Expert Systems With Applications 199 (2022) 116916

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Type-1 fuzzy forecasting functions with elastic net regularization

Nihat Tak [a,*], Deniz İnan [b]

[a] Department of Econometrics, Kirklareli University, 39700, Kirklareli, Turkey
[b] Department of Statistics, Marmara University, 34700, Istanbul, Turkey

ARTICLE INFO

Keywords:
Type-1 fuzzy functions
Elastic-net regularization
Forecasting
Non-linear forecasting

ABSTRACT

Fuzzy functions have recently been used for forecasting problems. The main concepts behind a fuzzy functions are to cluster the inputs using a fuzzy clustering method and to include the obtained membership grades and their non-linear transformations as new variables in the input matrix. Then, multiple linear regression models are solved for different clusters. However, adding related variables to the input matrix leads to the multicollinearity problem. Thus, the main contribution of the proposed method is to employ an elastic net in fuzzy functions to overcome the aforementioned problem. Two regularization terms occur in an elastic net that come from the ridge and the lasso regression. These regularization terms are optimized using the nested cross-validation approach to overcome the multicollinearity problem in the fuzzy functions method. Twelve practical time-series datasets are analyzed to evaluate the performance of the proposed fuzzy functions. The outstanding performance of the proposed method has been verified in terms of root mean squared errors and mean absolute percentage errors for the selected datasets.

1. Introduction

Forecasting is usually the first step of planning and plays a crucial role in many fields (i.e., economics, finance, health science, meteorology, energy). Thus, many efforts have occurred in the literature to improve the prediction accuracy of forecasting methods. These methods can be defined under probabilistic and non-probabilistic approaches. Exponential smoothing (ES) (Brown, 2004) and autoregressive integrated moving average (ARIMA) (Box et al., 2015) are two of the most popular probabilistic forecasting methods commonly used by researchers in the literature. These methods account for very strong assumptions such as stationarity and constant variance, which are not always able to be provided by real-life datasets. Thus, probabilistic methods fail to give satisfactory prediction accuracy when assumptions are violated. In order to improve prediction accuracy using computer technology developments, non-probabilistic methods have been commonly studied in the recent literature. Artificial neural networks (ANNs) and fuzzy sets (FSs) are the two alternative approaches researchers have mainly used as non-probabilistic forecasting methods.

ANNs were first introduced by McCulloch and Pitts (1943) and in short imitate the human brain. The goal of ANNs is to explore the relationship between the input and the output for complex dataset structures. Due to its success in modeling complex datasets and its assumption-free property, ANNs have been commonly used for forecasting problems in the recent literature. Feedforward, recurrent, and multiplicative neural networks have been widely used for forecasting purposes (Meshram et al., 2019; Svozil et al., 1997; Talaat et al., 2020).

Another alternative approach is based on fuzzy set theory. Fuzzy set theory was proposed by Zadeh (1965) as a generalization of classical set theory. The main idea of fuzzy set theory is to consider uncertainty. Forecasting approaches-based fuzzy set theories are gathered mainly under two roofs: fuzzy time series (FTS) and fuzzy inference systems (FIS). Although huge efforts have occurred regarding FTS methods, the main focus of this study will be on FISs. FIS was first introduced by Zadeh (1973) by defining linguistic variables. Linguistic variables are defined as variables whose values are words in an artificial language. Later, Mamdani and Assilian (1975) proposed FIS in the context of linguistic variables. Another, widely used FIS was proposed by Takagi and Sugeno (1985). Jang (1993) proposed the adaptive neuro fuzzy inference system(ANFIS) for classification problems. ANFIS was later adapted to time series prediction models. Some recent studies on forecasting that adopt ANFIS have been proposed by Catalao et al. (2011), Chang (2008), Cheng et al. (2009, 2013), Hailin and Miao (2020) and Pousinho et al. (2012). However, all these FISs use rules-based systems. Thus, an expert opinion must be present for defining the rules of the problems.

The need for expert opinion is usually the challenge for these FISs. In order to overcome this challenge, Turksen (2008) proposed type-1 fuzzy functions (T1FFs) as a non-rule-based FIS. T1FFs were first

* Corresponding author.
E-mail addresses: nihattak@gmail.com (N. Tak), denizlukuslu@marmara.edu.tr (D. İnan).

https://doi.org/10.1016/j.eswa.2022.116916
Received 7 November 2021; Received in revised form 14 February 2022; Accepted 12 March 2022
Available online 21 March 2022

introduced for regression and classification problems. Later, they were adapted to forecasting purposes. One early study to employ T1FFs in forecasting problems was conducted by Beyhan and Alci (2010), in which T1FFs were used to improve the performance of the autoregressive model through exogenous input. Later, Aladag et al. (2014) used the autoregressive model in T1FFs. The proposed method was able to search for the best model. Dalar and Eğrioğlu (2018) proposed bootstrapped T1FFs to better estimate the model coefficients. Aladag et al. (2016) combined a fuzzy time series model with fuzzy functions, using binary particle swarm optimization to obtain the coefficients with this method. Tak et al. (2018) adapted the moving average autoregressive model in T1FFs to improve the prediction accuracy. Because the objective function was non-derivative, this method also employed particle swarm optimization. Tak (2018) also proposed another forecasting approach to stabilize and improve the forecasting ability of recurrent type-1 fuzzy functions. Goudarzi et al. (2016) proposed an interactively recurrent fuzzy functions model for chaotic time series prediction.

T1FFs' performance depends directly on the clustering algorithm, and T1FFs use FCM to partition the inputs. The thought of using the better clustering algorithm to improve the prediction ability of a time series problem led researchers to adopt different clustering algorithms; intuitionistic FCM, possibilistic FCM, and picture fuzzy sets in T1FFs. Yolcu et al. (2019) used intuitionistic FCM based on the hesitation margin in T1FFs, which also took non-membership values into consideration. Tak (2020b) employed the intuitionistic FCM and autoregressive moving average model alongside the gray wolf optimizer to calculate the coefficients of the models with better calculation times for T1FFs. Picture fuzzy sets have been adapted to T1FFs by Bas et al. (2020). Tak (2020a) proposed possibilistic type-1 fuzzy functions in order to overcome some of FCM's disadvantages (i.e. coincident cluster centers). Studies have shown better calculations of membership grades to increase T1FFs' forecasting performance.

Another concern in T1FFs has been identified as the multicollinearity problem. Multicollinearity problems occur because T1FFs add numerous related variables based on membership grades into the input matrix. Some penalized regression models have been used in the literature to overcome this concern. Bas et al. (2019) adapted ridge regression to T1FFs to eliminate the multicollinearity problem. Kizilaslan et al. (2019) also used ridge regression along with intuitionistic FCM in T1FFs to improve the prediction ability. Bas et al. (2020) proposed picture fuzzy regression functions based on ridge regression.

