((1 - \mu_{i2})/\mu_{i2}) & x_2 & \dots & x_{p+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ \mu_{in} & \mu_{in}^2 & \exp(\mu_{in}) & \ln((1 - \mu_{in})/\mu_{in}) & x_{n-p} & \dots & x_{n-1} \end{bmatrix} \tag{17}$$

$$Y^{(i)} = \begin{bmatrix} x_{p+1} \\ x_{p+2} \\ \vdots \\ x_n \end{bmatrix} \tag{18}$$

Maximum likelihood estimators of regression parameters are obtained as follows.

$$\hat{\beta}_R^{(i)} = (X^{(i)'} X^{(i)} + kI)^{-1} X^{(i)'} Y^{(i)} \tag{19}$$

$$\hat{Y}^{(i)} = X^{(i)} \hat{\beta}_R^{(i)} \tag{20}$$

Here shrinkage parameter (k) is obtained using Eq. (9).

Step 4 The results obtained from fuzzy functions are weighed according to the membership values and the output values are calculated as follows.

$$\hat{y}_k = \frac{\sum_{i=1}^c \hat{y}_{ik} \mu_{ik}}{\sum_{i=1}^c \mu_{ik}}, \quad i = 1, 2, \dots, c, \quad k = 1, 2, \dots, n - p \tag{21}$$

where, \hat{y}_{ik} represents the predicted value obtained from i^{th} cluster for the observation k and \hat{y}_k is the forecasted value of the approach for the observation k . The flowchart of proposed method is given in Fig. 1.

5 Applications

In this study, twelve time series data are analyzed for the evaluation of the forecasting performance of the proposed method. The first five time series data of these datasets are Istanbul Stock Exchange Market (BIST100) data observed daily between years 2009 and 2013. The next five time series dataset are Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) data observed daily between years 2000 and 2004. The next dataset is Australian Beer Consumption (AUST) data observed quarterly between years 1956 and 1994. Finally, the last dataset is Turkey Electricity Consumption (TEC) data observed monthly between first month of 2002 and last month of 2013.

These time series and their features are presented in Table 1. The methods are compared by root mean square error (RMSE) and mean absolute percentage error (MAPE) criteria. RMSE and MAPE are calculated using Eqs. (22) and (23).

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (y_k - \hat{y}_k)^2} \tag{22}$$

$$MAPE = \sqrt{\frac{1}{n} \sum_{k=1}^n \frac{|y_k - \hat{y}_k|}{y_k}} \times 100 \tag{23}$$

