



www.sjm06.com

Serbian Journal of Management 12 (2) (2017) 217 - 236

Serbian  
Journal  
of  
Management

## A NEW CONSENSUS BETWEEN THE MEAN AND MEDIAN COMBINATION METHODS TO IMPROVE FORECASTING ACCURACY

Serkan Aras\* and Emrah Gülay

*Dokuz Eylül University, Department of Econometrics, Dokuzcesmeler-Buca, İzmir, Turkey*

*(Received 02 January 2017; accepted 28 April 2017)*

---

### Abstract

To improve the forecasting accuracies, researchers have long been using various combination techniques. In particular, the use of dissimilar methods for forecasting time series data is expected to provide superior results. Although numerous combination techniques have been proposed until date, the simple combination techniques —such as mean and median —maintain their strength, popularity, and utility. This paper proposes a new combination method based on the mean and median combination methods so as to combine the advantages of both these methods. The proposed combination technique attempts to utilize the strong aspects of each method and minimize the risk that arises from the selection of the combination method with poor performance. In order to depict the potential power of the proposed combining method, well-known six real-world time series data were used. Our results indicate that the proposed method presents with promising performances. In addition, a nonparametric statistical test was exploited to reveal the superiority of the proposed method over the single methods and other forecast combination methods from all of the investigated data sets.

*Keywords:* Forecasting, time series, combined forecasts, least support vector machines, neural networks, SETAR, LSTAR, ARIMA

---

### 1. INTRODUCTION

The analysis and forecasting of the time series data plays an important role in several scientific and engineering practices. The development of accurate forecasting

methods constitutes the main subject of concern of the researchers. For this purpose, several linear and non-linear forecasting models have been devised in the literature as a result of numerous studies in this field (Box & Jenkins, 1970; Franses et al., 2014).

---

\* Corresponding author: emrahgulay2011@gmail.com

DOI: 10.5937/sjm12-13091

It is also a well-known fact that no single forecasting method have given the best results for modeling all features of the data generating-process of a time series (Makridakis et al., 1982). Therefore, it is quite risky and irrational to resort to only one of the existing forecasting methods. After the first seminal works (Reid, 1968; Bates & Granger, 1969), which effectively combined multiple forecasts to improve forecasting accuracy, combining forecasting methods has become the most popular topic among researchers from across every field of forecasting. Several researchers have claimed that combining forecasts would have lower error measures than forecasting by a single method (Clemen, 1989; Rapach et al., 2010). The expectations from combining single models are to obtain better forecasts than from single models by applying distinctive modeling abilities of each model in approximating different patterns in the data and reducing the risk of selecting the wrong forecasting method.

Some researchers oppose forecast combinations. Statisticians oppose this because it harms the traditional statistical procedures such as the use of statistical significance, while others believe that a comprehensive single method, comprising of the complete relevant information on the matter, can be more effective for forecasting (Hibon & Evgeniou, 2005; Larrick & Soll, 2003). Despite all these reservations, the combination of forecasting methods remains an interesting approach to achieve more accurate forecasts, with numerous applications (Clemen, 1989; Armstrong, 2001; Zou et al., 2007). Selecting the constituent forecasting models that can be used in combination and the manner of combining them to produce the final combined forecasts for determining the

method of combination are the two most important problems that affect the results obtained. With regard to the first problem, several papers have reported that selecting single models as dissimilar models or methods based on different information as possible can make the combination method superior for forecasting (Armstrong, 2001; Newbold & Granger, 1974; Zhang, 2003).

Among the combination methods, a simple average method that weights all the forecasts equally is the easiest and the most common one. In addition, it can be easily understood, applied, and interpreted, with no need of calculation of any parameters. Moreover, some researchers have indicated that the simple average produces considerably better forecasting results than other complicated combination techniques in most cases (Clemen, 1989; De Gooijer & Hyndman, 2006; Jose & Winkler, 2008; Genre et al., 2013). Nevertheless, simple average is quite susceptible to extreme values and, therefore, the variation in the forecast errors can be high (Armstrong, 2001; Jose & Winkler, 2008). Therefore, some studies use the median, which is less susceptible to extreme values. However, there is no precise result for forecasting superiority between simple average and median. The simple average method produced better results in one study (Stock & Watson, 2004), the worse results in others (Larrece & Moinpour, 1983; Agnew, 1985), and the same result in another study (McNees, 1992). Thus, it is almost impossible to detect any superiority between these two methods, both of which are quite easy in computation.

The motivation for the paper is to show that an individual model may not identify the true process, but combining forecasts from several models may play an important role in

achieving better predictive performance (Terui & Van Dijk, 2002). We leveraged this idea to achieve better predictive performance and studied several different time series models for analyzing the data sets. The special focus of this study was the development of the proposal by Adhikari and Agrawal (2013) regarding the combination of simple average and median. The forecasts of median and simple average were linearly combined in a previous study (Adhikari & Agrawal, 2013), while their weights were fixed at the whole test data point and, additionally, noise was added to each combined forecast. In this study, the weights of simple average and median at all test data points were found one by one in accordance with the forecast values of the single forecasting methods instead of fixing them; moreover, the addition of noise was not needed. Considering the difference between the average and median of the forecasts produced by constituent forecasting models, if the difference was high, the weight of the median was increased; otherwise, the mean was increased. Thus, by establishing a balance between the two combination methods, the variation in the forecasts could be reduced and the risk caused by selection of the wrong combination methods could be minimized.

The rest of the paper is organized as follows. The subsequent section provides a literature review concerning combination techniques. The proposed combination method is presented in detail in Section 3. Section 4 deals with the single forecasting methods used in combination. Section 5 provides datasets and model parameters of the methods used. The results obtained are reported in Section 6. Finally, conclusions and further discussions are given in Section 7.

## 2. LITERATURE REVIEW

In the wake of the pioneering studies by Reid (1968) and Bates and Granger (1969), there was a great rise in the number of studies for increasing the forecasting accuracy and decreasing the error variance by combining different forecasting methods. The articles by Clemen (1989) and De Menezes et al. (2000) summarize the studies performed on this topic across the literature. The reason for combining forecasts from different models arises from the assumption that a single model may not be sufficient to capture all patterns in a dataset. Hence, some studies related to combination of forecasts of various single models were performed. Makridakis and Winkler (1983) showed the effects of combining several forecasting methodologies covering a large number of time series. Armstrong (1989) indicated that combining forecasts provides consistent, but small gains in forecasting accuracy. Diebold and Pauly (1990), Stock and Watson (1999), Chan et al. (1999), and Marcellino (2004) indicated that the combination of forecasts of several different models can improve the forecasting performance. Avoiding difficulties and risks by selecting the right model is another compelling factor for combining forecasts. To this, Zhang (2003) indicated that, “the final selected model is not necessarily the best for future uses due to many potential influencing factors such as sampling variation, model uncertainty and structure change. By combining different methods, the problem of model selection can be eased with little extra effort” (p.160). Zou and Yang (2004) proposed an algorithm to convexly combine the models for obtaining better performance of prediction, although they mentioned that combining all models may not be a good idea. Stock and Watson

(2004), who used linear and non-linear forecasting models, demonstrated that pooled forecasts are superior to the single best model. Hibon and Evgeniou (2005) proposed a simple model criterion based on selecting forecasts and stated that the accuracy of the selection of combination is much better than that of selected individual forecasts.

Several linear combination methods have been proposed on how to combine forecasts (De Gooijer & Hyndman 2006; Newbold & Granger, 1974; Zhang, 2003; Bunn, 1975; Lemke & Gamrys, 2010). Although forecast combinations were performed with non-linear techniques (Timmermann, 2006; Deutsch et al., 1994; Fiordaliso, 1998), these studies are limited. The main problem of the non-linear combination techniques is the dearth of the number of effective studies that can be documented to make a design successful (Timmermann, 2006). In addition, the extensive combination literature reveals that simple methods produce better forecasting results than complicated ones (Clemen, 1989, Timmermann, 2006, Miller et al., 1992; Graefe et al., 2014).