Ridge regression penalizes the sum of squared coefficients (Hoerl & Kennard, 1970). As a result, high values for shrinkage parameters shrink coefficients towards zero. Thus, ridge regression overcomes the multicollinearity problem in a model; however this does not reduce the number of variables, it only reduces their effects. In addition, the presence of irrelevant variables in the model may result in an overfitting problem.

Some of the variables that T1FFs add to the membership matrix may be irrelevant. In addition, a high number of variables increases a model's complexity. Therefore, instead of the ridge regression method that uses all variables by reducing their effects, using a regression method that makes the variable selection while solving the multicollinearity problem may be more effective.

The least absolute shrinkage and selection operator (Lasso) regression is quite similar structurally to ridge regression (Tibshirani, 1996). It also adds a penalty term for non-zero coefficients. Unlike ridge regression, Lasso regression penalizes the sum of their absolute values. As a result, many coefficients are zeroed out under lasso regression for high values of the shrinkage parameter, which is never the case in ridge regression.

The performances of these two regression methods depends on the structure of the data. Ridge regression shows good performance if most predictors influence the response. On the contrary, Lasso regression performs better when few predictors influence the response. The variable selection for Lasso regression is very dependent on the data and therefore unstable. In the presence of predictor groups with high pairwise correlations, Lasso tends to choose only one of these predictors and does not care which predictor this is. However, the prediction performance of the ridge regression was experimentally observed to be better than Lasso for high correlations between variables (Melkumova & Shatskikh, 2017). Considering these problems, elastic net regression has been proposed, which combines the penalties of ridge regression and Lasso regression to get the best of both (Zou & Hastie, 2005).

Recently, elastic net regression has been used for forecasting purposes in the literature. Plakandaras et al. (2015) proposed a novel hybrid forecasting methodology using elastic net regression as a variable selection tool for forecasting U.S house prices. Uniejewski et al. (2016) used various penalized regression models for electricity cost datasets. They concluded the elastic net to outperform the others. Jiang and Dong (2018) used elastic net for irrelevant variable selection for global solar radiation forecasting. Tian et al. (2019) proposed a forecasting method for short-term prediction of electric load based on the Kalman filter using elastic net to improve the forecasting accuracy. Liu et al. (2018) adapted elastic net regression in the group method for the data handling method to improve the performance of short-term load forecasting.

By considering the performance of elastic net-based forecasting methods in the literature, we have employed elastic net regression in T1FFs. The aim of the study is to improve the forecasting accuracy of T1FFs by solving the multicollinearity problem more efficiently by eliminating the unnecessary predictors and reducing the model complexity.

This paper presents the elastic net regularization and T1FFs in Section 2. Section 3 covers the proposed elastic net regression-based T1FFs with the pseudo-codes. Section 4 compares the performance of the proposed method is compared with various existing forecasting methods for 12 practical time series datasets, and Section 5 lastly discusses the remarks and results from the proposed method.

2. Preliminaries

2.1. T1FFs

Rule-based FISs establish a relation between input and output based on expert opinion. However, the fuzzy functions approach proposed by Turksen (2008) does not require any expert opinion while establishing the relation between input and output by using the cluster centers of the FCM to establish the hidden rules. The main idea of T1FFs is to include an extra variable based on the dataset given in the input matrix. Type-1 fuzzy functions start with constructing the input matrices(Z). First, the lagged variables of the time series are obtained. Next, they are clustered using FCM, and the cluster centers are obtained. Afterward, the input matrices(Z_i) are constructed for all clusters using the lagged variables, the obtained membership grades, and their various nonlinear transformations. The least-square forecasts are obtained for all clusters, and the final predictions are weighted using the membership grades for the new time points. The detailed steps of T1FFs are given below.

- Step 1. The inputs are obtained as the lagged variables (X_t) of a given time series dataset.

$$Z = \begin{bmatrix} Z_{t-1} & Z_{t-2} & \dots & Z_{t-p} \end{bmatrix}$$

- Step 2. The fuzziness index value(m), cluster numbers(c), and cluster centers(v) are initialized.
- Step 3. Membership grades(μ) are calculated in Eq. (1).

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

Step 4. Cluster centers are updated in Eq. (2).

$$v_i = \frac{\sum_{k=1}^{n} \mu_{ik}^{m} z_k}{\sum_{k=1}^{n} \mu_{ik}^{m}}, i = 1, 2, ..c \tag{2}$$

- Step 5. Steps 3–4 are repeated until the difference of cluster centers between two iterations drops under a certain threshold.
- Step 6. The obtained membership grades and their non-linear transformations are included in the inputs($Z^{(i)}$) for each cluster as new variables.

$$Z^{(i)} = \begin{bmatrix} \mu_{i1} & \log(\frac{1-\mu_{i1}}{\mu_{i1}}) & exp(\mu_{i1}) & x_1 & \dots & x_p \\ \mu_{i2} & \log(\frac{1-\mu_{i2}}{\mu_{i2}}) & exp(\mu_{i2}) & x_2 & \dots & x_{p+1} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mu_{ik} & \log(\frac{1-\mu_{ik}}{\mu_{ik}}) & exp(\mu_{ik}) & x_{k-p} & \dots & x_{k-1} \end{bmatrix}, Y^{(i)} = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_{k-p} \end{bmatrix}$$

- Step 7. Then, multivariate regression model ($Y^{(i)} = Z^{(i)}\beta^{(i)} + \varepsilon^{(i)}$) is solved for all clusters.
- Step 8. Outputs of the T1FFs are weighted averages of the results obtained from the clusters as follows:

$$\hat{y}_i = \frac{\sum_{i=1}^{c} \hat{y}_{ik}\mu_{ik}}{\sum_{i=1}^{c} \mu_{ik}}, k = 1, 2, \dots, n \tag{3}$$

2.2. Elastic net

Linear regression model ($Y = X\beta + \varepsilon$ with $\varepsilon \sim N(0, \sigma^2)$ that aims to estimate the $nx1$ response vector Y as a linear combination of nxp matrix of explanatory variables X, and an $nx1$ vector of normally distributed error term with variance σ^2, where n is the number of observations, p is the number of explanatory variables and β is the $px1$ parameter vector.

Because we do not know the actual parameter values (β), we have to estimate them from the sample. In the Ordinary Least Squares (OLS) approach, we estimate parameter values by minimizing the residual sum of squares. ($L_{OLS}(\hat{\beta}) = \sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2$).

The OLS estimator has the desired property of being unbiased. However, in cases where there is a high linear relationship between the explanatory variables or the presence of too many explanatory variables, it may have a huge variance. Consequently, the estimates obtained using OLS approach may be too far from the actual parameter values.