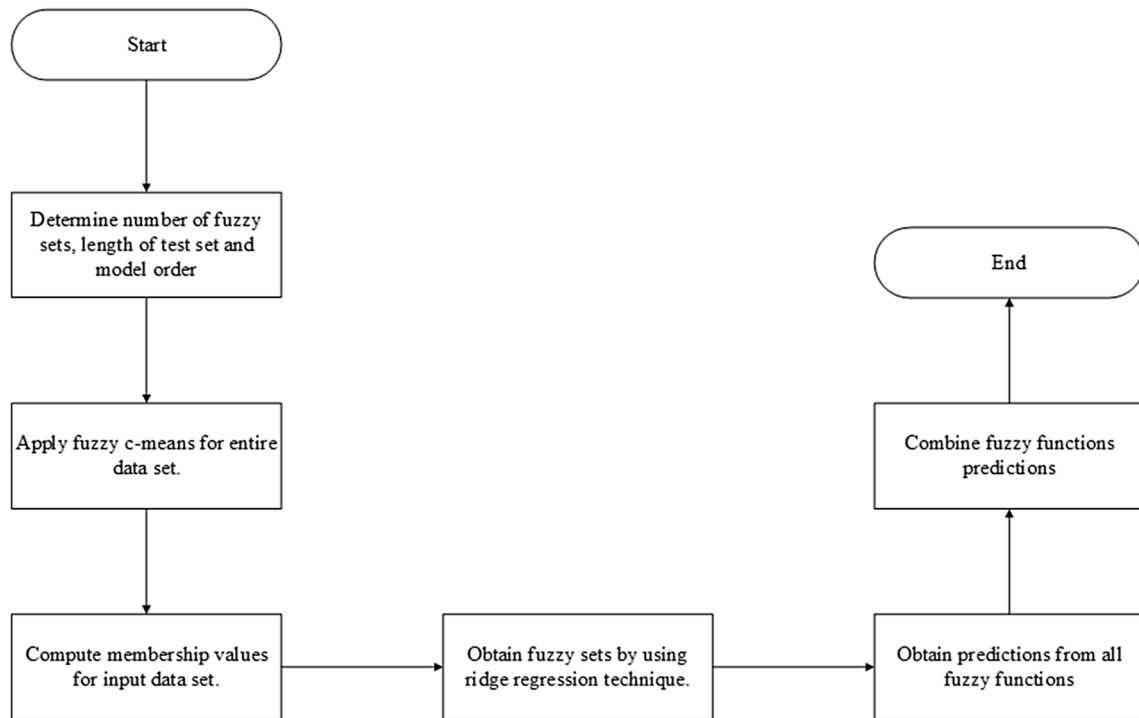


Fig. 1 The flowchart of proposed method

Table 1 The names and features of time series

Series no.	Series/year	The number of			α -cut	ntest
		Obs.	Lag	Cluster		
1	BIST100/2009	103	1:5	3:7	0, 0.1, 0.2	7, 15
2	BIST100/2010	104	1:5	3:7	0, 0.1, 0.2	7, 15
3	BIST100/2011	106	1:5	3:7	0, 0.1, 0.2	7, 15
4	BIST100/2012	106	1:5	3:7	0, 0.1, 0.2	7, 15
5	BIST100/2013	106	1:5	3:7	0, 0.1, 0.2	7, 15
6	TAIEX/2000	271	1:5	5:15	0, 0.1, 0.2	47
7	TAIEX/2001	244	1:5	5:15	0, 0.1, 0.2	43
8	TAIEX/2002	248	1:5	5:15	0, 0.1, 0.2	43
9	TAIEX/2003	249	1:5	5:15	0, 0.1, 0.2	43
10	TAIEX/2004	250	1:5	5:15	0, 0.1, 0.2	45
11	AUST	148	4:8	3:10	0, 0.1, 0.2	16
12	TEC	144	2:16	3:10	0, 0.1, 0.2	12

Obs the number of observations, Lag the number of lags in a time series, Cluster the number of clusters, ntest the length of test set

In the first case, the proposed method was implemented on BIST100 datasets. Alternative forecasting methods except proposed method used in the analysis of BIST 100 datasets are listed below.

ARIMA: Autoregressive Integrated Moving Average Model, The best model was determined by Box and Jenkins (1970).

ES: Exponential Smoothing, Simple, Holt and Winters exponential smoothing methods were applied and the best model was selected (Brown 1963).

MLP-ANN: Multi-Layer Perceptron Artificial Neural Network, The number of inputs and hidden layer neurons were changed from 1 to 5 and the best architecture was selected by trial and error method. Levenberg Marquardt training algorithm was used as learning algorithm (Wilamowski et al. 2007).

Table 2 The obtained results for series 1–5 and when $n_{test}=7$ and 15

Series/ n_{test}	ARIMA (Box and Jenkins 1970)		ES (Brown 1963)		MLP-ANN (Wilamowski et al. 2007)		SC (Song and Chissom 1993)		T1FF (Turksen 2008)		T1FFRR	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
Series 1/7	345	0.0087	345	0.0087	325	0.0083	1402	0.0396	446	0.0101	319	0.0077
Series 1/15	540	0.0120	540	0.0120	525	0.0114	1754	0.0438	534	0.0122	495	0.0112
Series 2/7	1221	0.0183	1208	0.0185	1077	0.0143	1128	0.0182	1180	0.0179	1080	0.0155
Series 2/15	1612	0.0220	1612	0.0220	1603	0.0220	1742	0.0265	1852	0.0264	1575	0.0213
Series 3/7	1058	0.0144	1057	0.0144	920	0.0128	1396	0.0200	1083	0.0153	915	0.0115
Series 3/15	1130	0.0150	1130	0.0150	1096	0.0146	1360	0.0189	1146	0.0156	1028	0.0143
Series 4/7	651	0.0084	651	0.0084	775	0.0111	1292	0.0183	1034	0.0162	720	0.0106
Series 4/15	621	0.0088	621	0.0088	783	0.0117	1047	0.0153	1038	0.0161	676	0.0100
Series 5/7	1362	0.0116	1362	0.0116	1315	0.0109	1450	0.0108	1512	0.0131	1251	0.0102
Series 5/15	1269	0.0109	1269	0.0109	1233	0.0107	1931	0.0176	1279	0.0108	1237	0.0103