Despite a large body of research concerning combination techniques, the efficiency and robustness of the main statistical combination methods have been emphasized in numerous studies (De Menezes et al., 2000; Timmermann, 2006; De Gooijer & Hyndman, 2006; Schauburger & Tutz, 2014). The simplest method is to use the arithmetical average of single model forecasts to obtain a combined forecast. It was often found in most studies that detailed combination methods, which require estimation of many parameters, perform worse than the combination method using equal weights (Clemen, 1989; Stock &

Watson, 2004; Jose & Winkler, 2008; Winkler & Clemen, 1992; Smith & Wallis, 2009). Armstrong stated (2001) that when there is an uncertainty about a problem at hand, using simple methods is a good strategy. A robust alternative to the simple average is trimmed mean, which is an average value calculated by excluding the highest and the lowest forecasts with equal percent (Armstrong, 2001; Jose & Winkler, 2008; Lemke & Gabrys, 2010; Hassan et al., 2012). Stock and Watson (2004) utilized trimmed mean, with symmetric 5% trimming, and showed that it resulted in the same performance as that by using simple average. In a recent study, Jose and Winkler (2008) showed that trimmed mean can produce slightly more accurate results than simple average, as well as reduce the risk of high error. However, until date, there is no method available to determine the trimming amount. Moreover, median was studied as an alternative to the simple average, with mixed results (Armstrong, 2001; Stock & Watson, 2004; Jose & Winkler, 2008; Larreche & Moinpour, 1983; Agnew, 1985). Therefore, combining the simple average and median correctly can offer an advantage when high forecasting accuracy is required. The combination approach implemented in this study attempts to combine the beneficial properties of simple average and median combination techniques. The detailed formulization of the proposed combination is presented in the Section 3.

### 3. THE PROPOSED COMBINING METHOD

Let the actual testing dataset of a time series forecasted using  $n$  different models be  $Y=[y_1, y_2, \dots, y_N]^T$  and let the  $i^{\text{th}}$  model

forecast of Y be  $\hat{Y}^{(i)} = [\hat{y}_1^{(i)}, \hat{y}_2^{(i)}, \dots, \hat{y}_N^{(i)}]^T$ , ( $i=1,2,\dots,n$ ), where superscripts represent different single models and subscripts correspond to data points in a test set. Let us assume that  $u_j$  and  $v_j$  represent the mean and median of  $\{\hat{y}_j^{(1)}, \hat{y}_j^{(2)}, \dots, \hat{y}_j^{(n)}\}$ ,  $j=1,2,\dots,N$ , respectively. Then the combined forecast of Y will be found as follows:

$$\hat{y}_j = \alpha_j u_j + (1 - \alpha_j) v_j \quad , \quad 0 \leq \alpha_j \leq 1, \\ j = 1, 2, \dots, N \quad (1)$$

The determination of  $\alpha_j$  forms the focus point of this study. In a similar study, Adhikari and Agrawal (2013) fixed this  $\alpha$  value for all data points in a test set. They defined the formula in two ways as median dominating ( $0 \leq \alpha < 0.5$ ) and mean dominating ( $0.5 < \alpha \leq 1$ ) and added noise to the forecasts. Noise addition to real data set, which already includes noises, seemed unnecessary in the present study. In addition, instead of fixing  $\alpha$  at each point in a test set, it is preferred to adjust  $\alpha$  considering the mean and median values at that point. As known from basic statistics, the simple mean is influenced more by outliers than by median. Therefore, if there is an outlier forecast at a data point of the test set, it is logical not to rely on the simple mean. Utilizing that basic knowledge, the following procedure considers the difference between the mean and median at each data point, and, when the difference is high, it reduces the weight of the mean. When the difference is less, it increases the effect of mean on the combination. First, scaling is required to search for the possible maximum difference between the mean and median. For this purpose, the used scaling is given as follows:

$$\hat{y}_j^{min} = \min\{\hat{y}_j^{(1)}, \hat{y}_j^{(2)}, \dots, \hat{y}_j^{(n)}\} \\ \hat{y}_j^{max} = \max\{\hat{y}_j^{(1)}, \hat{y}_j^{(2)}, \dots, \hat{y}_j^{(n)}\} \\ \hat{y}_j^{(i)'} = \frac{\hat{y}_j^{(i)} - \hat{y}_j^{min}}{\hat{y}_j^{max} - \hat{y}_j^{min}}, \quad (i = 1, 2, \dots, n), \\ (j = 1, 2, \dots, N) \quad (2)$$

Now, the forecasts of different models at each data point are scaled, such that the smallest is 0 and the biggest is 1. From now on,  $m_j$  values to evaluate the difference between median and mean can be calculated with the help of Equation 3, as given below:

$$m_j = \frac{|u_j' - v_j'|}{0.4 - |u_j' - v_j'|} \quad (3)$$

Where,  $u_j'$  and  $v_j'$  are respectively the mean and median of the scaled values.  $\{\hat{y}_j^{(1)'}, \hat{y}_j^{(2)'}, \dots, \hat{y}_j^{(n)'}\}$  Owing to the fact that five different single models were employed in this study, 0.4 represents the maximum difference between mean and median occurring in the scaled data. For example, there are two extreme cases that can arise in the scaled data: 0,0,0,1,1 and 0,0,1,1,1. In these two situations, the absolute difference between the mean and median is 0.4.

After  $m_j$  values are found,  $\alpha_j$ , which determines the mean and median weights at each data point, can be calculated with the help of Equation 4, as given below:

$$\alpha_j = e^{-m_j} \quad (4)$$

With the help of the above equation,  $\alpha_j$  remains between 0 and 1, and as the difference between the mean and median becomes closer to 0, the weight of the mean increases, while, in the opposite situation, the weight of the median increases.



#### 4. THE SINGLE FORECASTING MODELS

As is known to all, in forecasting literature, the combination of forecasts from dissimilar and competent models can lead to much better forecasting performance (Armstrong, 2001; Newbold & Granger, 1974; Zhang, 2003). With regard to the number of individual models that will be used in combination, Armstrong (2001) suggested that using at least five forecasts would be a good choice. He also stated that adding more forecasts might improve the forecasting performance of the combination method, but with a decreasing rate of improvement. Jose and Winkler (2008) recommended using five, seven or nine forecasts in combination, in parallel with Armstrong (2001). Hence, following these studies, we use the following single models in our proposed combination method:

- Self-Exciting Threshold Autoregressive (SETAR)
- Logistic Smooth Transition Autoregressive (LSTAR)
- Autoregressive Integrated Moving Average (ARIMA)
- Artificial Neural Network (ANN)
- Least Square Support Vector Machines (LSVM)

##### 4.1. SETAR Model

The first idea on forecasting models of multi-regimes dates back to Bacon and Watts (1971). Tong (1978) proposed the Threshold Autoregressive Model (TAR). In the TAR model, a regime that happened at time  $t$  can be determined by observable  $q_t$  variable about a threshold value. SETAR model, on the other hand, assumes the time series of the

threshold value variable  $q_t$  to be selected by its delay values. For instance, when  $y_t$  variable needs to be modeled,  $q_t = y_{t-d}$  ( $d$  is an integer  $> 0$ ) (Feng & Liu, 2003).