The solution to this problem is to reduce the variance at the expense of adding some bias to the model; that is called regularization. One of the most famous approaches used for regularization is the Ridge Regression approach. Loss function ($L_{Ridge}(\hat{\beta}) = \sum_{i=1}^{n}(y_i - x_i')^2 + \lambda \sum_{j=1}^{p} \hat{\beta}_j$) of Ridge Regression, in addition to minimizing the sum of squared residuals, also penalizes the size of parameter estimates in order to shrink them toward to zero. λ parameter is called the regularization penalty.

Another regularization approach is Least Absolute Shrinkage and Selection Operator (Lasso). Lasso and Ridge Regression approaches are conceptually very similar. Unlike Ridge Regression, Lasso penalizes the sum of the absolute values of the coefficients ($L_{enet}(\hat{\beta}) = \frac{\sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2}{2n} + \lambda(\frac{1-\alpha}{2}\sum_{j=1}^{m}\hat{\beta}_j^2 + \alpha\sum_{j=1}^{m}\left|\hat{\beta}_j\right|)$) rather than the sum of their squares. The λ parameter is called the regularization penalty.

Both methods deal with the multicollinearity problem in different ways. In Lasso, one of the correlated explanatory variables has a larger coefficient, while the rest are zeroed. Thus, Lasso may also make variable selection. Ridge regression uses all of the correlated explanatory variables. In the Ridge regression, the coefficients of correlated explanatory variables are similar. Ridge regression is successful in data structures where most of the explanatory variables effect the response, while Lasso is successful in data structures where few explanatory variables have an effect on the response. However, the actual parameter values are, of course, not known during the analysis.

Zou and Hastie (2005) proposed elastic nets as a new regularization and variable selection method in 2005. Elastic net is a convex combination of the ridge and Lasso regressions and aims to minimize the following loss function.

$$\sum_{i=1}^{n}(y_i - \beta_0 \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda[(1-\alpha)\left\|\beta\right\|_2^2/2 + \alpha\left\|\beta\right\|_1] \tag{4}$$

where Y is the $n * 1$ response vector, X is the $n * p$ matrix for the standardized predictors, n is the number of observations, P is the number of predictor variables, and β is a vector of parameters.

$\|\beta\|_1$ is called l_1 norm ($\|\beta\|_1 = \sum_{j=1}^{p}\left|b_j\right|$) and is the l_2 or Euclidean norm ($\|\beta\|_2 = \sqrt{\sum_{j=1}^{p}\beta_j^2}$).

The loss function of an elastic net has two shrinkage parameters: α and λ. α controls the mixing between the ridge and Lasso regressions and varies from 0 to 1. Using $\alpha = 1$ gives the Lasso, and $\alpha = 0$ similarly gives the ridge. λ controls the amount of penalization. Although high values for λ shrinks the coefficients toward 0 in ridge regression, they shrink many coefficients to exactly 0 in Lasso regression.

3. Proposed Elastic net based Fuzzy Regression Functions (E-FRFs)

T1FFs includes membership grades and their nonlinear transformations into the input matrix and use the least square method to estimate the model coefficients. However, including such related variables into the input matrix violates the assumption in the least square method that no linear relationship exists between explanatory variables and causes overfitting problems. The multicollinearity problem was overcome in the studies by Bas et al. (2019, 2020), and Kizilaslan et al. (2019). While ridge regression handles the multicollinearity problem, it does not account for the overfitting problem. Considering these aforementioned limitations, this study proposes elastic net-based fuzzy regression functions for overcoming the problems of multicollinearity, complexity of the model, and overfitting.

Two penalty terms (α, λ) need to be determined in elastic nets. The optimum values for α and λ in the objective function are obtained using the nested cross validation approach. In order to minimize the objective function, the cyclical coordinate descent algorithm is used. Section 3.1 provides detailed steps for the proposed method, and Algorithm 1 provides the pseudo code.

3.1. Algorithm of E-FRFs

- Step 1. Determine the parameters of the proposed method.
  c = number of clusters
  p = lag length
  nctrain = the length of a fold
  nctest = the length test set in cross validation
- Step 2. The inputs are the lagged values of the time series.
  $Z = [Y_{t-1} Y_{t-2} \dots Y_{t-p}]$
- Step 3. Cluster the inputs using FCM and determine the centers of the clusters and membership grades of the observations in each cluster by considering $alpha - cut$.
  - Step 3.1 Determine the fuzziness index parameter (m), number of clusters (c), and initial cluster centers (v).
  - Step 3.2 Calculate the membership values using Eq. (5).

$$\mu_{ik} = \left[\sum_{j=1}^{c}\left(\frac{d(z_k, v_i)}{d(z_k, v_j)}\right)^{\frac{2}{m-1}}\right]^{-1}, i = 1, 2, \dots, c; k = 1, 2, \dots, n \tag{5}$$

under the constraint: $\sum_{i=1}^{c}\mu_{ik} = 1, if\, \mu_{ik} < alpha - cut$, then μ_{ik} value will be taken as zero. Z is the input matrix, v are the cluster centers, $d(.)$ stands for the Euclidean distance function, c is the number of clusters, and m is the fuzziness parameter in Eqs. (5)–(6).

– Step 3.2 Update the new cluster centers.

$$v_i = \frac{\sum_{k=1}^{n} \mu_{ik}^m z_k}{\sum_{k=1}^{n} \mu_{ik}^m} \tag{6}$$

– Step 3.4 Repeat Steps 2 and 3 until the difference of clusters between two iterations drops under a certain threshold or a certain number of iterations has been reached.

- Step 4. Membership grades, functions of the membership grades, and the lagged values of the time series are combined in X_i.
- Step 5. The optimum α and λ are searched using the nested cross validation approach.

– Step 5.1 The sequence of α and λ are initialized.
– Step 5.2 k folds are obtained by the definition of nested cross validation approach. In this sense, k is calculated in Eq. (7).

$$k = \frac{n - nctrain}{nctest} \tag{7}$$

where n is the number of observations in the time series, $nctrain$ is the length of the first fold, and $nctest$ is the length of the test set for all folds. The inputs of the first fold are given below.

$$X_{(j)}^{(i)(l)} = [\mu_{(j)}^{(i)(l)} \mu_{(j)}^{(i)(l)^2} log(\mu_{(j)}^{(i)(l)}) exp(\mu_{(j)}^{(i)(l)}) Y_{t-1(j)}^{(l)} Y_{t-2(j)}^{(l)} \dots Y_{t-p(l)}^{(i)}] \tag{8}$$

$$Y^{(l)} = [Y_{(j)}^{(l)}] \tag{9}$$

where $l\ :\ 1,2,\dots,k$; $i\ :\ 1,2,\dots,c$; $j\ :\ 1,2,\dots,fn$; $fn = nctrain + (l-1)nctest$.