n_{test} the length of test set

Table 3 The comparison of VIF values obtained from T1FF and T1FFRR methods for series 1 when $n_{test}=7$

VIF values	
T1FF	T1FFRR
1963.121	2.225606
1545.953	5.193898
1368.769	5.433400
1969.465	1.776532
4.01E-08	3.85E-08
6.97E-08	6.73E-08
6.81E-08	6.68E-08
6.74E-08	6.41E-08
3.48E-08	3.47E-08

SC: Song and Chissom time invariant fuzzy time series method (Song and Chissom 1993), the number of fuzzy sets were changed from 5 to 15 and the best number of fuzzy sets were selected.

FF: Fuzzy Function Approach (Turksen 2008). The model order and the number of fuzzy sets were changed from 1 to 5 and from 5 to 15, respectively.

The best situations were determined for each series. Table 2 summarizes the results obtained from test set for Series 1–5.

The proposed method has 70% success for BIST 100 datasets. And also, the proposed method has a noticeable improvement on the forecasting performance. Besides, although the proposed method has 30% failures the multicollinearity problem was eliminated using the proposed method. An example of multicollinearity problem in T1FF approach is given in Table 3 for Series 1 when $n_{test}=7$. It was shown that T1FF method has multicollinearity problem because some VIF values are greater than 10. And also there is no multicollinearity problem when using T1FFRR method because all VIF values are less than 10.

Table 4 The best situations for series 1–5

Series/ n_{test}	m	cn	α -cut
Series 1/7	4	5	0.1
Series 1/15	3	3	0.2
Series 2/7	4	3	0.1
Series 2/15	2	3	0.0
Series 3/7	2	5	0.1
Series 3/15	4	4	0.2
Series 4/7	2	5	0.1
Series 4/15	2	5	0.0
Series 5/7	2	5	0.1
Series 5/15	2	5	0.0

m the number of inputs, cn the number of clusters

And also, the best situations for series 1–5 for the best results are given in Table 4.

Moreover, Kruskal–Wallis H test was applied for RMSE values of test set from different methods given in Table 2. The significance value is smaller than 0.05 and it is clear that there is a significant difference between all applied methods.

In the second case, the proposed method was implemented on TAIEX datasets. Table 5 summarizes the results obtained from test set for series 6–10.

Analysis of Table 5 reveals that the proposed method exhibit more successful and superior forecasting performance compared to other methods in terms of RMSE performance criteria. And also, the best situations for series 6–10 for the best results are given in Table 6.

In the next case, the proposed method was implemented on AUST dataset (series 11). Table 7 summarizes the results obtained from test set for Series 11. Series 11 was analyzed by Seasonal Autoregressive Integrated Moving Average (SARIMA), Winters' multiplicative exponential smoothing

Table 5 All obtained results for TAIEX

Methods	Series					Mean
	6	7	8	9	10	
Chen (1996)	176.32	147.84	101.18	74.46	84.28	116.816
Chen et al. (2010)	129.42	113.33	66.82	53.51	60.48	84.712
Chen and Chen (2011)	123.62	115.33	71.01	58.06	57.73	85.15
Chen et al. (2012)	119.98	114.47	67.17	52.49	52.27	81.276
T1FFRR	119.73	113.17	62.55	48.73	51.66	79.168

Table 6 The best situations for series 6–10

Series	<i>m</i>	<i>cn</i>	α -cut
Series 6	5	9	0.1
Series 7	2	9	0.2
Series 8	4	13	0.1
Series 9	2	5	0.2
Series 10	4	11	0.0

m the number of inputs, *cn* the number of clusters

method (WMES) proposed by Winters (1960), linear and nonlinear artificial neural network model (L&NL-ANN) proposed by Yolcu et al. (2013), multiplicative neuron model-based fuzzy time series method (MNM-FTS) proposed by Aladag (2013) and T1FF proposed by Turksen (2008) method.