A SETAR model with  $k$  regime ( $d; p_1, p_2, \dots, p_k$ ) can be defined as follows (Chan et al., 2004):

$$Y_t = \begin{cases} \phi_0^{(1)} + \sum_{j=1}^{p_1} \phi_j^{(1)} Y_{t-j} + a_t^{(1)}, & Y_{t-d} \leq q_1 \\ \phi_0^{(2)} + \sum_{j=1}^{p_2} \phi_j^{(2)} Y_{t-j} + a_t^{(2)}, & q_1 < Y_{t-d} \leq q_2 \\ \vdots \\ \phi_0^{(k)} + \sum_{j=1}^{p_k} \phi_j^{(k)} Y_{t-j} + a_t^{(k)}, & q_{k-1} < Y_{t-d} \end{cases} \quad (5)$$

Here,  $k$  is the number of regimes,  $d$  is the delay parameter, and  $p_i$  is the degree of autoregressive model in the  $i^{\text{th}}$  regime model. Threshold parameters must provide the following restrictions.

$$-\infty = q_0 < q_1 < \dots < q_{k-1} < q_k = \infty \quad (6)$$

In each  $i^{\text{th}}$  regime,  $a_t^{(i)}$  are independently and identically distributed normal random variables having 0 mean and constant variance ( $\sigma_i^2 < \infty$ ). In these models, superscripts show the regimes. Dynamic behavior of the times series variable in each regime is assumed to follow a linear autoregressive process; and parameter estimations can be easily performed by the least square method.

##### 4.2. STAR Model

Economic theory most often asserts that the economy behaves differently when the values of certain variables follow different regimes. This very common situation is the assumption that variable changes between two regimes and a smooth transition from one regime to another is estimated from the

dataset. The dependent variable is given with the linear combination of the predetermined variables and random error term, in which each coefficient is a function of the state variable. The single variable version of smooth transition models is the smooth transition autoregressive model (STAR). All of the predetermined variables are due to the delays in the dependent variable. Regimes are endogenously produced. For the stationary and ergodic  $y_t$  process, STAR model of the  $p^{\text{th}}$  degree is defined as follows (Cancelo & Mourelle, 2005):

$$y_t = \pi_0 + \sum_{i=1}^p \pi_i y_{t-i} + F(y_{t-d}) \left[ \theta_0 + \sum_{i=1}^p \theta_i y_{t-i} \right] + u_t \quad (7)$$

Here,  $F(y_{t-d})$  is a transition function and is  $0 \leq F \leq 1$  while  $d$  shows the delay of the transition parameter.

The transition function is defined as  $G(y_{t-d}; \lambda, c)$ . The regime that occurs at  $t$ -time may be determined by the value relative to the observable variables  $y_{t-d}$  and  $G(y_{t-d}; \lambda, c)$ . The different options for transition function  $G(y_{t-d}; \lambda, c)$  lead to the different kinds of change in the behavior of regime change. The most popular option for transition function  $G(y_{t-d}; \lambda, c)$  is the logistic function, as given below:

$$G(y_{t-d}; \lambda, c) = (1 + \exp\{-\lambda(y_{t-d} - c)\})^{-1}, \lambda > 0 \quad (8)$$

The resultant model is termed logistic STAR or LSTAR. The parameter  $c$  can be interpreted as the threshold value between the two regimes. The parameter  $\lambda$  determines the smoothness of the change in the value of the logistic function (Dijk et al., 2002).

There exists two different transition functions in the STAR models. On one hand, STAR models with logistic transition

functions are identified as LSTAR models. On the other hand, STAR models with exponential transition functions are identified as ESTAR models. Before determining the model that can be used in the study, there occurs a decision rule for choosing between LSTAR and ESTAR models. For a thorough overview of the decision rule on selection models, the readers are encouraged to refer to the study of Terasvirta and Anderson (1992). After applying this decision rule, we decided to use LSTAR model for all data sets.

### 4.3. ARIMA Model

In some cases, the time series dealt with was observed to indicate a feature of non-stationary process. It was found, in particular, that the series did not ensure the stationarity conditions of the series in the event of studying with financial time series, such as stock return. For this reason, the ARIMA Model is introduced, allowing the process to become stationary in this part.

Because the ARIMA approach was popularized by Box and Jenkins (1970), it is most often named as the Box-Jenkins model. The ARIMA model is defined as follows:

AR( $p$ ):  $p$  = auto-correlation degree

I( $d$ ):  $d$  = Integration degree (taking the difference)

MA( $q$ ):  $q$  = moving average degree

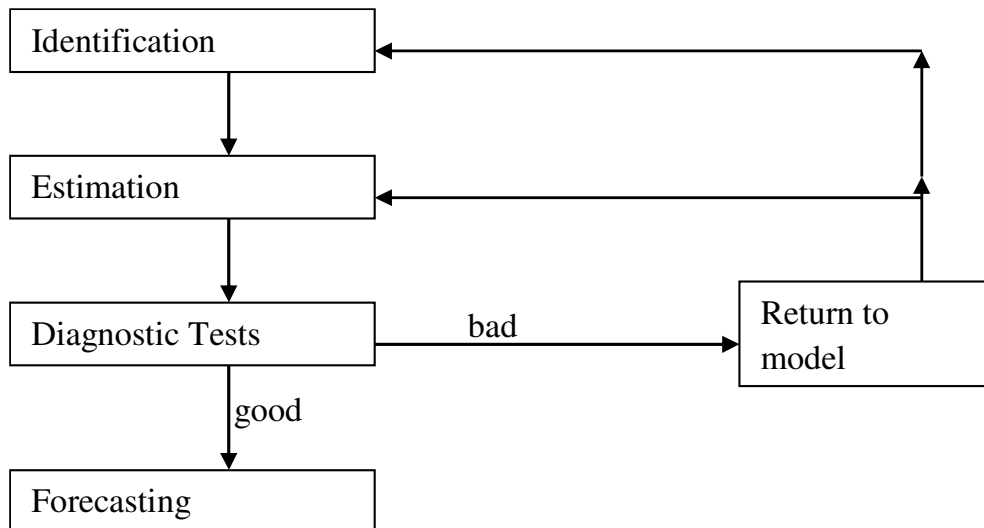
The general view of the ARIMA model is given in the following Equation 9:

$$\underbrace{(1-\Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p)}_{AR(p)} \underbrace{(1-B)^d}_{I(d)} y_t = c + \underbrace{(1-\theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)}_{MA(q)} e_t \tag{9}$$

The process of obtaining the ARIMA model comprises of four stages: i) the determination of the integration degree (d), which makes the series stationary, of p value, showing auto-correlation degree with the help of auto-correlation function and partial auto-correlation function, and of q value, that is, moving average degree; ii) the estimation of coefficients through the least squares method or maximum likelihood method; iii) the interpretation of a model by obtaining diagnostic statistics to provide the validity of the model, and repetition of all these stages performed by returning to the first stage, which has a restatement condition if the model is not valid; and iv) the determination of the accuracy of forecasting by using simple statistics and confidence intervals. The stages cited above are depicted in Figure 1 with the help of a flow chart (Makridakis et al., 1998).