– Step 5.3 The outputs of the k folds for all c clusters are calculated in Eq. (10).

$$\hat{\beta}_{elastic}^{(i)(l)} = \min_{(\beta_0^{(i)(l)}, \beta^{(i)(l)}) \in R^{p+1}} \sum_{j=1}^{fn} (Y_{(j)}^{(l)} - \beta_0^{(i)(l)} - X_{(j)}^{(i)(l)} \beta^{(i)(l)})^2 + \lambda [\frac{(1-\alpha) \left\| \beta^{(i)(l)} \right\|_2}{2} + \alpha \left\| \beta^{(i)(l)} \right\|_1] \tag{10}$$

where $\hat{\beta}_{elastic}^{(i)(l)}$, which is minimized via the objective function in Eq. (10), is calculated using the "glmnet" (Friedman et al., 2009) function in R. The estimated coefficients of a fold are used to calculate the forecasts of the test set as given in Eq. (11).

$$\hat{Y}_{ForeFold}^{(i)(l)} = X_{(j)}^{(i)(l)} \hat{\beta}_{elastic}^{(i)(l)} \tag{11}$$

– Step 5.4 The outputs of the c clusters are next weighted using the membership values in Eq. (12).

$$\hat{Y}_{ForeFold}^{(i)*} = \frac{\sum_{i=1}^{c} \hat{Y}_{ForeFold}^{(i)(l)} \mu_{(j)}^{(i)(l)}}{\sum_{i=1}^{c} \mu_{(j)}^{(i)(l)}} \tag{12}$$

where $\hat{Y}_{ForeFold}^{(i)*}$ represents the outputs of the proposed method for the lth fold.
– Step 5.5 Calculate the evaluation criteria (i.e. RMSE) and save the calculated evaluation criteria to $RMSE_{folds}$.

$$RMSE_{folds} = [RMSE_l], l = 1,2,\dots,k \tag{13}$$

– Step 5.6 Repeat Steps 5.3–5.5 for all folds.
– Step 5.7 Average the RMSE values and save them to $RMSE_{mean}$.

$$RMSE_{mean_z} = [meanRMSE_{folds}], z = 1,2,\dots,q, q = n\alpha * n\lambda \tag{14}$$

where $n\alpha$ is the length of the α sequence, and $n\lambda$ is the length of the λ sequence.

– Step 5.8 Repeat Steps 5.3–5.7 for all α and λ pairs.
– Step 5.9 Return the optimum value of the α and λ pair that gives the minimum evaluation criteria value (RMSE).

- Step 6. Calculate the outputs from the T1FFs for all clusters using Eqs. (15)–(16).

$$\hat{\beta}_{elastic}^{(i)} = \min_{(\beta_0^{(i)(l)}, \beta^{(i)(l)}) \in R^{p+1}} \sum_{j=1}^{fn} (Y_{(j)} - \beta_0^{(i)} - X_{(j)}^{(i)} \beta^{(i)})^2 + \lambda_{opt} [\frac{(1-\alpha_{opt}) \left\| \beta^{(i)} \right\|_2}{2} + \alpha_{opt} \left\| \beta^{(i)} \right\|_1] \tag{15}$$

$$\hat{Y}_{elastic}^{(i)} = X^{(i)} \hat{\beta}_{elastic}^{(i)} \tag{16}$$

where α_{opt} and λ_{opt} are estimated using the nested cross validation approach; $\hat{\beta}_{elastic}^{(i)}$ are the estimated parameters for the time series in the ith cluster, $\hat{Y}_{elastic}^{(i)}$ is the output (forecasts) obtained from the proposed method for the ith cluster, and $X^{(i)}$ is the actual inputs for the ith cluster.

- Step 7. Calculate the outputs of the proposed method using the forecasts and membership grades in Eq. (17).

$$\hat{Y}_{elastic}^{*} = \frac{\sum_{i=1}^{c} \hat{Y}_{elastic}^{(i)} \mu_{(.)}^{(i)}}{\sum_{i=1}^{c} \mu_{(.)}^{(i)}} \tag{17}$$

Algorithm 1: Pseudo-code of the E-FRFs

```
Initialize c, p, nctest, and nctrain
Input matrix constituted from the lagged values of the time series
Calculate the degrees of membership (μ) by using FCM for the
  training dataset.
Include μ, log((1−μ)/μ), exp(μ) into the input matrix as new variables.
Determine the folds for cross validation.
while (i < the length of alphas(α)) do
    while (j < the length of lambdas (λ)) do
        while (l < k) do
            Calculate RMSE values for the l^th fold
            Store in RMSE_folds
        Average RMSEs in RMSE_folds for i^th alpha and j^th lambda
          in RMSE_mean
Return α_opt and λ_opt which have the minimum RMSE
β_α = P_α
Calculate the forecasts by using β_α with Equations (14)–(16)
```

4. Evaluation

Twelve practical time series datasets are used for evaluating the elastic net fuzzy regression functions. The first five datasets are taken from the Istanbul Stock Exchange Index (ISEX) annually between 2009 and 2013. The last 7 datasets and 15 observations have been left out of the sample for testing purposes. The observations from the Taiwan Stock Exchange (TAIEX) sets have been measured daily between 1999 and 2004. The observations measured over the last two months in the TAIEX datasets have been left out of the sample for testing purposes. The ABC dataset has been selected for evaluating the proposed method as the last application in the study. The last 16 observations have been left out of the sample.

Five parameters must be identified in order to perform the proposed method: the optimum values for alpha and lambda in the elastic net regression, the optimum number of clusters, the optimum value for the fuzzifier in the FCM algorithm, and the optimum lag length in the AR(p) process. The proposed method employs the nested cross validation

Table 1
Search space for ISEX datasets.

Series	n	ntest	Alpha	Lambda	c	p	m
ISEX/2009	103	7/15	0–1	0–1000	2–10	1–10	1.3–3.0
ISEX/2010	104	7/15	0–1	0–1000	2–10	1–10	1.3–3.0
ISEX/2011	106	7/15	0–1	0–1000	2–10	1–10	1.3–3.0
ISEX/2012	106	7/15	0–1	0–1000	2–10	1–10	1.3–3.0
ISEX/2013	106	7/15	0–1	0–1000	2–10	1–10	1.3–3.0

method for obtaining the optimum values of alpha and lambda. The grid search space for alpha and lambda is determined as 0–1 and 10^{-2}–10^{3}, respectively, and both are moved at increments of 0.1 around the space.