Analysis of Table 7 reveals that the proposed method exhibit more successful and superior forecasting performance compared to other methods in terms

of MAPE and RMSE performance measures. We conclude that the best result is obtained in the case where “*m* = 8, *cn* = 5, and α -cut = 0”. In Table 7, the results of methods compared with proposed method was taken from [20].

And also, the graph of the real observations and the forecasts obtained from proposed method for the test set is presented in Fig. 2. According to this graph, it is clearly seen that the forecasts obtained from proposed method are very accurate.

Finally, the proposed method was implemented to TEC dataset (series 12). The dataset was forecasted using SARIMA, MLP-ANN, Multiplicative Neuron Model Artificial Neural Network (MNM-ANN), L&NL-ANN, and T1FF approach.

Table 8 summarizes the results obtained from test set for Series 12. Table 8 reveals that the proposed method exhibit more successful and superior forecasting performance compared to other methods in terms

Table 7 All obtained results for Series 11

Test data	SARIMA (Box and Jenkins 1970)	WMES (Winters 1960)	L&NL-ANN (Yolcu et al. 2013)	MNM-FTS (Aladag 2013)	T1FF (Turksen 2008)	T1FFRR
430.50	452.72	453.91	449.92	437.50	446.20	446.64
600.00	578.29	575.22	574.28	537.50	580.12	580.95
464.50	487.70	502.32	481.47	437.50	483.04	481.19
423.60	446.28	444.73	442.79	437.50	442.97	442.76
437.00	456.77	459.66	445.12	437.50	444.74	445.13
574.00	583.51	582.48	571.97	537.50	579.90	579.87
443.00	492.13	508.64	472.76	487.50	468.01	465.80
410.00	450.36	450.31	416.36	437.50	418.98	418.72
420.00	461.01	465.40	428.63	437.50	431.60	431.85
532.00	588.96	589.74	559.89	562.50	559.41	559.10
432.00	496.77	514.96	445.75	462.50	444.08	442.34
420.00	454.64	455.89	390.25	412.50	394.99	394.65
411.00	465.46	471.15	412.38	437.50	409.72	410.03
512.00	594.71	597.00	533.19	537.50	525.60	525.92
449.00	501.67	521.28	442.13	437.50	438.91	436.09
382.00	459.17	461.46	405.08	412.50	409.07	408.81
RMSE	47.0367	53.3295	18.7888	29.1381	17.3926	17.0845
MAPE	0.0949	0.1072	0.0357	0.0532	0.0345	0.0340

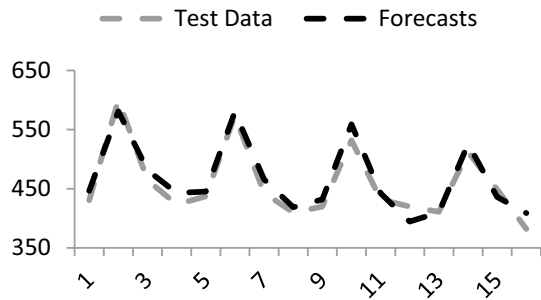


Fig. 2 The graph of the real observations and the forecasts obtained from proposed method for series 11

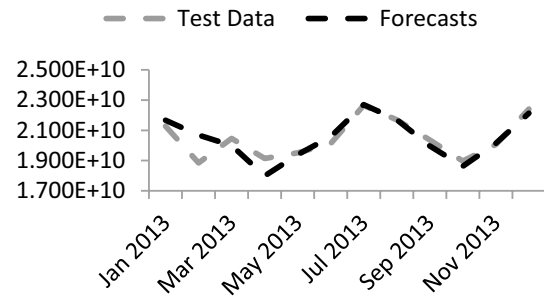


Fig. 3 The graph of the real observations and the forecasts obtained from proposed method for series 12

of MAPE and RMSE performance measures. We conclude that the best result is obtained in the case where “ $m = 16$, $cn = 4$, and α -cut = 0.1”.