**4.4. ANN Model**

ANN is a mathematical model with parallel data processing structure. Its development was inspired by the structure and function of brain cells. This model possesses the ability to perform non-linear mapping between the input and output. Because the general and flexible modeling abilities of neural networks allow them to find non-linear structures and to model linear processes, this ability makes them an appealing approach in forecasting applications (Zhang, 2001). Several successful applications has shown that neural networks is extremely useful in modeling and forecasting time series (Zhang et al., 1998). A feed forward network with a single hidden layer is often preferred in forecasting time series by considering the problem of overfitting. In forecasting time series, the



Reference: (Makridakis et al., 1998)

Figure 1. The ARIMA Model process



main process performed by Time-Delay Neural Networks is given in Figure 2. Figure 2 shows how  $\{y(t), y(t-1), y(t-2), \dots, y(t-m)\}$  finite time series is matched with  $y$ , which is composed of a single output. The functional form of the network with time-delay inputs is stated in the following equation:

$$\hat{y}_{t+1} = f(y_t, y_{t-1}, \dots, y_{t-m}) = \varnothing(\vartheta_{c0} + \sum_{i=1}^n w_i g(w_{ci} + \sum_{j=0}^m w_{ij} y_{t-j})) \quad (10)$$

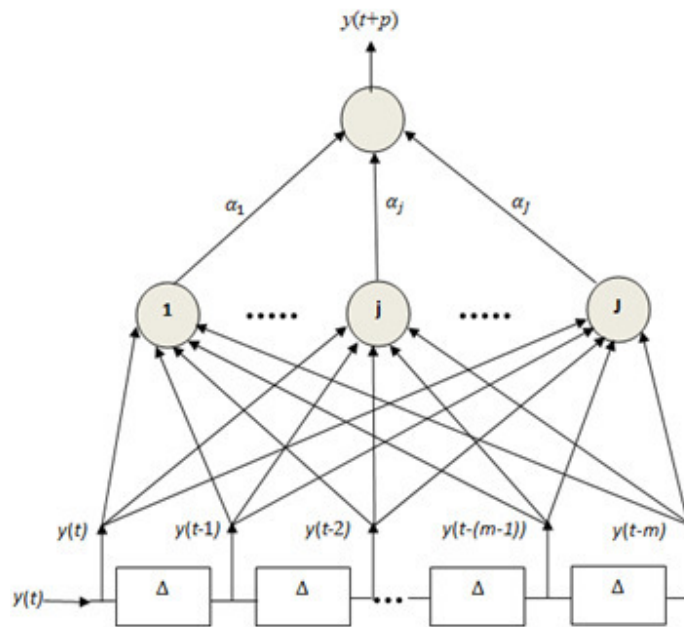
Where,  $y_{t-j}$  is the  $j$  time before observations of the series,  $m$  is the number of inputs or delays in the model,  $n$  indicates the number of hidden units,  $w_{ij}, \{i=1,2,\dots,n, j=0,1,\dots,m\}$  is the weight matrix from input units to hidden units,  $w_i, \{i=1,2,\dots,n\}$  is the weight vector from the  $i$ th hidden unit to output unit,  $g$  and  $\varnothing$  functions, respectively, indicate logistic and linear transfer functions, considering the suggestion of Faraway and Chatfield (1998)

for forecasting applications.  $w_{ci}$  and  $\vartheta_{c0}$ , respectively, represent the constant units of input and hidden layers.

Different learning algorithms can be used for training ANN. Among the learning algorithms available, the most popular is the backpropagation algorithm (Zou et al., 2007). Backpropagation algorithm was used in this study, and for changing the connection weights, Levenberg-Marquardt optimization algorithm was employed. More detailed explanations about neural networks can be found elsewhere (Hagan et al., 1996).

### 4.5. LSSVM Model

LSSVM algorithm is the improved version of Support Vector Machines (SVM) algorithm. Let us suppose that we have  $\{x_i, y_i\}_{i=1}^N$  training data set while the input pattern of  $i$ th sample is  $x_i \in R^n$  and the desired output pattern is  $y_i \in R^n$ . For function estimation, the optimization problem in the



Reference: (Hansen et al., 1999)

Figure 2. Time-Delay Neural Networks

LSSVM approach is formulated as follows (Suykens et al., 2002):

$$f(x_i) = w^T \phi(x_i) + b \tag{11}$$

Where  $\phi(x)$  performs non-linear mapping input data to feature space.  $w$  and  $b$  are the parameters minimize the following objective function:

$$J(w, e_i) = \frac{1}{2} \|w\|^2 + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2 \tag{12}$$

Where,  $\gamma$  represents the regularization constant and  $e_i$  corresponds to the training set error. The constraints of this objective function are as follows:

$$y_i = w^T \phi(x_i) + b + e_i, \quad i = 1, \dots, N \tag{13}$$

The equality constraint in LSSVM is taken instead of the inequality constraint in SVM. In addition,  $e_i$  error was converted into  $e_i^2$  in the objective function. Thus, the solution of the problem becomes easier. The Lagrange function, established for the solution, is as stated below:

$$L(w, \alpha_i, b, e_i) = J + \sum_{i=1}^N \alpha_i [y_i - w^T \phi(x_i) - b - e_i] \tag{14}$$

Where,  $\alpha_i$  is Lagrange multiplier. According to Karush-Kuhn-Tucker (KKT) conditions, we partially differentiate L and obtain the following equations:

$$\begin{aligned} \frac{\partial L}{\partial w} = 0 &\rightarrow w = \sum_{i=1}^N \alpha_i \phi(x_i) \\ \frac{\partial L}{\partial b} = 0 &\rightarrow \sum_{i=1}^N \alpha_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 &\rightarrow \alpha_i = \gamma e_i \\ \frac{\partial L}{\partial \alpha_i} = 0 &\rightarrow y_i - w^T \phi(x_i) - b - e_i = 0 \end{aligned} \tag{15}$$

The optimization problem takes the form of the following linear system by eliminating  $w$  and  $e_i$ :

$$\begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & K(x_1, x_1) + \frac{1}{c} & \dots & K(x_k, x_k) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & K(x_N, x_1) & \dots & K(x_N, x_N) + \frac{1}{c} \end{bmatrix} \begin{bmatrix} b \\ \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ \vdots \\ y_N \end{bmatrix} \tag{16}$$

where  $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$  is known as the kernel function.  $\alpha_i$  and  $b$  are obtained by solving the linear equations and, finally, the following LSSVM model is attained for function estimation:

$$y(x) = \sum_{i=1}^N \alpha_i K(x, x_i) + b \tag{17}$$

Any function that satisfies Mercer’s condition can be used as the kernel function. In this study, the radial basis function (RBF) given below was frequently utilized as the kernel function:

$$K(x_i, x_j) = \exp\left(-\frac{|x_i - x_j|^2}{2\sigma^2}\right) \tag{18}$$

The main difficulty of LSSVM algorithm is the selection of free parameters, such as kernel parameter and trade-off parameter. In this study, a grid search was employed, which is a common approach (Hsu & Lin, 2002) to overcome this selection problem.

## 5. EMPIRICAL RESULTS

### 5.1. Datasets and Model Parameters

Performance evaluation of the proposed method was carried out using six well-known data series. Time plots of the series under investigation are given in Figure 3. Lynx data are used by many researchers for

comparing the performances of linear and non-linear methods (Subba Rao & Sabr, 1984; Priestley, 1988). Before modeling, the series' logarithm to the base 10 was taken, as suggested by Priestley (1988). The last 14 observations of the series, composed of a total of 114 observations, were divided as the test set. Sunspot series was used to assess the performance of several forecasting methods by several researchers (Subba Rao & Sabr, 1984; De Groot & Wurtz, 1991). The last 67 observations of the Sunspot series, forming a total of 289 data, were taken as the test set. The third series is the annual Real GNP of the USA, which has trended over a period of time, and can be noted in the earlier part of the series, where there is a significant cyclical behavior (Hipel & McLeod, 1994). The GNP series is non-stationary by means of cyclic fluctuations and has no seasonal component. The total number of observations of the series is 85, and while we use 70 observations for modeling the series, the rest of the observations were used for the test set.