Determining the optimum number of clusters is an important and challenging problem for unsupervised clustering. Although numerous studies are found to have determined the optimum number of clusters for FCM, no consensus exists on which method always performs better. For this reason, the optimum number of clusters is searched iteratively. Selecting the fuzzifier is also another challenging task. Selecting the fuzzifier is a common studied topic in the FCM literature. The fuzzifier determines the membership grades of an object in the clusters. Selecting a high fuzzifier means an object belongs to all sets equally. Selecting a small fuzzifier tends to determine whether an object belongs to a set or not. Thus, the optimum value of the fuzzifier may differ from one dataset to another. In this sense, the current study also searches iteratively for the optimum value of the fuzzifier. Root mean squared errors (RMSE; see Eq. (17)) and mean absolute percentage errors (MAPE; see Eq. (18)) have been used to compare the forecasting accuracy of the proposed method with existing methods.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(actual - fitted)^2} \quad (18)$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{(actual - fitted)}{actual}\right| \quad (19)$$

where n is the number of obtained forecasts.

4.1. ISEX datasets

For the first five applications, the ISEX datasets have been evaluated to compare the prediction ability of the elastic net fuzzy regression functions using the selected existing methods. Daily ISEX sets have been obtained for the first two quarters annually from 2009 to 2013. In order to evaluate the performance of E-FRFs, the last 7 and 15 observations are left out of the sample. The search intervals for the alpha, lambda, cluster numbers, fuzzy index value, and lag length are given in Table 1.

The proposed method searched for the best outcomes in terms of the MAPE value, with the RMSE values being calculated accordingly. Table 2 represents the optimum value for α, λ, c, p, and m. The prediction ability of the elastic net fuzzy regression functions have been compared using the autoregressive integrated moving average (ARIMA) (Box et al., 2015), exponential smoothing (ES) (Brown, 2004), multilayer perceptron artificial neural networks (MLP-ANN) (Wilamowski et al., 2007), type-1 fuzzy functions (Turksen, 2008), and type-1 fuzzy functions based on ridge regression (T1FFRRs) (Bas et al., 2019). The outcomes of the selected methods have been obtained from Bas et al. (2019) and Tak (2020b). The methods are compared in terms of MAPE and RMSE values.

Table 3 presents the outcomes from the proposed and selected forecasting methods, with the corresponding RMSE values for the ISEX sets for the length of test set being 7. E-FRFs outperformed the others for 3 datasets and obtained the second-best outcomes for 2 datasets. However, when looking at the mean of the RMSE values over the years, E-FRFs have the best forecasting accuracy. Fig. 1 represents the forecasting accuracy of E-FRFs and other methods in terms of RMSE values.

Table 2
The optimum values of the coefficients for ISEX datasets.

Series/ntest	Alpha	Lambda	c	p
2009 - 7	0	316.2278	8	3
2010 - 7	0	125.8925	10	8
2011 - 7	0	1000	2	3
2012 - 7	0	1	9	3
2013 - 7	0.7	794.3282	8	5
2009 - 15	0	50.11872	3	3
2010 - 15	0.1	1000	9	10
2011 - 15	0	398.1072	8	9
2012 - 15	0	1	3	4
2013 - 15	0	1	5	3

Table 3
The comparison table of selected methods and E-FRFs in terms of RMSE (ntest = 7).

RMSE	ARIMA	ES	MLP-ANN	T1FFs	T1FFRRs	Proposed
2009/7	345	345	325	446	319	**240.2516**
2010/7	1221	1208	1077	1180	1080	**1044.867**
2011/7	1058	1057	920	1083	**915**	946.9358
2012/7	**651**	**651**	775	1034	720	662.3993
2013/7	1362	1362	1315	1512	1251	**832.0063**
Mean	927.4	924.6	882.4	1051	857	**745.292**

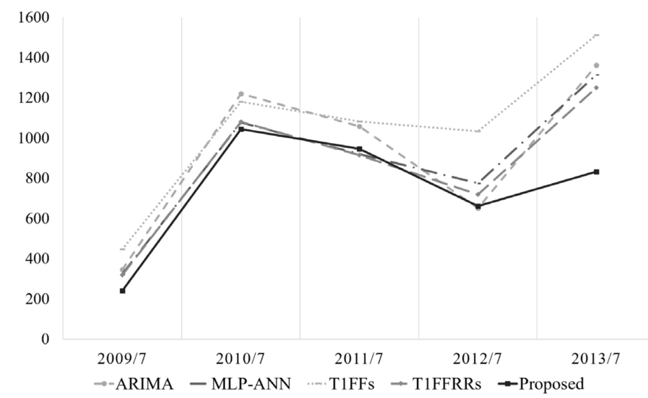

Fig. 1. Line plot of RMSE values for ISEX datasets ($ntest = 7$).

Table 4
The comparison table of selected methods and E-FRFs in terms of MAPE (ntest = 7).

MAPE	ARIMA	ES	MLP-ANN	T1FFs	T1FFRRs	Proposed
2009/7	0.0087	0.0087	0.0083	0.0101	0.0077	**0.0050**
2010/7	0.0183	0.0185	0.0143	0.0179	0.0155	**0.0142**
2011/7	0.0144	0.0144	0.0128	0.0153	0.0115	**0.0114**
2012/7	**0.0084**	**0.0084**	0.0111	0.0162	0.0106	0.0102
2013/7	0.0116	0.0116	0.0109	0.0131	0.0102	**0.0069**
Mean	0.01228	0.01232	0.01148	0.01452	0.0111	**0.0095**

Another metric used to evaluate the performance of E-FRFs is MAPE. The MAPE values for the proposed and selected methods are given in Table 4. Inspecting the table, E-FRFs are seen to have outperformed the other methods for four time series. The proposed E-FRFs also clearly outperformed the other methods in terms of the mean for MAPE.

Table 5 presents the outcomes of the proposed and selected forecasting methods, with the corresponding RMSE values for ISEX sets for the length of test set being 15. The proposed elastic net fuzzy regression functions outperformed the other forecasting methods in terms of the mean for the RMSE metric.

Inspecting Table 6, the proposed elastic net fuzzy regression functions clearly has the best forecasting accuracy in terms of the means for the MAPE values when leaving the last 15 observations out of sample. Fig. 2 allows the forecasting accuracy of the proposed method and the other methods to be visualized in terms of RMSE values.

Table 5
The comparison table of selected methods and E-FRFs in terms of RMSE (ntest = 15).