And also, the graph of the real observations and the forecasts obtained from proposed method for the test set is presented in Fig. 3. According to this graph, it is clearly seen that the forecasts obtained from proposed method are very accurate.

6 Conclusion and discussions

Fuzzy function approach is a kind of fuzzy inference system that can produce successful results for the analysis of forecasting problems. In fuzzy function approach, a fuzzy function corresponding to each fuzzy set is generated using multiple regression analysis. It can be found a high correlation between the non-linear functions of membership functions.

To overcome this problem, a new fuzzy function approach using ridge regression instead of multiple linear regression method in Type 1 fuzzy function approach is proposed and this is the contribution of this paper. The multicollinearity problem was eliminated using our new proposed method. And also, to show the superior forecasting performance of the proposed method, the proposed method was applied to many real life datasets and its superior forecasting performance was proved. The reason for superior forecasting performance is the elimination of multicollinearity problem. The proposed method can be used in different areas which need time series forecasting such as economy, finance, meteorology, hydrology, etc.

In the future studies, the new proposed approach can be applied to Type 2 fuzzy function approach and also different regression techniques can be used for Type 1 fuzzy function.

Table 8 All obtained results for Series 12

Test data	MLP-ANN (Wilamowski et al. 2007)	SARIMA (0,1,1) (0,1,1) 12	MNM-ANN (Yadav et al. 2007)	(L&NL-ANN) (Yolcu et al. 2013)	T1FF (Turksen 2008)	Proposed method
21,275,408,487	21,504,260,392	21,690,314,907	21,929,976,335	22,005,935,580	21,666,285,710	21,675,649,037
18,841,712,637	21,407,932,217	19,879,748,318	19,674,336,447	21,273,737,442	20,685,602,131	20,715,993,541
20,463,933,683	19,934,741,061	20,679,491,118	20,759,353,666	20,386,359,964	19,996,267,563	20,048,429,544
19,139,248,871	17,339,157,313	18,616,749,529	18,177,421,217	19,068,902,087	17,987,700,997	18,030,742,336
19,511,728,912	19,132,107,582	18,166,751,109	19,502,618,441	19,337,714,416	19,394,021,620	19,427,466,506
20,132,602,347	20,521,392,811	19,476,718,063	19,961,628,854	20,493,224,715	20,584,842,565	20,555,636,973
22,648,523,194	21,559,624,073	22,996,373,189	21,466,603,203	22,493,619,106	22,702,026,988	22,707,403,125
21,698,207,982	21,608,712,088	22,602,364,226	20,970,484,983	22,642,925,619	21,714,016,063	21,762,357,930
20,358,717,408	20,450,417,618	19,372,692,274	20,819,409,088	20,387,962,879	19,965,826,638	20,025,776,421
18,964,661,109	19,608,790,983	19,145,368,564	18,842,139,533	19,384,311,943	18,624,614,607	18,648,025,760
20,061,232,838	20,798,229,864	18,281,990,717	20,341,486,661	20,668,793,095	20,140,094,911	20,160,893,745
22,405,662,577	21,396,071,971	21,319,220,705	20,523,920,152	22,329,629,382	22,138,069,735	22,109,509,894
RMSE	1,065,870,606	917,321,409.3	813,259,007.7	820,978,567.5	687,192,130.1	681,722,004.5
MAPE	0.039841988	0.038842	0.030114538	0.025444349	0.023662007	0.023219348