The fourth series is the number of childbirths annually per 10,000 of 23-year-old women in the USA between 1917 and 1975. As shown in the time plot of the series, the wide-ranging trends in the birthrate—declining during the Depression and increasing from World War II onward, followed by a drop after 1960—are clearly detectable (Velleman & Hoaglin, 1981). We can see from the time plot that the childbirth series is non-stationary and non-seasonal. A total of 59 observations were made for the total size of the series. The testing set contained the total 10 forecast observations used for comparing forecasts of different combining techniques. Other series taken from Box and Jenkins (1976) is the monthly numbers of passengers in international air

travel between 1949 and 1960. We have followed the suggestions of Box and Jenkins (1976) by taking the logarithm base 10 of the number of passengers. This series is an example of a seasonal time series. Owing to the seasonal nature of air travel, we expected heavier travel in summer months, as shown in the time plot of the series, where a 12-month annual pattern with, an upward trend can be clearly seen (Woodward et al., 2011). The series contains 144 observations, which are monthly totals of international airline passengers. The last 12 observations are obtained for the test set. The final series used in the study is quarterly new plant/equipment expenditures in the USA for 1964–1976. The seasonality of new plant/equipment expenditures can be easily seen in the time plot of the series. The original series is split into a dataset containing 44 observations for building the models, as well as a dataset that includes eight observations to test the models. All datasets mentioned here can be easily found on the website of Time Series Data Library (Hyndman, 2012).

Table 1 provides the model parameters of single models that are found in all experiments of datasets. For SETAR ( $k; p_1, p_2$ ) model,  $k$  denotes the number of regimes and  $p_1$  and  $p_2$  denote the degree of autoregressive model in low and upper regimes, respectively. In LSTAR ( $p_1, p_2$ ) model,  $p_1$  and  $p_2$  stand for the degree of autoregressive model in low and upper regimes respectively. Regarding the ARIMA model, it is known from previous studies (Zhang, 2003; Subba Rao & Sabr, 1984; Priestley, 1988), that a subset autoregressive model of order 12 and 9 are the simplest ARIMA models for Lynx and Sunspot datasets respectively. All ARIMA models used for other series are presented in Table 1.

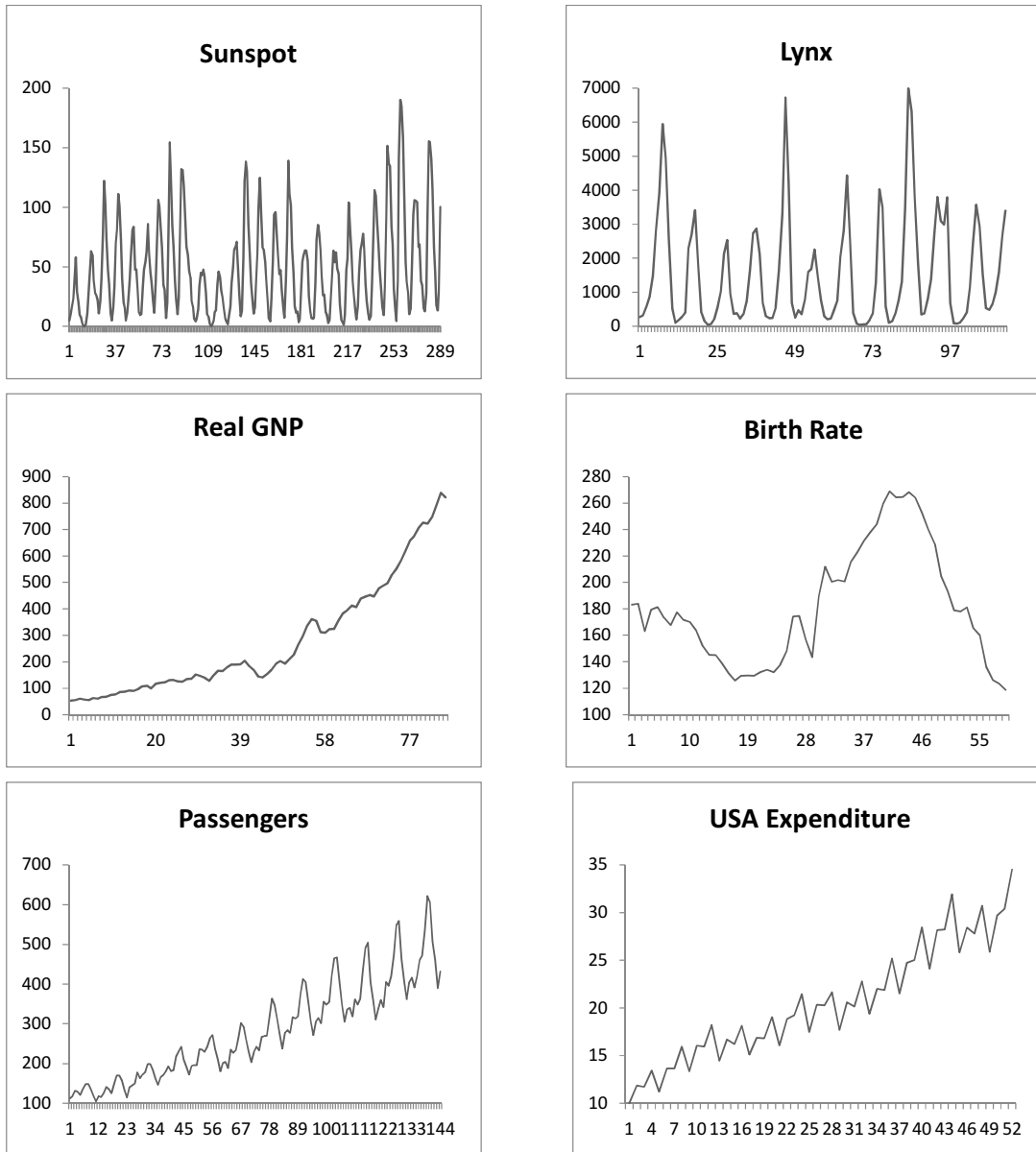


Figure 3. Time plots of the series

For ANN models, an experiment was conducted to determine the numbers of input and hidden units, and the input and hidden unit numbers producing the least square error in the validation data set were used as the final model. In ANN (i,h,o) model, i corresponds to the number of input units, h represents the number of hidden units, and o represents the number of output units in the

final model. Finally, to find free parameters of the LSSVM method, a grid search was employed, conducting a 10-fold cross-validation. Also, the analyses were based on the one-step-ahead forecast errors, which were differences between the data value at time t and the forecast of that value made at time t-1. All experiments in this study were implemented on Matlab and R software.

Table 1. Parameters of the single models for the six real time series

Data Sets	SETAR	LSTAR	ARIMA	ANN	LSSVM	
					$\gamma$	$\sigma^2$
<i>LYNX</i>	SETAR(2;4,3)	LSTAR(3,3)	ARIMA(12,0,0)	ANN(2-6-1)	19.86	3.67
<i>SUNSPOT</i>	SETAR(2;2,6)	LSTAR(6,6)	ARIMA(9,0,0)	ANN(3-9-1)	246.54	11.68
<i>RGNP</i>	SETAR(2;1,1)	LSTAR(1,5)	Random Walk	ANN(4-1-1)	288.91	8176.9
<i>BIRTHS</i>	SETAR(2;1,2)	LSTAR(1,2)	Random Walk	ANN(4-2-1)	1.199	0.019
<i>AP</i>	SETAR(2;1,1)	LSTAR(1,3)	SARIMA (0,1,1) $\times$ (0,1,1) <sup>12</sup>	ANN(12-4-1)	63.96	2793.6
<i>UE</i>	SETAR(2;1,4)	LSTAR(5,2)	SARIMA (0,1,1) $\times$ (0,1,1) <sup>4</sup>	ANN(4-4-1)	3.63	1.88

## 5.2. Results

Table 2 indicates the results of the single models in terms of MSE and MAE values. As seen in Table 2, LSSVM produced the best model results in Lynx data. However, the ANN model gave close results. These two models are superior to other single models in the Lynx dataset. In Sunspot data, the ARIMA model gave better error performance than all other models. LVSSM and ANN models followed the ARIMA model. For the RGNP data set, from a non-linear view-point, ANN model demonstrated better forecasting performance than other individual models. LSTAR model is the second-best model, according to the forecasting performance. Regarding other childbirths series that show fluctuations of birth rates through the years, the models that best describe the behavior of the series are SETAR and LSTAR. The forecasting performances of the two models relative to other models were superior. For the next series, that is, airline passengers, which was affected by increasing trend and seasonal variations from January 1949 to December 1960, the best values of MSE and MAE were reported for LSTAR and ARIMA models. For the last series, expenditures in the USA, seasonally adjusted, SETAR and ANN models showed better results in terms of forecasting performance. The results for all

datasets showed that, if a dataset has a trend or cyclical pattern, non-linear models such as SETAR, LSTAR, and ANN models showed promising forecasting performance with respect to other models. As observed from Table 2, it is not possible to declare any one of the individual models as the best for all datasets.