RMSE	ARIMA	ES	MLP-ANN	T1FFs	T1FFRRs	Proposed
2009/15	540	540	525	534	495	**499.6106**
2010/15	1612	1612	1603	1852	1575	**1299.632**
2011/15	1130	1130	1096	1146	1028	**999.6103**
2012/15	**621**	**621**	783	1038	676	762.276
2013/15	1269	1269	1233	1279	1237	**1207.17**
Mean	1034.4	1034.4	1048	1169.8	1002.2	**953.6598**

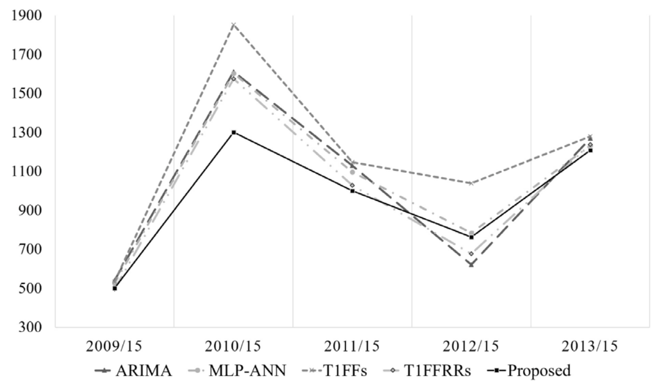

Fig. 2. Line plot of RMSE values for ISEX datasets (*ntest* = 15).

Table 6
The comparison table of selected methods and E-FRFs in terms of MAPE (ntest=15).

MAPE	ARIMA	ES	MLP-ANN	T1FFs	T1FFRRs	Proposed
2009/15	0.012	0.012	0.0114	0.0122	0.0112	**0.0108**
2010/15	0.022	0.022	0.022	0.0264	0.0213	**0.0200**
2011/15	0.015	0.015	0.0146	0.0156	0.0143	**0.0135**
2012/15	**0.0088**	**0.0088**	0.0117	0.0161	0.0123	0.0116
2013/15	0.0109	0.0109	0.0107	0.0108	0.0103	**0.0102**
Mean	0.01374	0.01374	0.01408	0.01622	0.01342	**0.0132**

Table 7
Search space for TAIEX datasets.

Series	n	ntest	Alpha	Lambda	c	p	m
TAIEX/1999	266	45	0–1	0–1000	2–15	1–10	1.3–3.0
TAIEX/2000	271	47	0–1	0–1000	2–15	1–10	1.3–3.0
TAIEX/2001	244	43	0–1	0–1000	2–15	1–10	1.3–3.0
TAIEX/2002	248	43	0–1	0–1000	2–15	1–10	1.3–3.0
TAIEX/2003	249	43	0–1	0–1000	2–15	1–10	1.3–3.0
TAIEX/2004	250	45	0–1	0–1000	2–15	1–10	1.3–3.0

4.2. TAIEX datasets

Six datasets of daily observations for each year from 1999 to 2004 have been used to confirm the performance of the proposed E-FRFs. The search interval for selecting the best parameters of E-FRFs and the length of test sets (ntest) for each year are given in Table 7.

The proposed method searched for the best results in terms of the RMSE values. Table 8 presents the optimum values for α, λ, c, p, and m obtained after the search. The forecasting ability of the elastic net fuzzy regression functions have been compared with those from Bas et al. (2019), Chen (1996), Chen and Jian (2017), Chen and Phuong (2017), Cheng et al. (2016), Tak (2020b), Tak et al. (2018), Yu and Huarng (2006) and Yu et al. (2016).

The outcomes indicate the best results to have been obtained from the E-FRFs for 1999, 2000, 2001, and 2004. In addition, the prediction ability of the proposed method is measured best in terms of the mean of the RMSE values (see Table 9). Fig. 3 represents the forecasting accuracy of the proposed method and selected forecasting methods in terms of RMSE.

Table 8
The optimum parameter specifications for TAIEX datasets.

	1999	2000	2001	2002	2003	2004
c	8	10	11	6	15	15
p	3	9	13	3	13	8
Alpha	0	0.127	0	0.2	0.1	0
Lambda	1	1.584893	100	15.84893	7.943282	316.2278

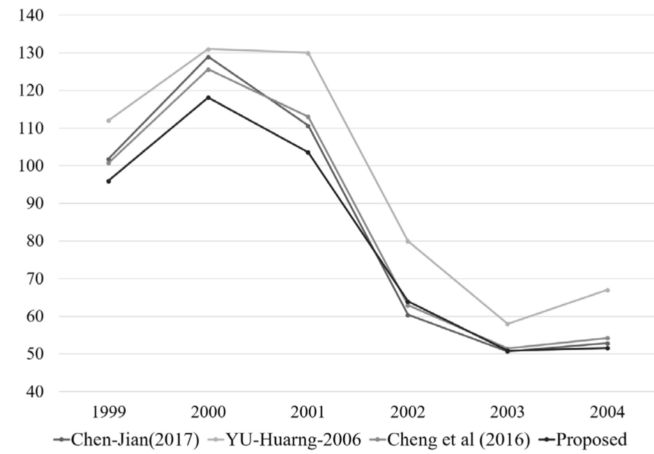

Fig. 3. Graph of RMSE values of T1-R-IFFs and existing methods (*ntest* = 7).

4.3. ABC datasets

Another dataset commonly used to evaluate the performance of E-FRFs in the literature involves beer consumption in Australia. Observations for this have been measured quarterly from 1956 to 1994. The MAPE metric is used to search for the best outcomes from the E-FRFs. Cluster centers vary between 2 to 15 and the lag lengths for the AR process between 1–10; the optimum values for α and λ are searched between 0 to 1 and 0 to 1000 in increments of 0.1. The last 16 observations are left out of sample for testing purpose in the ABC dataset.

The SARIMA, WMES, L-NL-ANN, MNM-FTS, T1FF, and T1FFRR methods have been used to compare the prediction ability of the elastic net fuzzy regression functions. Table 10 represents the actual values from the last 16 observations and one-step-ahead forecasts for the existing forecasting and the proposed methods. Upon inspecting Table 10, the best forecasting accuracy is clearly obtained by the proposed method in terms of both MAPE and RMSE values. Fig. 4 presents a line plot of the actual values and forecasts from the E-FRFs.

5. Conclusions

This study introduces a new forecasting method based on T1FFs and elastic nets. T1FFs start clustering the time series datasets using FCM. Afterward, the obtained membership grades and their nonlinear transformations are included in the T1FFs input matrix. However, including such related variables in the input matrix results in multicollinearity. Although some unnecessary variables are present, the least square method allows all the variables to remain in the model. This also results in overfitting and complexity problems. In order to overcome these problems, the proposed method employs an elastic net. In order to validate the forecasting accuracy of the E-FRFs, 12 practical time series (ISEX, TAIEX, and ABC) datasets are used. The results demonstrate the proposed E-FRF to be quite a competitive forecasting method for certain time series datasets. Overall, E-FRFs have the following contributions and advantages.

- E-FRFs require no assumptions such as stationarity.

Table 9
The comparison table of selected methods and E-FRFs for TAIEX.