References

- Aladag CH (2013) Using multiplicative neuron model to establish fuzzy logic relationships. *Expert Syst Appl* 40(3):850–853
- Aladag CH, Turksen IB, Dalar AZ, Egrioglu E, Yolcu U (2014) Application of Type 1 fuzzy functions approach for time series forecasting. *Turkish J Syst* 5(1):1–9
- Aladag CH, Yolcu U, Egrioglu E, Turksen IB (2016) Type 1 fuzzy time series function method based on binary particle swarm optimization. *Int J Data Anal Tech Strat* 8(1):2–13
- Beyhan S, Alci M (2010) Fuzzy functions based ARX model and new fuzzy basis function models for nonlinear system identification. *Appl Soft Comput* 10:439–444
- Bezdek JC (1981) *Pattern recognition with fuzzy objective function algorithms*. Plenum Press, New York
- Box GEP, Jenkins GM (1970) *Time series analysis: Forecasting and control*. San Francisco Holden Day (reviseded. 1976)
- Brown RG (1963) *Smoothing, forecasting, prediction*. Englewood Cliffs, N.J.: Prentice-Hall
- Celikyilmaz A, Turksen IB (2008a) Enhanced fuzzy system models with improved fuzzy clustering algorithm. *IEEE Trans Fuzzy Syst* 16(3):779–794
- Celikyilmaz A, Turksen IB (2008b) Uncertainty modeling of improved fuzzy functions with evolutionary systems. *IEEE Trans On system Man Cybernetics* 38(4):1098–1110
- Celikyilmaz A, Turksen IB (2009) Modeling uncertainty with fuzzy logic, studies in fuzziness and soft computing. Springer
- Chen SM (1996) Forecasting enrollments based on fuzzy time-series. *Fuzzy Sets Syst* 81:311–319
- Chen SM, Chang YC (2010) Multi-variable fuzzy forecasting based on fuzzy clustering and fuzzy rule interpolation techniques. *Inf Sci* 180(24):4772–4783
- Chen SM, Chen CD (2011) TAIEX forecasting based on fuzzy time series and fuzzy variation groups. *IEEE Trans Fuzzy Syst* 19(1):1–12
- Chen SM, Chu HP, Sheu TW (2012) TAIEX forecasting using fuzzy time series and automatically generated weights of multiple factors. *IEEE Trans Syst Man Cybern Part A Syst Humans* 42(6):1485–1495
- Hoerl AE, Kennard RW, Baldwin KF (1970) Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12:55–67
- Hoerl AE, Kennard RW, Baldwin KF (1975) Ridge regression: some simulations. *Commun Stat* 4:105–123
- Jang JSR (1993) ANFIS: Adaptive network based fuzzy inference system. *IEEE Trans On system Man Cybern* 23(3):665–685
- Maciel L, Ballini R, Gomide F (2016) Evolving granular analytics for interval time series forecasting. *Granul Comput* 1:213–224
- Pedrycz W, Chen SM (2011) *Granular computing and intelligent systems: design with information granules of high order and high type*. Springer, Heidelberg
- Pedrycz W, Chen SM (2015a) *Granular computing and decision-making: interactive and iterative approaches*. Springer, Heidelberg
- Pedrycz W, Chen SM (2015b) *Information granularity, big data, and computational intelligence*. Springer, Heidelberg
- Song Q, Chissom BS (1993) Forecasting enrollments with fuzzy time series—Part I. *Fuzzy Sets Syst* 54:1–10
- Takagi T, Sugeno M (1985) Fuzzy identification of systems and its applications to modeling and control. *Man Cybernetics* 15:116–132
- Turksen IB (2008) Fuzzy function with LSE. *Appl Soft Comput* 8:1178–1188
- Turksen IB (2009) Fuzzy system models. *Encyclopedia Complex Syst Sci* 4080–4094
- Wilamowski BM, Cotton NJ, Kaynak O, Dundar G (2007) Method of computing gradient vector and Jacobean matrix in arbitrarily connected neural network. In: *IEEE Symposium on Industrial Electronics*, Vigo, Spain, pp 3298–3303
- Winters PR (1960) Forecasting sales by exponentially weighted moving averages. *Manage Sci* 6:324–242
- Yadav RN, Kalra PK, John J (2007) Time series prediction with single multiplicative neuron model. *Appl Soft Comput* 7:1157–1163
- Yolcu U, Aladag CH, Egrioglu E (2013) A new linear and nonlinear artificial neural network model for time series forecasting. *Decis Support Syst J* 54:1340–1347
- Zarandi MHF, Zarinbal M, Ghanbari N, Turksen IB (2013) A new fuzzy functions model tuned by hybridizing imperialist competitive algorithm and simulated annealing, application: stock price prediction. *Inf Sci* 222(10):213–228

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.