In Table 3, the results of the combination methods were provided. In addition to the mean, median, and trimmed combination methods, the results of other group combination methods (we can call this “group the meta-combining” methods, because they are based on the combination of combining methods) were also reported in Table 3. Combining methods of  $\alpha = 0.5$ ,  $\alpha = 0.25$ , and  $\alpha = 0.75$  indicate that all of them were obtained by means of combining mean and median combination methods, but at different levels, which indicate the weights of mean and median in Equation 1. Thus,  $\alpha = 0.25$  represents that the weight of median is bigger than the mean in the combination,  $\alpha = 0.5$  corresponds to equal weights for mean and median, and  $\alpha = 0.75$  shows that the weight of the mean is bigger than median in the combination method. Other meta-combining methods in the table are Med\_Dom (Median Dominating) and Mean\_Dom (Mean Dominating) techniques proposed by Adhikari and Agrawal (2013).

It is difficult to judge which method is



Table 2. Results of the single model

Models	Lynx		Sunspot		RGNP		Births		AP		UE	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
<b>ARIMA</b>	0.030	0.133	311.97	13.36	526.43	17.29	136.59	10.15	360.03	14.47	0.617	0.670
<b>SETAR</b>	0.028	0.137	402.71	15.27	516.06	18.40	96.14	7.51	455.98	15.74	0.574	0.564
<b>LSTAR</b>	0.027	0.115	433.24	15.77	426.18	17.06	99.63	8.10	329.36	12.87	2.047	1.247
<b>ANN</b>	0.012	0.093	389.87	14.73	422.18	17.32	115.60	6.94	429.76	16.10	0.594	0.604
<b>LSSVM</b>	0.011	0.076	380.86	14.52	458.75	17.98	148.10	9.98	416.97	15.77	1.184	0.857

Table 3. Results of the combination model

Models	Lynx		Sunspot		RGNP		Births		AP		UE	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
<b>Median</b>	0.0090	0.075	325.44	13.33	426.86	17.09	86.84	6.27	407.08	14.97	0.587	0.596
<b>Mean</b>	0.0108	0.083	304.44	12.71	417.70	16.64	88.63	6.25	359.09	14.10	0.477	0.540
<b>Proposed</b>	0.0087	0.074	312.48	12.95	413.89	16.76	87.48	6.19	385.61	14.56	0.494	0.528
<b><math>\alpha=0.5</math></b>	0.0096	0.077	312.81	12.98	418.90	16.86	87.59	6.25	381.88	14.53	0.522	0.544
<b>Trimmed</b>	0.0097	0.073	321.01	13.14	436.52	17.21	86.74	6.25	411.12	15.01	0.581	0.615
<b><math>\alpha=0.25</math></b>	0.0101	0.080	308.09	12.83	417.46	16.75	88.08	6.24	370.19	14.31	0.497	0.538
<b><math>\alpha=0.75</math></b>	0.0092	0.076	318.59	13.15	422.03	16.98	87.18	6.25	394.18	14.75	0.552	0.570
<b>Med_Dom</b>	0.0096	0.077	334.36	13.24	432.37	17.14	89.90	6.69	446.93	15.22	0.582	0.628
<b>Mean_Dom</b>	0.0107	0.082	324.59	12.92	426.70	16.77	90.98	6.66	420.69	14.79	0.532	0.576

better by screening the numbers in Table 3. We therefore used statistical tests to judge more objectively the superiority of the methods in comparison with others. However, mean combining is of great importance in literature, and several studies have used it as a benchmark, which is difficult to beat for more sophisticated approaches (Clemen, 1989; Timmermann, 2006; Makridakis & Winkler, 1983; Stock & Watson, 1999). Our results were in agreement with those in the literature; however, if some forecasts of the single models are much superior to those of single models, the proposed method tends to be slightly better than the simple average method in combining forecasts. For LYNX, RGNP, BIRTHS, and UE data results, we have one or more superior forecasts by individual models as compared with

forecasts of other individual models. In such situations, the proposed method tends to work better than simple average combining approaches in terms of MSE or MAE; however, if the forecasts of the component models are almost similar, as for SUNSPOT and AP data, the simple average combining approach tends to perform better than the proposed method.

The nonparametric Friedman test was employed to determine statistically significant performance differences between all forecasting methods, considering all datasets under investigation. This test included the null hypothesis that the different groups have been selected from populations with the same median; and the alternative hypothesis was that at least one median is different. The Friedman test results obtained were, as follows: for forecasting MSE, the

Friedman’s  $\chi^2$  statistic was 41.49 and  $p = 0.000079$ , and, for forecasting MAE, the Friedman’s  $\chi^2$  statistic was 44.76 and  $p = 0.000023$ . Significant differences were noted between forecasting methods, and then multiple comparison procedures were used to decide which groups were significantly different from others, based on the mean rank differences in the groups. Detailed information about these tests is given elsewhere (Hochberg & Tamhane, 1987).

In Figure 4, the mean rank of each group is indicated by a symbol, and the interval is represented by a line extending from the

symbol. Two group means were significantly different if their intervals were disjoint; they are not significantly different when their intervals overlapped. As shown in Figure 4, the forecasting accuracies obtained through combining methods are often better than individual models in terms of MSE and MAE. The proposed method produced statistically significant forecasting performance compared to all single models in terms of both MSE and MAE. The proposed combining method reduces the overall forecasting error in a more accurate manner than other combining methods in