	1999	2000	2001	2002	2003	2004	Mean
Chen (1996)	120	176.32	147.84	101.18	74.46	84.28	117.34
Chen and Jian (2017)	101.82	128.95	110.66	**60.41**	50.65	52.86	84.23
Tak (2020b)	97.81	122.23	106.81	64.24	51.5	52.79	82.563
Yu and Huarng (2006)	112	131	130	80	58	67	96.4
Chen and Phuong (2017)	99.97	126.59	110.17	61.62	53.01	53.28	84.11
Tak et al. (2018)	98.33	128.18	106.48	65.14	52.38	53.78	84.05
Cheng et al. (2016)	100.74	125.62	113.04	62.94	51.46	54.25	84.68
Yu et al. (2016)	101.29	125.42	113.22	63.99	52.99	52.4	84.88
Bas et al. (2019)	99.12	119.73	113.17	62.55	**48.73**	51.66	82.493
Proposed	**96.011**	**118.143**	**103.605**	63.928	50.883	**51.595**	**80.694**

Table 10
The comparison table of selected methods and E-FRFs for ABC dataset in terms of RMSE and MAPE.

Test	SARIMA	WMES	L&NL-ANN	MNM-FTS	T1FF	T1FFRR	Proposed
430.5	452.72	453.91	449.92	437.5	446.2	446.64	442.527
600	578.29	575.22	574.28	537.5	580.12	580.95	582.242
464.5	487.7	502.32	481.47	437.5	483.04	481.19	484.235
423.6	446.28	444.73	442.79	437.5	442.97	442.76	441.163
437	456.77	459.66	445.12	437.5	444.74	445.13	435.298
574	583.51	582.48	571.97	537.5	579.9	579.87	581.594
443	492.13	508.64	472.76	487.5	468.01	465.8	467.965
410	450.36	450.31	416.36	437.5	418.98	418.72	419.808
420	461.01	465.4	428.63	437.5	431.6	431.85	431.224
532	588.96	589.74	559.89	562.5	559.41	559.1	556.698
432	496.77	514.96	445.75	462.5	444.08	442.34	445.091
420	454.64	455.89	390.25	412.5	394.99	394.65	401.888
411	465.46	471.15	412.38	437.5	409.72	410.03	410.214
512	594.71	597	533.19	537.5	525.6	525.92	521.351
449	501.67	521.28	442.13	437.5	438.91	436.09	438.146
382	459.17	461.46	405.08	412.5	409.07	408.81	415.898
RMSE	47.04	53.33	18.79	29.14	17.39	17.08	16.823
MAPE	0.09	0.11	0.04	0.05	0.03	0.03	0.032

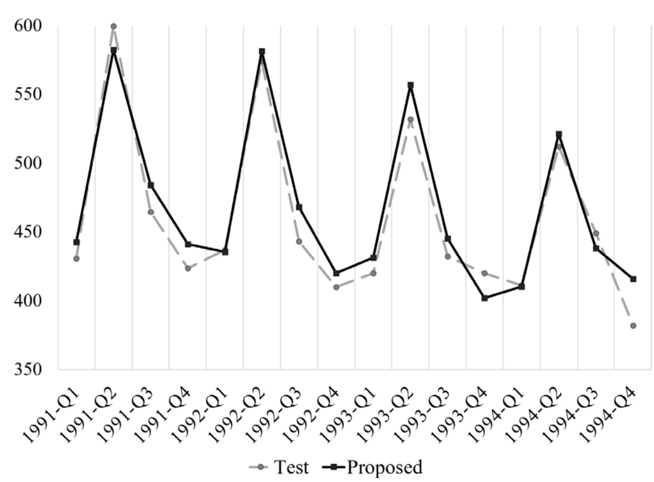

Fig. 4. Graph of RMSE values of T1-R-IFFs and existing methods ($ntest = 7$).

- E-FRFs require no expert opinion to define a rule because E-FRFs are a non-rule based system.
- Type-1 fuzzy functions add membership grades and their non-linear transformations to the model. Thus, elastic net regulation terms are added to the model in order to solve the multicollinearity problem.
- α and λ are optimized using the nested cross validation approach for elastic regression. Therefore, the multicollinearity problem is solved more efficiently by removing correlated variables from the model.
- E-FRFs obtain better forecasting accuracy than most other forecasting methods in the literature.

Although E-FRFs are a powerful forecasting method, their accuracy can be improved by also adding moving average terms to the model. Different fuzzy clustering methods other than FCM may also improve the forecasting ability of E-FRFs. These scenarios suggest future directions of the proposed method.

CRediT authorship contribution statement

Nihat Tak: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Deniz İnan:** Investigation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aladag, C. H., Turksen, I. B., Dalar, A. Z., Egrioglu, E., & Yolcu, U. (2014). Application of type-1 fuzzy functions approach for time series forecasting. *Turkish Journal of Fuzzy Systems*, *5*(1), 1–9.

Aladag, C. H., Yolcu, U., Egrioglu, E., & Turksen, I. B. (2016). Type-1 fuzzy time series function method based on binary particle swarm optimisation. *International Journal of Data Analysis Techniques and Strategies*, *8*(1), 2–13. http://dx.doi.org/10.1504/IJDATS.2016.075970.

Bas, E., Egrioglu, E., Yolcu, U., & Grosan, C. (2019). Type 1 fuzzy function approach based on ridge regression for forecasting. *Granular Computing*, *4*(4), 629–637.

Bas, E., Yolcu, U., & Egrioglu, E. (2020). Picture fuzzy regression functions approach for financial time series based on ridge regression and genetic algorithm. *Journal of Computational and Applied Mathematics*, *370*, Article 112656.

Beyhan, S., & Alci, M. (2010). Fuzzy functions based ARX model and new fuzzy basis function models for nonlinear system identification. *Applied Soft Computing*, *10*, 439–444. http://dx.doi.org/10.1016/j.asoc.2009.08.015.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Brown, R. G. (2004). *Smoothing, forecasting and prediction of discrete time series*. Courier Corporation.

Catalao, J. P. D. S., Pousinho, H. M. I., & Mendes, V. M. F. (2011). Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting. *IEEE Transactions on Power Systems*, *26*(1), 137–144. http://dx.doi.org/10.1109/TPWRS.2010.2049385.

Chang, B. R. (2008). Resolving the forecasting problems of overshoot and volatility clustering using ANFIS coupling nonlinear heteroscedasticity with quantum tuning. *Fuzzy Sets and Systems*, *159*(23), 3183–3200. http://dx.doi.org/10.1016/j.fss.2008.04.003.

Chen, S. M. (1996). Forecasting enrollments based on fuzzy time series. *Fuzzy Sets and Systems*, *81*(3), 311–319. http://dx.doi.org/10.1016/0165-0114(95)00220-0.

Chen, S. M., & Jian, W. S. (2017). Fuzzy forecasting based on two-factors second-order fuzzy-trend logical relationship groups, similarity measures and PSO techniques. *Information Sciences*, *391–392*, 65–79. http://dx.doi.org/10.1016/j.ins.2016.11.004.