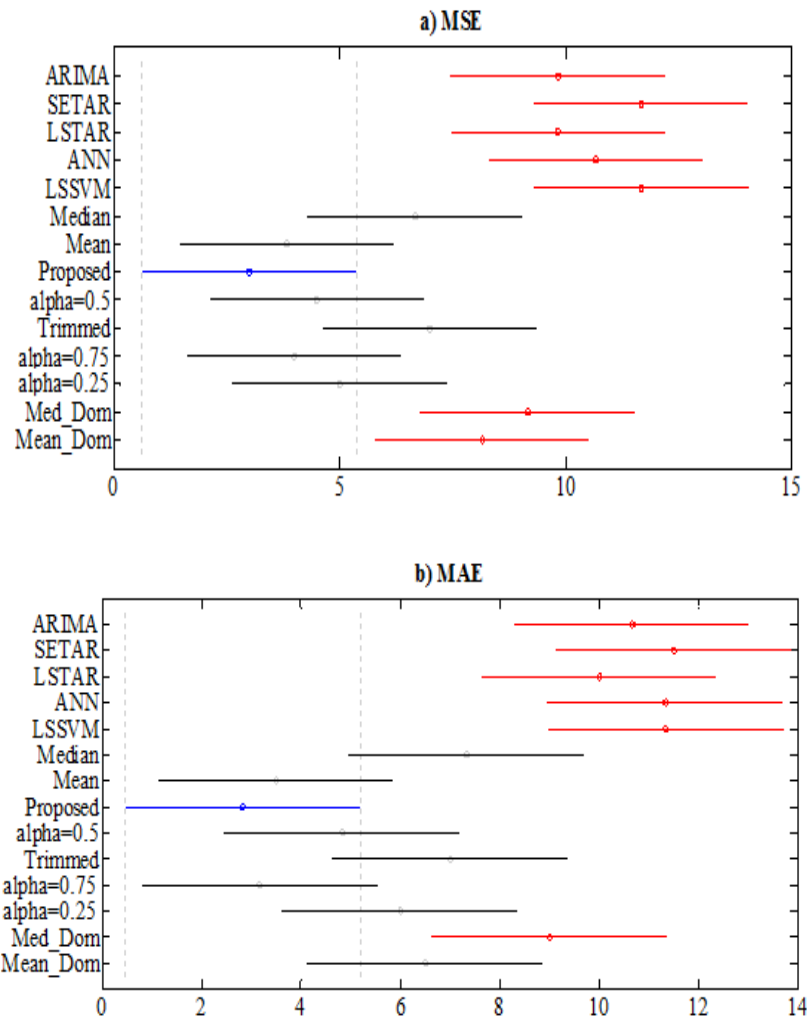


Figure 4. Friedman test results for MSE and MAE

question. As a great extent of error was already reduced with the help of simple average and median combination methods, in our method, we only combined two combination methods effectively to improve the forecasting accuracy, as well as to decrease the model selection risk.

## 6. CONCLUSION

After Bates and Granger's optimal weights approach, which did not work well in practice, there has been reports of several improvement in several unique combining methods. In the literature, several methods are available, varying from simple average methods to more complex ones for combining weights. Methods to increase forecasting accuracy have attracted the attention of researchers over the past few decades. For this purpose, the use of a combination of forecasting methods has been commonly used. Although there are several different and enhanced combination methods available, simple combination methods, such as mean and median, have generally produced better results than more complicated methods. However, the simple mean method also possess a disadvantage as, it is sensitive to extreme forecasts. These extreme forecasts have a detrimental effect on the combined forecast in the course of combining all forecasts. Hence, a consensus approach between combination methods of the mean and median has been introduced in this study. By the proposed method, it is hoped that the mentioned disadvantages of the mean and median combination methods have alleviated. The difference of this study from other studies in literature is to use dynamic weights in each forecasting point instead of using fixed weights in combining

the mean and median combination methods.

In this study, a linear combination method was used, which is partially responsive to the question as to which of the so-called combination methods be selected, and in which weight values change at every data point. Five single models, including ARIMA, SETAR, LSTAR, ANN, and LSSVM, were selected in accordance with the purpose of the study. In our analysis, we considered eight combining methods to compare our proposed approaches. Promising results were achieved in the consequence of the research of the proposed method, which was implemented on six real known datasets. Through the use of the proposed method, a performance between the mean and median combination was achieved in the worst case, although superiority was established to both combination methods in the best case. Moreover, we applied a nonparametric Friedman test to demonstrate that significant differences exist in combined forecasts by different methods under the study. Based on the Friedman test, our proposed approach for combined forecasts seems slightly better than other combining methods.

The decision makers in both private and public organizations must make effective plans to survive and increase their market shares in today's competitive global economy. Every plan such as capacity plan, production and inventory plan, purchasing plan, manpower plan, and financial plans depends highly on forecasting future conditions accurately. To improve forecasting accuracy, a considerable number of methods have been proposed. Based on the findings reported in this study, managers in all levels can employ the proposed combination method in reducing risk of selecting poor combination method.

## НОВИ КОНСЕНЗУС У КОМБИНАЦИЈИ МЕТОДА СРЕДЊЕ ВРЕДНОСТИ И МЕДИЈАНЕ У ЦИЉУ ПОВЕЋАЊА ТАЧНОСТИ ПРЕДВИЂАЊА

Serkan Aras, Emrah Gülay

### Извод

Како би се увећала тачност предвиђања, истраживачи већ дуће време користе различите комбинације техника. У суштини, паралелно коришћење различитих метода за предвиђање података временских низова требало би да омогући супериорне резултате. Иако су до данас предложене бројне комбинације техника, једноставна комбинација - као што је средња вредност и медијана - задржавају своју примењивост и популарност код истраживача. Овај рад предлаже нови комбиновани метод, који се заснива на комбинацији средње вредности и медијане, како би се комбиновале предности обеју метода. Предложена техника комбинације има за циљ коришћење јаких аспеката аваке од метода и смањење ризика који се може појавити из селекције комбинације метода са лошим перформансама.. Како би се показала потенцијална снага предложеног комбинованог метода, коришћене су добро познате временске серије из реалне праксе. Добијени резултати показују да се предложени метод може адекватно применити, са погодним перформансама. Додатно, коришћен је непараметрички статистички тест који предочава супериорност предложеног метода над појединачним методима и другим комбинацијама које се користе за предвиђање, на истраживаном сету података.

*Кључне речи:* Предвиђање, временска серија, комбиновано предвиђање, векторске машине и подршка, неуронске мреже, СЕТАР, ЛСТАР, АРИМА

### References

- Adhikari, R., & Agrawal, R.K. (2013). A linear hybrid methodology for improving accuracy of time series forecasting. *Neural Computing and Applications*, 25, 269-281.
- Agnew, C. (1985). Bayesian consensus forecasts of macroeconomic variables. *Journal of Forecasting*, 4, 363-376.
- Armstrong, J.S. (2001). Combining forecasts. In: J.S. Armstrong (Eds.). Chapter 13 in *Principles of forecasting: a handbook for researchers and practitioners*. Kluwer Academic Publishers: Norwell, MA.
- Armstrong, S. (1989). Combining forecasts: The end of the beginning or the beginning of the end? *International Journal of Forecasting*, 5, 585-588.
- Bacon, D.W., & Watts, D.G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika*, 58, 525-534.
- Bates, J.M., & Granger, C.W. (1969). The combination of forecasts. *Operation Research*, 20, 451-468.
- Box, G.E.P., & Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control*, San Francisco, Holden Day.
- Box, G.E.P., & Jenkins, G.M. (1976). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.
- Bunn, D.W. (1975). A Bayesian approach to the linear combination of forecasts. *Operational Research Quarterly*, 26, 325-329.
- Cancelo, J.R., & Mourelle, E. (2005).

- Modeling cyclical asymmetries in GDP: international evidence. *Atlantic Economic Journal*, 33, 297-309.
- Chan, Y.L., Stock, J.H., & Watson, M.W. (1999). A dynamic factor model framework for forecast combination. *Spanish Economic Review*, 1, 91-121.
- Chan, W., Wong, A.C.S., & Tong, H. (2004). Some Nonlinear Threshold Autoregressive Time Series Models for Actuarial Use. *North American Actuarial Journal*, 8, 37-61.
- Clemen, R.T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5, 559-583.
- De Gooijer, J.G., & Hyndman, R.J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, 22, 443-473.
- De Groot, C., & Wurtz, D. (1991). Analysis of Univariate Time Series with Connectionist Nets: A Case Study Of Two Classical Examples. *Neurocomputing*, 3, 177-192.
- De Menezes, L.M., Bunn, D.W., & Taylor, J.W. (2000). Review of Guidelines for the Use of Combined Forecasts. *European Journal of Operational Research*, 120, 190-204.
- Deutsch, M., Granger, C.W.J., & Terasvirta, T. (1994). The combination of forecasts using changing weights. *International Journal of Forecasting*, 10, 47-57.
- Diebold, F.X., & Pauly, P. (1990). The use of prior information in forecast combination. *International Journal of Forecasting*, 6, 503-508.
- Dijk, D.V., Terasvirta, T., & Franses, P.H. (2002). Smooth Transition Autoregressive Models-A Survey of Recent Developments. *Econometric Reviews*, 21 (1), 1-47.
- Faraway, J., & Chatfield, C. (1998). Time Series Forecasting with Neural Networks: A Comparative Study Using the Airline Data. *Applied Statistics*, 47, 231-250.
- Feng, H., & Liu, J. (2003). A SETAR Model for Canadian GDP: Non-linearities and Forecast Comparisons. *Applied Economics*, 35 (18), 1957-1964.
- Fiordaliso, A. (1998). A nonlinear forecast combination method based on Takagi-Sugeno fuzzy systems. *International Journal of Forecasting*, 14 (3), 367-379.
- Franses, P.H., Dijk, D.V., & Opschoor, A. (2014). *Time Series Models For Business and Economic Forecasting*. New York, NY, USA: Cambridge University Press.
- Genre, V., Kenny, G., Meyler, A., & Timmermann, A. (2013). Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting*, 29 (1), 108-121.
- Graefe, A., Armstrong, J.S., Jones Jr., R.J., & Cuzán, A.G. (2014). Combining forecasts: An application to elections. *International Journal of Forecasting*, 30 (1), 43-54.
- Hagan, M.T., Demuth, H.B., & Beale, M. (1996). *Neural Network Design*. Boston: PWS.
- Hansen, J.V., McDonald, J.B., & Nelson, R.D. (1999). Time Series Prediction with Genetic-Algorithm Designed Neural Networks: An Empirical Comparison with Modern Statistical Models. *Computational Intelligence*, 15 (3), 171-184.
- Hassan, S., Khosravi, A., Jaafar, J., & Belhaouari, S. (2012). Load forecasting accuracy through combination of trimmed forecasts, in *Neural Information Processing*, (pp. 152-159). Springer Berlin/ Heidelberg.
- Hibon, M., & Evgeniou, T. (2005). To combine or not to combine: Selecting among forecasts and their combinations.



- International Journal of Forecasting, 21 (1), 15-24.
- Hipel, K.W., & McLeod, A.I. (1994). *Time Series Modelling of Water Resources and Environmental Systems* (Vol. 45). Elsevier.
- Hochberg, Y., & Tamhane, A.C. (1987). *Multiple Comparison Procedures*. Hoboken, NJ: John Wiley & Sons.
- Hsu, C.W., & Lin, C.J. (2002). A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Transactions on Neural Networks*, 13 (2), 415-425.
- Hyndman, R.J. (2012). *Time Series Data Library*. [www.robjhyndman.com/TSDL/](http://www.robjhyndman.com/TSDL/). (12/06/2012).
- Jose, V.R.R., & Winkler, R.L. (2008). Simple robust averages of forecasts: some empirical results. *International Journal of Forecasting*, 24 (1), 163-169.
- Larreche, J.C., & Moinpour, R. (1983). Managerial judgment in marketing: the concept of expertise. *Journal of Marketing Research*, 20, 110-121.
- Larrick, R., & Soll, J. (2003). Intuitions about combining opinions: Misappreciation of the averaging principle. Working paper INSEAD, 2003/09/TM.
- Lemke, C., & Gabrys, B. (2010). Meta-learning for time series forecasting and forecast combination. *Neurocomputing*, 73 (10), 2006-2016.
- Makridakis, S., Wheelwright, S.C., & Hyndman, R.J. (1998). *Forecasting Methods and Applications*. New York: John Wiley and Sons Ltd.
- Makridakis, S., & Winkler, R. (1983). Averages of forecasts: Some empirical results. *Management Science*, 29 (9), 987-996.
- Makridakis, S., Anderson, A., Carbone, R., Fildes, R., Hibdon, M., Lewandowski, R., Newton, J., Parzen, E., & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: results of a forecasting competition. *Journal of Forecasting*, 1 (2), 111-153.
- Marcellino, M. (2004). Forecast Pooling for Short Time Series of Macroeconomic Variables. *Oxford Bulletin of Economic and Statistics*, 66 (1), 91-112.
- McNees, S.K. (1992). The uses and abuses of 'consensus' forecasts. *Journal of Forecasting*, 11 (8), 703-710.
- Miller, C.M., Clemen, R.T., & Winkler, R.L. (1992). The effect of nonstationarity on combined forecasts. *International Journal of Forecasting*, 7 (4), 515-529.
- Newbold, P., & Granger, C.W. (1974). Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society*, 137, 131-165.
- Newbold, P., & Harvey, D.I. (2002). Forecast combination and encompassing". Pp. 268-283 in M.P. Clements & D.F. Hendry (Eds.), *A Companion to Economic Forecasting*, Oxford: Blackwell Press.
- Priestley, M.B. (1988). *Non-linear and Non-Stationary Time Series Analysis*. San Diego, CA: Academic Press.
- Rapach, D.E., Strauss, J.K., & Zhou, G. (2010). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *Review of Financial Studies*, 23 (2), 821-862.
- Reid, D.J. (1968). Combining three estimates of gross domestic product. *Economica*, 35 (140), 431-444.
- Schauberger, G., & Tutz, G. (2014). *Regularization Methods in Economic Forecasting, Advanced Studies in Theoretical and Applied Econometrics. Empirical Economic and Financial Research: Theory, Methods and Practice* (pp. 61-80).

- Switzerland: Springer International Publishing.
- Smith, J., & Wallis, K.F. (2009). A Simple Explanation of the Forecast Combination Puzzle. *Oxford Bulletin of Economics and Statistics*, 71 (3), 331-355.
- Stock, J.H., & Watson, M. (1999). A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series. Pp. 1-44 in R.F. Engle & H. White (Eds.). *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W. J. Granger*, Oxford, U.K: Oxford University Press.
- Stock, J.H., & Watson, M.W. (1999). Forecasting inflation. *Journal of Monetary Economics*, 44 (2), 293-375.
- Stock, J.H., & Watson, M.W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23 (6), 405-430.
- Subba Rao, T., & Sabr, M.M. (1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. New York: Springer-Verlag.
- Suykens, J.A., De Brabanter, J., Lukas, L., & Vandewalle, J. (2002). Weighted least squares support vector machines: Robustness and sparse approximation. *Neurocomputing*, 48 (1), 85-105.
- Terasvirta, T., & Anderson, H.M. (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics*, 7, 119-136.
- Terui, N., & Van Dijk, H.K. (2002). Combined forecasts from linear and nonlinear time series models. *International Journal of Forecasting*, 18 (3), 421-438.
- Timmermann, A. (2006). Forecast combinations. Pp. 135-196 in G. Elliott, C. Granger & A. Timmermann (Eds.), *Handbook of Economic Forecasting*. Elsevier.
- Tong, H. (1978). On a threshold model. Pp. 575-586 in C.H. Chen (Eds.). *Pattern Recognition and Signal Processing*. Sijhoff & Noordoff. Amsterdam.
- Velleman, P.F., & Hoaglin, D.C. (1981). *Applications, Basics and Computing of Exploratory Data Analysis*. Boston: Duxbury Press.
- Winkler, R.L., & Clemen, R.T. (1992). Sensitivity of weights in combining forecasts. *Operations Research*, 40 (3), 609-614.
- Woodward, W.A., Gray, H.L., & Elliott, A.C. (2011). *Applied Time Series Analysis*. London, U.K.: CRC Press.
- Zhang, G.P. (2001). An Investigation of Neural Network Model for Linear Time-Series Forecasting. *Computers & Operations Research*. 28 (12), 1183-1202.
- Zhang, G.P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159-175.
- Zhang, G.P., Patuwo, B.E., & Hu, M.Y. (1998). Forecasting with Artificial Neural Networks: The State of the Art. *International Journal of Forecasting*, 14 (1), 35-62.
- Zou, H., & Yang, Y. (2004). Combining time series models for forecasting. *International Journal of Forecasting*, 20 (1), 69-84.
- Zou, H.F., Xia, G.P., Yang, F.T., & Wang, H.Y. (2007). An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. *Neurocomputing*, 70 (16), 2913-2923.