Chen, S. M., & Phuong, B. D. H. (2017). Fuzzy time series forecasting based on optimal partitions of intervals and optimal weighting vectors. *Knowledge Based Systems*, *118*, 204–216. http://dx.doi.org/10.1016/j.knosys.2016.11.019.

Cheng, S. H., Chen, S. M., & Jian, W. S. (2016). Fuzzy time series forecasting based on fuzzy logical relationships and similarity measures. *Information Sciences*, *327*, 272–287. http://dx.doi.org/10.1016/j.ins.2015.08.024.

Cheng, C. H., Wei, L. Y., & Chen, Y. S. (2009). Fusion ANFIS models based on multi-stock volatility causality for TAIEX forecasting. *Neurocomputing*, *72*(16–18), 3462–3468. http://dx.doi.org/10.1016/j.neucom.2008.09.027.

Cheng, C. H., Wei, L. Y., Liu, J. W., & Chen, T. L. (2013). OWA-based ANFIS model for TAIEX forecasting. *Economic Modelling*, *30*, 442–448. http://dx.doi.org/10.1016/j.econmod.2012.09.047.

Dalar, A. Z., & Eğrioğlu, E. (2018). Bootstrap type-1 fuzzy functions approach for time series forecasting. In *Trends and perspectives in linear statistical inference* (pp. 69–87). Springer.

Friedman, J., Hastie, T., & Tibshirani, R. (2009). Glmnet: Lasso and elastic-net regularized generalized linear models. R Package Version, 1.

Goudarzi, S., Khodabakhshi, M. B., & Moradi, M. H. (2016). Interactively recurrent fuzzy functions with multi objective learning and its application to chaotic time series prediction. *Journal of Intelligent & Fuzzy Systems*, *30*(2), 1157–1168. http://dx.doi.org/10.3233/IFS-151839.

Hailin, L., & Miao, W. (2020). Fuzzy clustering based on feature weights for multivariate time series. *Knowledge-Based Systems*, Article 105907.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Jang, J. S. R. (1993). ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, *23*(3), 665–685. http://dx.doi.org/10.1109/21.256541.

Jiang, H., & Dong, Y. (2018). A novel model based on square root elastic net and artificial neural network for forecasting global solar radiation. *Complexity*, *2018*.

Kizilaslan, B., Egrioglu, E., & Evren, A. A. (2019). Intuitionistic fuzzy ridge regression functions. *Communications in Statistics. Simulation and Computation*, 1–10.

Liu, W., Dou, Z., Wang, W., Liu, Y., Zou, H., Zhang, B., & Hou, S. (2018). Short-term load forecasting based on elastic net improved GMDH and difference degree weighting optimization. *Applied Sciences*, *8*(9), 1603.

Mamdani, E. H., & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, *7*(1), 1–13. http://dx.doi.org/10.1016/S0020-7373(75)80002-2.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115–133.

Melkumova, L., & Shatskikh, S. Y. (2017). Comparing ridge and LASSO estimators for data analysis. *Procedia Engineering*, *201*, 746–755.

Meshram, S. G., Ghorbani, M. A., Shamshirband, S., Karimi, V., & Meshram, C. (2019). River flow prediction using hybrid PSOGSA algorithm based on feed-forward neural network. *Soft Computing*, *23*(20), 10429–10438.

Plakandaras, V., Gupta, R., Gogas, P., & Papadimitriou, T. (2015). Forecasting the US real house price index. *Economic Modelling*, *45*, 259–267.

Pousinho, H. M. I., Mendes, V. M. F., & Catalao, J. P. D. S. (2012). Short-term electricity prices forecasting in a competitive market by a hybrid PSO–ANFIS approach. *International Journal of Electrical Power & Energy Systems*, *39*(1), 29–35. http://dx.doi.org/10.1016/j.ijepes.2012.01.001.

Svozil, D., Kvasnicka, V., & Pospichal, J. (1997). Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems*, *39*(1), 43–62.

Tak, N. (2018). Meta fuzzy functions: Application of recurrent type-1 fuzzy functions. *Applied Soft Computing*, *73*, 1–13.

Tak, N. (2020a). Type-1 possibilistic fuzzy forecasting functions. *Journal of Computational and Applied Mathematics*, *370*, Article 112653.

Tak, N. (2020b). Type-1 recurrent intuitionistic fuzzy functions for forecasting. *Expert Systems with Applications*, *140*, Article 112913.

Tak, N., Evren, A. A., Tez, M., & Egrioglu, E. (2018). Recurrent type-1 fuzzy functions approach for time series forecasting. *Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies*, *48*(1), 68–77. http://dx.doi.org/10.1007/s10489-017-0962-8.

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, *15*(1), 116–132. http://dx.doi.org/10.1109/TSMC.1985.6313399.

Talaat, M., Farahat, M., Mansour, N., & Hatata, A. (2020). Load forecasting based on grasshopper optimization and a multilayer feed-forward neural network using regressive approach. *Energy*, Article 117087.

Tian, Y., Shen, H., & Gao, W. (2019). Short-term forecasting of electric load based on Kalman filter with elastic net method. In *IOP conference series: Earth and environmental science, Vol. 354*. IOP Publishing, Article 012112.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *58*(1), 267–288.

Turksen, I. B. (2008). Fuzzy functions with LSE. *Applied Soft Computing*, *8*(3), 1178–1188. http://dx.doi.org/10.1016/j.asoc.2007.12.004.

Uniejewski, B., Nowotarski, J., & Weron, R. (2016). Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies*, *9*(8), 621.

Wilamowski, B. M., Cotton, N. J., Kaynak, O., & Dundar, G. (2007). Method of computing gradient vector and Jacobean matrix in arbitrarily connected neural networks. In *2007 IEEE international symposium on industrial electronics* (pp. 3298–3303). IEEE.

Yolcu, O. C., Bas, E., Egrioglu, E., & Yolcu, U. (2019). A new intuitionistic fuzzy functions approach based on hesitation margin for time-series prediction. *Soft Computing*, 1–12.

Yu, T. H. K., & Huarng, K. (2006). Corrigendum to a bivariate fuzzy time series model to forecast the TAIEX. *Expert Systems with Applications*, *37*(4), 3366–3372. http://dx.doi.org/10.1016/j.eswa.2007.05.016.

Yu, F., Zhang, L., Zhang, D., Fujita, H., & Gong, Z. (2016). A novel forecasting method based on multi-order fuzzy time series and technical anaylsis. *Information Sciences*, *367–368*, 41–57. http://dx.doi.org/10.1016/j.ins.2016.05.038.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353.

Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision process. *IEEE Transactions on Systems, Man, and Cybernetics*, *3*(1), 28–44. http://dx.doi.org/10.1109/TSMC.1973.5408575.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *67*(2), 301–320.